G01: Are we there yet?

A road map to the human past

WC 7958

If we could trace our family trees back far enough, we would all find our roots in Africa. Once upon a time we all looked more like the Khoisan — or as they were called when we were young, the Pigmies or Bushmen — of the Kalahari, that is, short in stature with yellowish brown skin with brown eyes and dark pepper-corn hair. Whether we spoke a language with those strange !clicks! the Khoisan and some other Africans use, no one can tell because our remote ancestors left no records other than their bare bones.

What little we think we actually "know" about our remote ancestors back in Africa we have deduced from archaeological remains and from sciences such as geology, climatology, palaeobotany and increasingly, from genetics.

Khoisan of the Kalahari

Although genetics is not a new science — remember the old monk, Mendel and his peas? — its application to the study of genealogy and pre-history and other, related disciplines is very new and in many ways, very much the child of the 21st Century. As a "popular" discipline, DNA-based genealogy (that is, the use of genetics to ascertain historical human relationships) has now passed the point some call "take off" and we are seeing greatly increased numbers of people employing genetic laboratories to study their genes and feed back information which can be used — at one extreme — to tell them where and when their remote ancestors were back in Deep Time; and — at the other — to provide evidence that individuals are or are not related to one another. So, for example, among other uses, genetic "testing" is being used:

- to determine whether or not a person's remote ancestors were "Celtic" or "Viking" or "Native American";
- to clarify the genetic relationships among men who descend from Scots clans or the Ui Neil dynasty of Ireland;
- to help decide how and when the ancestors of the present-day Japanese, Australian Aboriginals or Maori came to their homelands;
- to track the advent of agriculture (ie, the Neolithic) into Europe;
- to determine whether the so-called "Celts" of Britain were the aboriginal inhabitants after the LGM (Last Ice Age) or an invasion from Continental Europe maybe a thousand years before the Roman occupation…..

Long-lost cousins

There are many similar questions to which genetics might be able to provide some answers, questions about times and places long before our ancient ancestors began to record their history. But there are other questions which perhaps the majority of people interested in this popular application of genetics are keener to have answered. These are questions about their own family trees in historical times. Of these, the most common question is "are there any long-lost 'cousins' out there?" where, by 'cousins', they mean people to whom they were related but who lived so long ago that paper-based records no longer exist.

This is the promise of genetic genealogy, the one which makes the laboratories their greatest profit and which probably causes more disappointment than any other aspect of the emerging discipline. The truth is that until now, not enough people have recorded their genetic data in the public data-bases to make it possible for all but the most lucky of clients to identity relatives, even of a generation or so ago let alone back at the time of the Domesday Book. But, that said, numbers are increasing and, if genetic genealogy sees an increase in resources similar to what we saw for paper-based genealogy in the last couple of decades, then it will eventually be better able to keep its promise and find at least some of those "long-lost cousins".

Who can help?

Whether you are hoping to find lost cousins or maybe discover if your remote ancestors could have had a hand in painting the bison in Lascaux[1], the process is the same: you engage one or other of the DNA testing laboratories to analyse your genes for you and report back with information you can use maybe to solve the riddle. Such laboratories are springing up like rabbits but it is best to find out more about them before you commit yourself and your money. Perhaps the best known of those specialising in the kind of analysis needed for genetic genealogy are the following:

[1] Lascaux I, of course: Lascaux II which is the one now open to the public, was created last century.

Ethnoancestry	*http://www.ethnoancestry.com*
Family Tree DNA	*http://www.familytreedna.com/default.asp*
DNA Heritage	*http://www.dnaheritage.com*
Oxford Ancestry	*http://www.oxfordancestors.com/index.html*

I do not recommend Oxford Ancestors even though you will be tempted if you read *The Seven Daughters of Eve*. The reason is simple: they charge much more than the other laboratories and provide too little information. On the other hand, OA do provide you with beautifully printed certificates, "suitable for framing"……

The other laboratories offer a smorgasbord of tests from which you can choose those which suit you best. If your main interest is to help find "long lost cousins" then probably the best way to go is a 37-marker test of your y-DNA (I will explain these terms later but note that only men have yDNA and so women must ask a near male paternal relative to test). Several laboratories offer such a package, some including more than 37 markers. As with mega-pixels in a digital camera, so with genetic genealogy — the more markers, the finer the resolution. To give you some idea, the most popular test among family historians looking for "matches" on their family tree is the 37-marker test offered by FamilyTree-DNA which currently costs about AUD 210 (USD 149). Alternatively, a similar test (but including 43 markers) from DNA-Heritage is priced at about AUD 282 (USD199). Ethnoancestry is probably less useful for family history purposes but in my view, is excellent for those of us interested in what is termed "Deep Ancestry"… that is, researching where one's ancestors were back in remote prehistory.

However, don't be guided by cost alone: choice of laboratory must also involve some assessment of your aims and intentions and the recognition that not only will you help the science advance but your interests will undoubtedly change as you gain experience. I have been tested by all of the above laboratories and with the exception of OA as already explained, I received excellent service from all. I started by wanting to find which of Sykes' "clans" I belonged to and hoping to find "long lost cousins" but gradually grew more and more interested in pre-history and Deep Time. For me, now, it is convenient and cheaper to patronise Ethnoancestry because they have my DNA on file, I don't have to pay new collections fees, and most importantly, they are very oriented to research into Deep Ancestry. However, if I were interested in family history and looking for "cousins", lost or otherwise, then maybe I would stick to FT-DNA or DNA-Heritage[2] because they run special projects (surname, Scottish Clans etc) which bring people together with the same surname or from a known locality to share test results and help each other analyse them. FamilyTree also offer a reduced price to members of surname and other projects so, if you are interested, join an appropriate project and order from there!

[2] In point of fact, I regularly use both of these laboratories for y-DNA testing but chose FT-DNA for mtDNA tests.

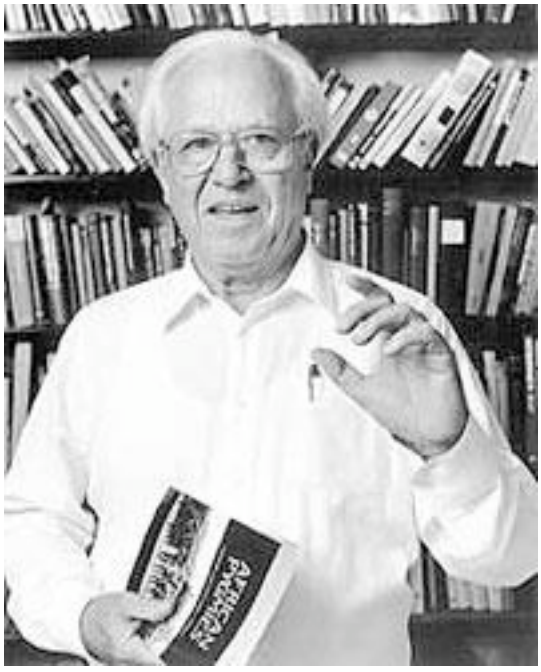
As in most areas where companies are marketing their products in a competitive field, even these ethical laboratories tend to present what can only be described as an "optimistic" view of what their service will achieve for you, the customer. For example, this is the marketing blurb for one company's 37-marker test:

> *The ideal beginners test. With the aid of this you are able to find relatives across the entire world and your geographic origin along the paternal line. The result shows you your Y-DNA profile of 37 markers and your Haplogroup. If 37 markers of 2 men match, this indicates with high probability that their relationship is recent.*[3]

This is not a matter of getting your results back, typing them into a data base and Bingo!....there are the lost cousins. The chances of a useful close "match" are remarkably slim, only slightly improved if you share the same surname. The best use of your test results in such a search is to join a "surname project". Further, the blurb does not say that the Haplogroup is estimated — only a SNP test can determine this definitively and if you do not belong to the same haplogroup then, no matter what the markers say, you cannot be related since Adam was a boy.

Before we get on to the actual genetics of it all, I want to digress a little into a couple of issues which I know you will raise if I don't first head you off at the pass…. These issues are *blood* and *race*.

Blood



Culturally we have all kinds of beliefs and values which have become traditionally associated with the word: for example, "blood is thicker than water", or "the blood which was spilled on ANZAC beach…." All cultures recognise the importance of blood to human life and institutionalise it one way or another into their language and most dramatically, into their religions. So, for example, on the one hand we have blood sacrifice while on the other, the ritual impurity of menstrual blood.

Luigi Luca Cavalli-Sforza

Perhaps more importantly for this course, blood is culturally assumed to be the mystical force which binds people together in kinship systems. So, for example, it is believed that members of a tribe are all descended from a single founding father,

[3] http://www.igenea.com/genealogy-dna-familytree-surname-17-haplogroup-13.htm

often expressed as "of the same blood"…. And of course, we have blood as an indicator of superior status, as in "blue blood"…..

As it happens, blood was the basis chosen by one of the founding fathers of genetic genealogy and population genetics, Luigi Luca Cavalli-Sforza. Until the 1960s, the science of demographics was fatally biased by the obsession of the times with race and jingoistic nationalism. Blood groups, it was thought, would be a more objective way of examining human groups. However, although more objective than other measures or "mis-measures" of Man, blood groups proved to be in many ways too broad to provide much information on ancient demographics. There have been studies which demonstrated, for example, the Japanese and Mongolians might have shared a common ancestry at some time in the remote past[4], but blood, for all its merits, lacked the resolution to determine finer points of relationship between peoples and so, as history shows, Cavalli-Sforza transferred his research to genes rather than blood.

The question of "race"

When I was an undergraduate student of anthropology we learned there were three main races among humans, Caucasoid, Negroid and Mongoloid. In more common parlance, these were the white, black and yellow divisions of human-kind. Although based obviously on skin colour, all manner of other properties were attributed to these groups. There was an assumption that these branches of the "human race" [5] were the descendents of earlier hominids, in the case of the Caucasoid or "white race", for example, from *Homo neanderthalensis*, or the Mongoloids, from so-called "Peking Man". Much of physical anthropology in those days — the early 50s — consisted in measuring length of femur, breadth vs length of skull and so on. This was known as "morphology" and, although by no means fool-proof, is still used extensively by archaeologists today.

Since the 1930s however, scientists have consistently demonstrated that the differences *within* so-called "races" are significantly greater than those *between* such groups. Genetics goes one step further and shows that despite the colour of our skins, we are all one species.

That however, has not put an end to the concept of "race": recent research shows that genetics is beginning to be able to distinguish not only race but ethnicity. For example, in their study *Discerning the ancestry of European Americans in genetic association studies*[6] Alkes L. Price and colleagues were able to distinguish

[4] Komatsu F, Hasegawa K, et al. : "Prevalence of diego blood group Dia antigen in Mongolians: comparison with that in Japanese.", PMID: 15062749

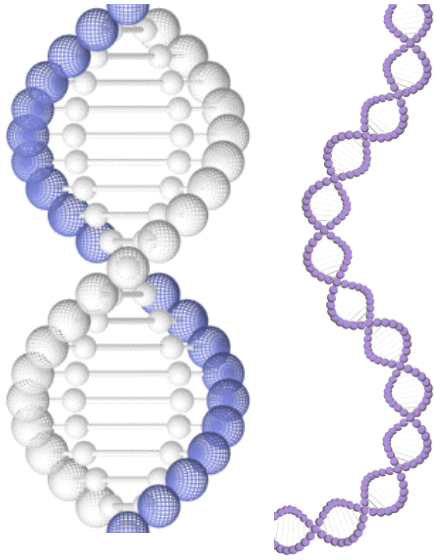
[5] Although many in earlier centuries believed the Negroid people were not human because they did not possess a soul.

[6] Price AL Butler J, Patterson N, Capelli C, Pascali VL, et al.,Discerning the ancestry of European Americans in genetic association studies. doi:10.1371/journal.pgen.0030221.eor (2007) - this is a pre-publication draft.

Ashkenazi Jewish Americans from other Caucasoids, presumably because of their Middle Eastern origins. This boils down to the fact that when a people spend long periods of time relatively isolated from other human groups and during that time, inter-breed more or less only among themselves, they develop genetic characteristics which, with our increasing knowledge, we are beginning to distinguish. The fact the Ashkenazi Jewish Americans can be distinguished from other Caucasoid Americans further suggests that it is not only geographical but also cultural isolation which can lead to the development of distinguishable genetic patterns — Ashkenazi Jews, for example, have retained a cultural isolation in that they intermarry more or less exclusively within their own cultural community.

Nonetheless, to take this further and attribute characteristics other than genetic ones to people is to take a much bigger step. The Nobel laureate, James Watson recently caused a scandal when he suggested Africans had evolved to be less intelligent than non-Africans[7].

However you define it, "race" demonstrably is a social construct which can have a very powerful effect upon people's health, wealth and quality of living. Geneticists suggest that instead of relying on such vague and emotionally-charged notions when devising drug trials or extracting population statistics, researchers should type people by their genetic differences rather than by superficial variables.



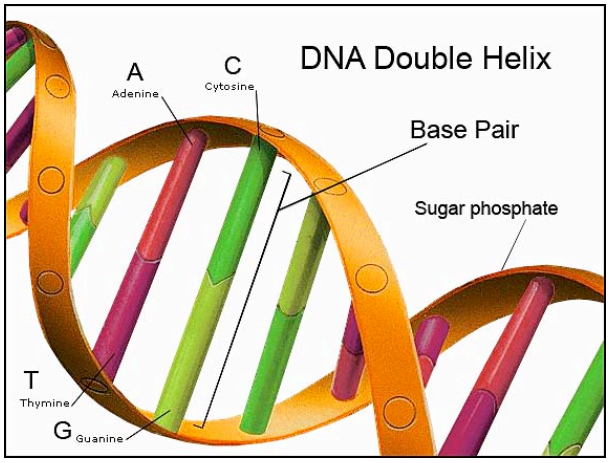
The famous "double helix" of DNA

Basic "Gene-speak"

If you have ever travelled in a country where the natives are clever enough to speak a foreign language, you have probably invested in one of those small books which are crammed with useful words and phrases. The classic, of course, was "La plume de ma tante est dans le jardin"…. The following are a few words and phrases which you might find more useful than my aunt's pen in finding your way around the genetic landscape. The whole science of genetics is very cryptic, no

[7] The controversy erupted up when James Watson, the Nobel Laureate who discovered the "Double Helix" and headed up the Genome mapping project, said that Africans are less intelligent than others. Elsewhere and less controversially he wrote: ... *there is no firm reason to anticipate that the intellectual capacities of peoples geographically separated in their evolution should prove to have evolved identically."* See Mallory, J: *James Watson Tells the Inconvenient Truth* at http://www.gnxp.com/blog/2007/10/james-watson-tells-inconvenient-truth_296.php

less so because much of it is based on complicated statistics. Fortunately, we don't have to understand the statistics, but there are a few words — such as DNA, chromosome, mtDNA and yDNA — which we must understand because they are part of the basic vocabulary of this new field of study.



Base pairs in the DNA double helix

DNA

DNA stands for DeoxyRibonucleic Acid, the essential genetic material of all organisms. It is found in every cell in your body (except for the red blood cells) and usually forms the famous double helix. Below are two artists' impressions of the "double helix".

DNA is a class of nucleic acids containing phosphate, deoxyribose (a sugar), and the four **bases**. These bases are substances derived from either *purine* or *pyrimidine* and are **adenine (A), cytosine (C), guanine (G)** or **thymine (T)**. There is no need to remember the names — geneticists refer to them simply by their initial: A, C, G or T. Of these, A customarily pairs off with T and G with C, forming what are called "base pairs" a term you will occasionally read in the literature and on Internet "Lists" devoted to sharing genetic genealogy information.

Another term you will hear which relates to the nature of DNA is "**Base sequence**" or more commonly, just "sequence" — this is simply the order in which the various bases are arranged at a particular point: for example, AGTA CGTA. This sequence might be repeated several times before a new sequence is initiated. As we will see later, it is counting these repetitions which forms the basis for much of genetic genealogy so that, when I say that my "DYS19=14", I mean that at that point on my y-chromosome, the sequence TAGA is repeated 14 times. Other people might have a different number of repeats of the TAGA sequence — for example, 13 or 15, to name a couple of alternatives of total repeats. These alternative totals are called ***alleles*.**

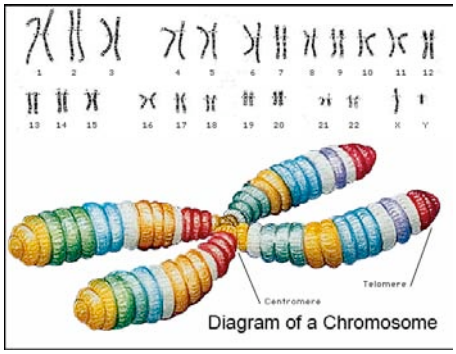
In a sense, DNA is the recipe book or blueprint for making the proteins which make up the organism to which the DNA belongs — but from now on, let's just stick to human beings unless there is reason to refer to some other creature. Its job doesn't end at conception or birth: DNA is also the code book with the instructions for replacing every bit of us throughout the rest of our life. Of course, this copying process is never 100% accurate so small mistakes in the DNA over time alter the structure of the person, a process we call "aging".

Recently, scientists mapped the entire human genome which is a bit like saying they have constructed the index to the code book. However, although they know that all of the pages exist and where to find them, as yet they often don't know what is written on them — they have the structure but not the content for much of the code manual. As it happens, much of our DNA has no known function and is called "junk DNA" although, as some researchers are now saying, this material is not actually useless as the name might imply because it looks as though it switches particular genes on or off!

The **genome**, since I mentioned it, is the total of all the genetic material in all the chromosomes. Its size is usually stated as the total number of base pairs which, in the case of human beings who have 46 chromosomes, amounts to 3 billion base pairs. If you were to stretch this DNA out in a straight line, it would extend for approximately 2 meters. All this is neatly packed into every cell in our bodies, cells so small that very powerful microscopes are needed to see them. Furthermore, the DNA resides in the nucleus of the cell — imagine a very small egg where the yolk is the "nucleus"…. This is packaging on a very small scale!

Chromosome

Chromosomes are responsible for physically transmitting hereditary information. There is a great deal of debate about what is actually transmitted this way. For example, one trait which might or might not be "genetically inherited" is homosexuality (as per the so-called "gay gene" debate which has been going on since 1994). However, there are many features we definitely know to be genetically transmitted from one generation to the next — for example, eye colour, ear shape, metabolic rate, and sadly, a host of nasty disabilities.[8]

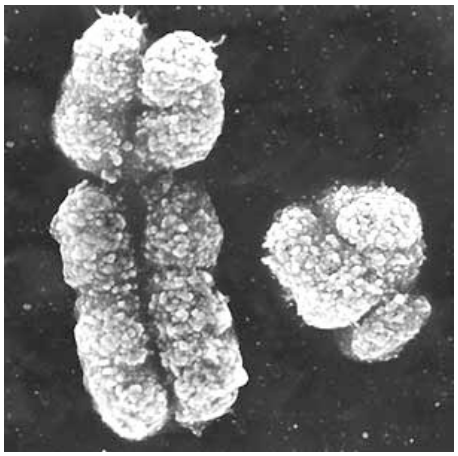


Human chromosomes.

Human beings have a total of 46 chromosomes which are to be found in the nucleus of each and every cell in our bodies. A chromosome is created by combining a strand of DNA with small proteins called **histones**. Once combined, these then wrap themselves up in such a way that they form a kind of elongated X.

Twenty-two of our chromosomes are common to us all, irrespective of gender, and are called *Autosomes*. Until now, autosomes have been sequenced mostly for forensic purposes — eg testing for paternity or other kinds of consanguinity, or

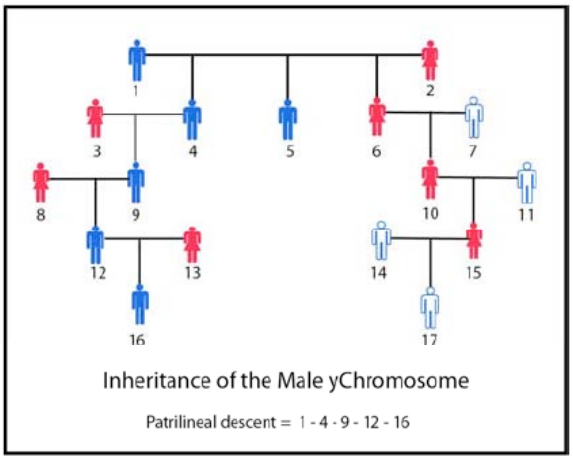
[8] For example, the Fragile X syndrome is the most common known single genetic cause of autism. I am indebted to the excellent web-page of the US National Fragile X Foundation for this information and for their illustrations reproduced here: *http://www.fragilex.org/html/home.shtml*



more familiar to us all, identifying murderers, rapists and the bodies of the victims of terrorists and tsunamis.[9] Recently, however, genetic genealogists have been starting to explore autosomes as a resource for information concerning their ancestry, but for the time being this is outside the scope of this course.

X and Y chromosomes

And of course 2 of our chromosomes at the point of conception determine whether we will be male or female. These are labelled the X and Y chromosomes. Of these, the X is the one which makes you your mother's daughter, while the Y makes a man of you. The Y chromosome is inherited by transmission in direct line from father to son, as in the diagram below:



Inheritance of y-chromosome

The X chromosome is effectively the "default" chromosome so, unless a Y shows up, the baby will be female. Hence, there is little on the X to do with sex determination. Most of the information encoded on it has to do with non-sexist human properties so that the X is relatively large compared to the Y. Although the Y chromosome[10] is really the runt of the litter, it is nonetheless approximately 60 million base-pairs long.

I will return to Y-DNA and how it is used in genetic genealogy a little later, but not so much on the principle of "ladies first", but rather because it was studied first, let's look at mitochondrial DNA….

Part II — Mitochondrial DNA

A few years ago a friend lent me a book to read while the rest of his house-guests went to the beach. I've never been a beach-type person because I don't like getting sun-burnt or sand in places sand ought not to be. The book was called "*The Seven Daughters of Eve*" and was written by Brian Sykes, Professor of Human Genetics

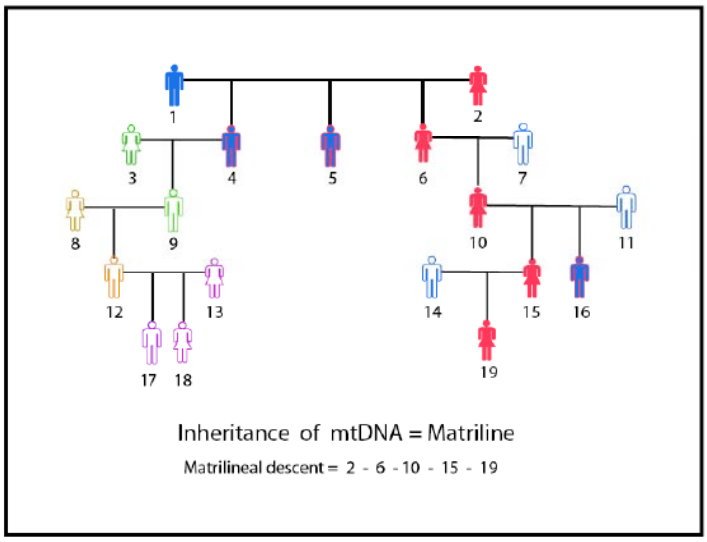
[9] However, the FBI apparently also use a part of the mtDNA, called the "control region", which is highly variable from person to person. A database of sequences from a diverse selection of people allows for fast and accurate identification. *ttp://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm*

[10] Incidentally, this is called "Y" because its arms are not more or less equal like the normal chromosome shape: y-chromosomes have one axis shorter than the other.

at the University of Oxford. Sykes had become famous outside academic circles for his investigations of ancient DNA in archaeological remains, such as the so-called "Cheddar Man" and "Ötzi the Iceman". *The Seven Daughters of Eve*[11] book however, became a best-seller and not only established him as the authority on the genetics of remote ancestors but also propelled genetic genealogy into the public domain.

I was totally enthralled by the book and spent the rest of my holiday reading and discussing it endlessly with my host who was equally captivated by the notion of being able to know when your remote ancestral mother first entered Europe. Sykes describes seven such women — hence the title of the book — who were descendants of that first mother whom scientists had affectionately called "Eve" when they found her fossilized remains back in Africa. These seven "daughters" were of course not the only women to enter Europe — there would have been many more hunting and gathering during the Pleistocene or digging and planting during the early Holocene, but in the generations to follow, sooner or later, their descendants had only sons or died childless. Of all the women in early Europe, only these seven are connected to us in an unbroken female line.

We will come back to these seven women later, but first we need to explain why it is necessary that our line of descent is maternal and not paternal or both. The secret lies in what is called Mitochondrial DNA — or mtDNA for short — which is the magical substance Professor Sykes extracted from ancient bones and from the blood or saliva of living people. mtDNA is inherited only from our mothers. Men do have mtDNA but we cannot pass it on to our children[12]. So, my mtDNA came from my mother and hers came from my grandmother and so on back until — as far back as I know — a woman living in rural Somerset passed her mtDNA on to a daughter who emigrated to Australia in the 1870s. My son does not carry my mtDNA — he got his from his mother and traces his mtDNA back to an ancestral mother elsewhere in 19th Century England and so of course, on and on back through the centuries to one of the Seven Daughters of Eve who may or may not have been the same as my most remote grandmother.



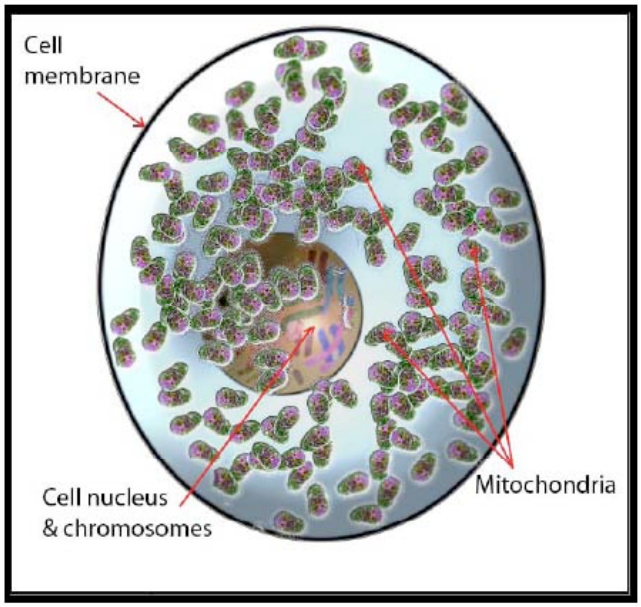
mtDNA inherited in female line

[11] Sykes, Bryan: T*he Seven Daughters of Eve*, 2001, ISBN 0-393-02018-5

[12] Sperm actually carry a small amount of mtDNA but this is apparently killed off by the mother's mtDNA after conception has occurred.

But what is mtDNA? As it happens, that is quite a story….

Once upon a time, and that was a very long time ago, a bacterium set up house inside a cell of one of the earliest animal life-forms on our planet. The host might have been just a single cell organism — certainly, it was not very complicated, nothing as sophisticated as a dinosaur or those mouse-like creatures which were the ancestors of all mammalian life on earth. The bacterium liked its new home so much it came to a permanent leasing arrangement with the cell which it was inhabiting. This symbiosis was so mutually agreeable, in fact, that successive generations of the bacteria set up house in later generations of cells and so began a process which has lasted until this day. Every living cell — except red blood cells — contains later generations of that original bacterium and, although the architecture of their accommodation might have altered over the millions of years — certainly whole cities of cells have evolved — the tenants have changed very little. Every so often a minute change can occur, a change we call a ***mutation***, which can be related one to another in a chain of events reflecting the history of the human race. That original bacterium, now lost in the mists of time, was the ancestor of what we nowadays call "*mitochondria*".



Organelles in a Human Cell.

The little symbiotic bugs — or more politely, "organelles" — occupy about 1/5th of the total volume of the cell in which it is living and, in return for a roof over its head, provides the cell with its energy. In a sense, the mtDNA are the batteries our cells use to keep on doing whatever is their function in life.

Three features of mtDNA are basic to genetic genealogy:

- Compared with nuclear DNA which is rearranged by about 50% with every new generation (because half of the chromosomes come from each parent and re-combine to form the new complement), mtDNA is what is called "non-recombinant DNA" and does not change from generation to generation. The DNA on the y-chromosome is also non-recombinant but we will deal with that later.

- Further, although the organelles which constitute our mtDNA have an ancestry much older than ours, their DNA accumulates mutations fairly rapidly — one about every 20,000 years — so that by noting the mutations, we can back-track down the evolutionary tree using these changes as the stepping stones — or mile-stones if you prefer. The speed of these mutations is important because human beings are basically so similar the world over, it is only by using these rapid mutations we can tell one person from another — or, for paleoanthropological purposes, one human group from another.
- And, fortunately for genetic genealogists, these changes or mutations tend to group together so that we are not faced with as many mutations as there are people! This "grouping together" is of course the feature which Sykes used to define his "clan mothers" in "*The Seven Daughters of Eve*".

While Sykes talks of "clan mothers" and gives them each a name associated with the part of the world into which they were presumably born[13], it is possible to equate these so-called "Ursula", "Xenia", "Helena", "Velda", "Tara", "Katrine" and "Jasmin" with the haplogroups the rest of the world label "U", "X", "H" and so on from the initial letter of their names.

CLAN MOTHER	POSITION OF CHANGE	NATURE OF CHANGE
Ursula	16270	C - T
Xenia	16223	C - T
Velda	16298	T - C
Tara	16126	T - C
	16294	C - T
Katrine	16224	T - C
	16311	T - C
Jasmine	16069	C - T
	16126	T - C

Oxford Ancestors' exposition of the mtHaplogroups

Haplogroups are large groups of people whose mtDNA is the same. At its simplest and historically, earliest level, the mtHaplogroup was defined by the position on what was called the Cambridge Reference Sequence at which a change or mutation was observed during genetic testing. Although we will look only at this earlier practise (for the sake of simplicity), it must be noted that these days the whole mtDNA can be — and preferably is — sequenced in order to obtain greater resolution and allow for greater certainty in associating the Haplogroup with a geographical region where it is assumed the person lived whose mtDNA mutated and thus founded the line.

The Cambridge Reference Sequence, against which all mtDNA sequences are compared in the earlier method, was adopted in Cambridge (hence the name)

[13] I am not certain if this is an "aide memoir" or just an annoying marketing strategy, "dumbing it down" for the sake of book sales. Steven Oppenheimer does the same name-dropping in his most recent book, *The Origins of the British,* although his male names are even more cryptic..

because it is the most common sequence of mtDNA found in Europeans. Those people whose mtDNA sequence matches the CRS are Haplogroup H (or, in Sykes' nomenclature, Clan Helena).

All those who diverge from the CRS, depending on where in the reference sequence their mutation occurred, are given another label, as in the table which has been taken from Oxford Ancestors and is Sykes' nomenclature. The "change" or mutation is where — as in the first line — a C changes to a T in the sequence.

The following is the section of the mtDNA sequence which is used to determine the haplogroup to which a person belongs. This is the simplest and oldest area of comparison and is generally referred to now as HVR1: other laboratories are now testing other areas of the mtDNA sequence which they have labelled HVR2 and HVR3[14] and, as just mentioned, genealogists strongly recommend testing the whole of the mtDNA. The whole sequence of mtDNA however is too long to include here.

ATTCTAATTT AAACTATTCT CTGTTCTTTC ATGGGGAAGC AGATTTGGGT
ACCACCCAAG TATTGACTCA CCCATCAACA ACCGCTATGT ATTTCGTACA
TTACTGCCAG CCACCATGAA TATTGTACGG TACCATAAAT ACTTGACCAC
CTGTAGTACA TAAAAACCCA ATCCACATCA AAACCCCCTC CCCATGCTTA
CAAGCAAGTA CAGCAATCAA CCCTCAACTA TCACACATCA ACTGCAACTC
CAAAGCCACC CCTCACCCAC TAGGATACCA ACAAACCTAC CCACCCTTAA
CAGTACATAG TACATAAAGC CATTTACCGT ACATAGCACA TTACAGTCAA
ATCCCTTCTC GTCCCCATGG ATGACCCCCC TCAGATAGGG GTCCCTTGAC

The Cambridge Reference Sequence for HVR1 (from OA)

When I got my results back from Oxford Ancestors in 2003, I received a print-out of the CRS above, BUT there was one letter printed in red, as below:

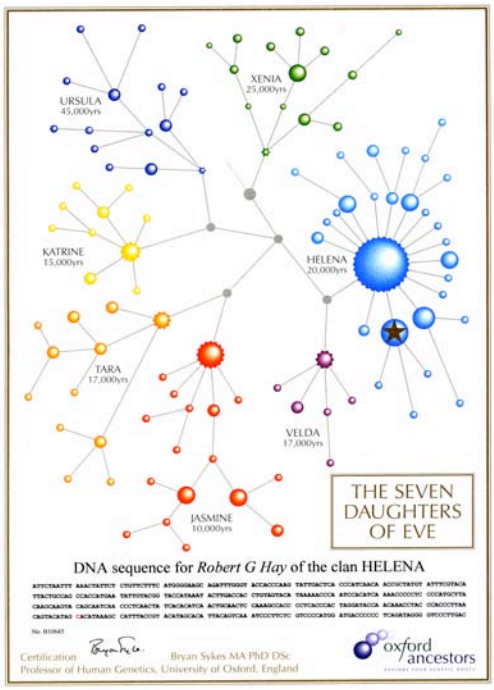
ATTCTAATTT AAACTATTCT CTGTTCTTTC ATGGGGAAGC AGATTTGGGT
ACCACCCAAG TATTGACTCA CCCATCAACA ACCGCTATGT ATTTCGTACA
TTACTGCCAG CCACCATGAA TATTGTACGG TACCATAAAT ACTTGACCAC
CTGTAGTACA TAAAAACCCA ATCCACATCA AAACCCCCTC CCCATGCTTA
CAAGCAAGTA CAGCAATCAA CCCTCAACTA TCACACATCA ACTGCAACTC
CAAAGCCACC CCTCACCCAC TAGGATACCA ACAAACCTAC CCACCCTTAA
CAGTACATAG CACATAAAGC CATTTACCGT ACATAGCACA TTACAGTCAA
ATCCCTTCTC GTCCCCATGG ATGACCCCCC TCAGATAGGG GTCCCTTGAC

Figure 10: My HV1 sequence from Oxford Ancestors

Since my sequence showed no changes at any of the points, I was judged to be "Clan Helena" — or Haplogroup H in normal genetics terminology. The red C however, indicated one mutation at position 16311 (normally the "16" is dropped) in which the original T was replaced with a C. My mtDNA signature then was

[14] HVR stands for hypervariable control region. Most recently, the sequencing of the entire mtDNA has been introduced. This gives much greater resolution which in turn allows for better understanding of geographical origins.

"311 T-C". To get any further I would need to do at least an HVR2 test to enable me to say more than that my remote ancestral mother was among the people who, about 20,000 years ago when the Last Ice Age was at its hight, retreated from the advancing cold to a refuge in the South of France.



My mtDNA Certificate from Oxford Ancestors: beautifully printed but not worth the expense.

There — or so OA told me — in the valley of the Dordogne, she lived with the people responsible for the magnificent paintings in many of the caves of the region.

And how do the geneticists know that?[15] In a feat of statistical wizardry and possibly some slight of hand, geneticists know approximately the time at which that mutation represented by my red C occurred. Furthermore, because they get a concentration of similar test results focussed on the particular geographic location, they feel confident to tell me that it was there, near the Dordogne, that one of my great^{n} grandmothers not only survived the chill dry winds and harsh conditions of the LGM, but gave birth to a daughter whose descendants resulted in me and about 45% of the rest of us who are of European descent.

Part III — Sons of Adam

As I have already said, the y-chromosome is handed down only from father to son and so, if women want to use y-DNA to study their remote ancestry, they must persuade their father or grandfather, and if they are very confident he is truly his father's son, then their brother to provide the saliva sample for the tests. I say this in all seriousness because geneticists estimate that about 5% of all births are what they call "a non-paternal event"… Of course, the same can be said of mothers and their children but people are usually more certain who their mother was except in cases of adoptions or, these days, donor eggs…. Genetic testing, especially when more than one member of a family is involved, can bring unwelcome surprises in its wake.

Just as there was a Mitochondrial Eve, so there was a y-Chromosome Adam, a real man who existed a long time ago to whom all living lineages can be traced. Once

[15] For more information about genetics, see the National Center for Biotechnology Information.at http://www.ncbi.nlm.nih.gov/About/primer/index.html

again, the ubiquitous Professor Sykes has got in on the act with his 2003 book *Adam's Curse – a future without men*[16], in which he tracks down his distant ancestor, a Mr. Sykes who lived in a village called Flockton, near Huddersfield in Yorkshire in the 13th Century. Sykes also used y-DNA tests to establish that men called Sykes in adjoining districts were descended from the same source. This book canvasses several topics, expanding on the topic of how men are made (biologically, not as a result of cold baths or life in the Army), a hypothesis about what causes homosexuality, raises the question whether sex is really necessary and finishes with the news that the y-chromosome is getting smaller as the X-chromosome cannibalises it.

While a good read and sometimes informative, *Adam's Curse* is not as inspirational as *The Seven Daughters of Eve*[17]. More informative is Steven Oppenheimer's 2004 book, *Out of Eden*[18] in which evidence derived from both mtDNA and y-DNA is used to trace the paths of our remote ancestors after they came out of Africa some 80 kya[19].

Most recently, both Sykes and Oppenheimer have published new books, *Blood of the Isles*[20] and *The Origins of the British*[21] respectively. Both raise interesting interpretations of the genetic and other data so far known about the remote ancestry of our species but both suffer by apparently using limited data and also by failing (for commercial reasons) to disclose their data bases. Also, both accept assumptions which although received wisdom when the books were first published, have recently been over-turned and thrown genetic genealogy of this kind on its head. We will discuss this in detail later on, but meanwhile, these are good "reads" but treat much as though it were fiction…..

There are currently two kinds of data obtained from y-chromosomes, one called "**Haplotype**" and the other "**Haplogroup**". The y-haplogroup, parallel to the mt-DNA one, refers to a mutation which a number of men share, having inherited it from a "founding father". The tests which determine which y-Haplogroup a man belongs to is called a **SNP-test** where SNP refers to *Single Nucleotide Polymorphism*. Many of these mutations have been located by geneticists and by the wizardry of science, the time and location at which they occurred has been estimated with some degree of certainty. I will return to this a little later on.
A haplotype refers to the genetic constitution of an individual chromosome. It is often described as a man's "*genetic signature*" although this is perhaps more

[16] Sykes, B. *Adam's Curse - a future without men*, Bantam Press, 2003
[17] op. cit.
[18] Oppenheimer, S: *Out of Eden - the peopling of the world*, Robinson, 2003/4
[19] a shorthand way of writing "thousand years ago"… can be either upper or lower case.
[20] Sykes, B: *Blood of the Isles - exploring the genetic roots of our tribal history*, Bantam Press, 2006. In the US this was published as *Saxons, Vikings, and Celts: The Genetic Roots of Britain and Ireland.*
[21] Oppenheimer, S: *The Origins of the Britiswh - A genetic detective story: the surprising roots of the English, Irish, Scottish and Welsh*, Carrol and Graf, 2006.

correctly reserved for comparing the pattern of genes of men who have a surname in common. Now, this starts to get complicated…..

When you order a y-DNA test you are sent a packet containing some kind of swab, a bottle with a special liquid inside it and instructions how to rub the end of the swab on the inside of your cheek and then place it in the bottle which you mail back to the testing laboratory (techniques vary: Enthnoancestry, for example, collect 10ml or so of saliva in a special little container).

The FT-DNA collection kit.

Then, some time later — depending on the demands on lab time and also on what they call "queueing" because they wait until they have a big enough batch of similar tests to run through at the same time — you get back an email with a list of numbers of which at first, you cannot make head nor tail.

I now have a total of 77 markers to puzzle over, but for the purposes of demonstration I will use my original Oxford Ancestors results for which I paid through the nose and for which I might as well get some benefit. And I will use them because there is only 12 of them and that makes the demonstration much easier than trying to deal with 37 or 67 even though, as everyone will tell you, the amount of information you can get from 12 markers is very, very limited.

DYS→	393	390	19	391	385a	385b	426	388	439	389i	392	389ii
BH	13	25	14	10	11	14	12	12	11	13	13	29

Well, what do all these numbers mean? From here on it is a matter of reading books and journal articles and most importantly, following the messages posted on the many Mailing Lists by experts in the field. Of these lists, RootsWeb-DNA[22] is by far the most informative but because of the complexity of much of what is discussed, it is also worth subscribing to others catering for "newbies" as newcomers to genetic genealogy are called[23].

So, let's briefly break down the above numbers. To start off with, "DYS" is shorthand for **DNA Y-chromosome Segment**. There are many of these used in

[22] http://lists.rootsweb.com/index/other/DNA/GENEALOGY-DNA.html
[23] eg, http://lists.rootsweb.com/index/other/DNA/DNA-NEWBIE.html or DNA-ANTHROGENEALOGY-subscribe@yahoogroups.com.

genetic genealogy, some apparently more informative than others. Some, such as DYS-390 in the above, are known to mutate slowly so they perhaps are more indicative of deep time demographics than are those which mutate more rapidly and thus bring us closer to our own time.

Let's look briefly at DYS 390 and use it to define a few other terms. At the location (or locus) DYS 390 you find the base pairs repeat themselves in the pattern TCTA TCTG in the range of 17 to 28 times. In my case, the sequence was repeated 25 times. That "25" is my **allele** — pronounced ['ali:l] — for DYS-390. In other words, an allele is the number of times a sequence is repeated on a particular **locus**. Your results when they come back from the laboratory tell you (1) the locus or DYS which was tested and (2) the allele at that point, i.e. the number of times the sequence was repeated.

Now, the frequency of DYS-390 varies from place to place. My allele of "25" is found most commonly in Russia and the Baltic region of Europe (about 28%) and only about 10% in the British Isles where my known paternal ancestors came from. That, however, does not mean my ancestry is Russian! It is not so much the alleles on individual STR markers which yield the most information but the patterns which emerge when you examine the whole haplotype and compare it with others.

There are two ways this can be done. The first, and the one most eagerly adopted by genealogists more interested in family history in historical times than in remote ancestry in Deep Time is to search the public data bases of haplotypes and hope to find what, in the jargon of genetic genealogy is called "**a match**". So, for example, when I first got my results from Oxford Ancestors, I eagerly searched through the Oxford Ancestors' data base of the haplotypes of people whom they had tested and was over-joyed to find 2 perfect matches, that is, two other men whose scores on the different **STR Markers** were identical with mine (STR Markers, by the way, stands for "**Short Tandem Repeat**" and refers to the allele on one of those DYS "markers"). One man was not all that communicative, but the other, Barry, and I emailed each other for a long time, developing quite a sense of kinship as we did so — after all, the perfect match indicated we were related some time in the past even though his family had always lived near Durham and mine in far-off Nairnshire, in Scotland!

Barry and I spoiled our genetic honeymoon, however, by both sending away for 37 markers from FamilyTree-DNA. When our respective results came back and we compared them, we had dropped off each other's trees — if we were related then it was back in the time when Adam was still in short pants…. that is, we differed on too many markers (there had been too many mutations) for there to be any hope that we had a common ancestor unless it was thousands and thousands of years ago.

When geneticists comb through whole data-bases of people's haplotypes they sometimes find patterns which repeat themselves, indicating that a number of people have haplotypes very similar to each other. It is thought that such a "cluster" indicates a common heritage derived most probably from an ancestor in whom the mutation occurred and whose descendents lived in the same locality long enough to produce a population of men with the same or similar haplotype. I say "similar" because there will likely be later mutations which will produce minor variations in the pattern. Statisticians calculate the "modal" allele for each STR marker (the mode is the highest frequency: so, in the case of DYS 390 the mode is 24, not my minority 25 score).

These patterns, called "**modal haplotypes**" are useful in genetic genealogy in pointing to likely common geographic origins. The largest and best known of these is called the **Atlantic Modal Haplotype** and there is a good chance that the majority of the people doing this course belong to this group. This is a pattern of alleles on STR markers which was previously assumed to indicate that ancestors of the individual lay in the Franco/Iberian refugia during the LGM[24] and that when their remote ancestors ventured north after the Ice retreated, they hugged the Atlantic coast, migrating up the west coast of France, into Ireland and western Britain. This was the assumption on which Oppenheimer, for example, wrote his book *The Origins of the British* but which as I have already said, has recently been overturned. The new implications have yet to be worked out, but according to recent calculations, the origins must lie in migrations of Haplogroup R people into Europe sometime *after* the LGM.

Although commonly expressed as the modal alleles on only 6 STR markers, (DYS # 19/388/390/391/392/393), more value can be found in using all of the 12 markers below. In this table, I have listed the 12 STR markers and their alleles on the AMH and below them, my test results. Below that are shown the differences in the two scores. The **GD or Genetic Distance** my scores are from the AMH is shown in the bottom line. (Arithmetic sign is ignored in assessing GD).

DYS	**393**	**390**	**19**	**391**	**385a**	**385b**	**426**	**388**	**439**	**389i**	**392**	**389ii**
AMH	13	24	14	11	11	14	12	12	12	13	13	29
BH	13	25	14	10	11	14	12	12	11	13	13	29
Diff.	0	**+1**	0	**-1**	0	0	0	0	**-1**	0	0	0

A GD = 1 on any one marker in any direction is regarded as not significant. GDs of 2 or more however, generally mean that the individual — in this case, me —

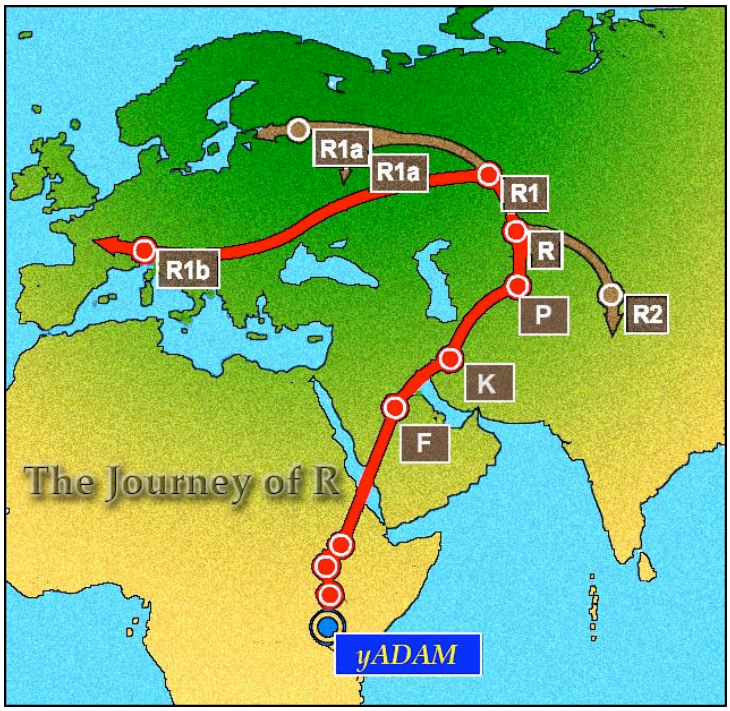
[24] "Last Glacial Maximum" - in common speech, "the Last Ice Age" but this is a misnomer.

does not belong to the Modal Haplotype. Closer to the end of this course we will go into this in much greater detail; suffice to say here that personally I suspected my DYS390=25 and DYS391=10, both relatively slow mutating markers, indicated some alternative ancient origin.

Haplogroups and SNPs

Although Haplotypes — that is, the individual person's set of alleles on STR markers — are potentially useful for family historical studies and have been the primary focus of research until recently, the emphasis is now shifting to y-Haplogroups as measured by SNPs (pronounced 'snips'). Although genetic genealogy "took off" because (mainly American) people wanted to extend the reach of paper-trail family history, books like *The Seven Daughters of Eve, Out of Eden* and the rapidly expanding *National Genographic Project* [25] run by Spenser Wells for National Geographic, among others, have all prompted a growing interest in remote ancestry. Haplogroups are of less use for family history but essential to the study of ancient human migrations.

Over time, the y-chromosome DNA experiences many mutations. A particular kind of mutation called a *Single Nucleotide Polymorphism* (or SNP for short) is so rare that it is considered unique and thus, definitive. If you have this SNP then you and all the other men in the world who have it have an ancestor in common who lived in a particular place at a particular time in history. You and all those other men constitute a **y-Haplogroup.** These haplogroups, as in the case of the mtHaplogroups, are designated by a capital letter; sub-divisions (or **clades** and sometimes sub-clades) by a number and then by lower-case letters. So, in my case, I was recently tested and found to be R-L2[26] because I tested positive to the SNP test L2 aka S139.



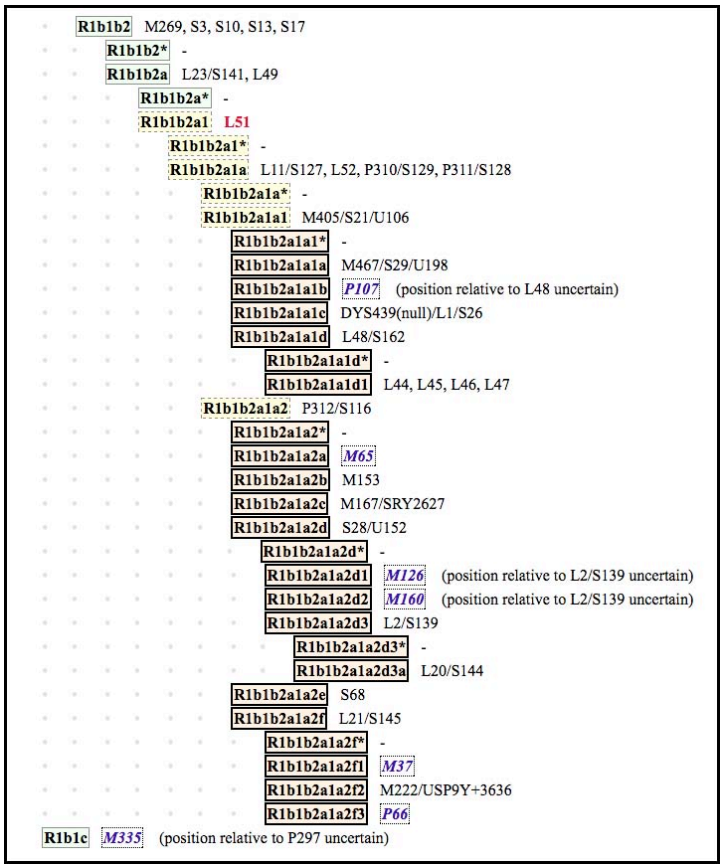
The migration out of Africa leading to R-M269 (Old terminology R1b) in Europe.

Haplogroup R is found in many parts of the world, particularly in India and Western Europe and is probably the most studied of all the Haplogroups of the y-chromosome. It is currently thought to have arisen ~27 KYA in Asia, later splitting into two

[25] https://www3.nationalgeographic.com/genographic/
[26] or R1b1b2a1a2d3 according to the 2009 ISOGG y-chromosome tree at the time of writing.

sub-clades, R1 and R2. Of these, R2 is these days most commonly found in Asia, especially on the Indian sub-continent and in central Asia.

The other mutation, the one which produced R1, probably occurred somewhere in south-western Asia during the LGM about 18,500 years ago. Further mutations produced subclades R1a and R1b, the former arising on the Eurasian Steppe. This sub-clade is today most frequently observed in eastern Europe and in western and central Asia.



Portion of the ISOGG y-Haplogroup phylogenetic tree showing sub-clades of R-M269[27]

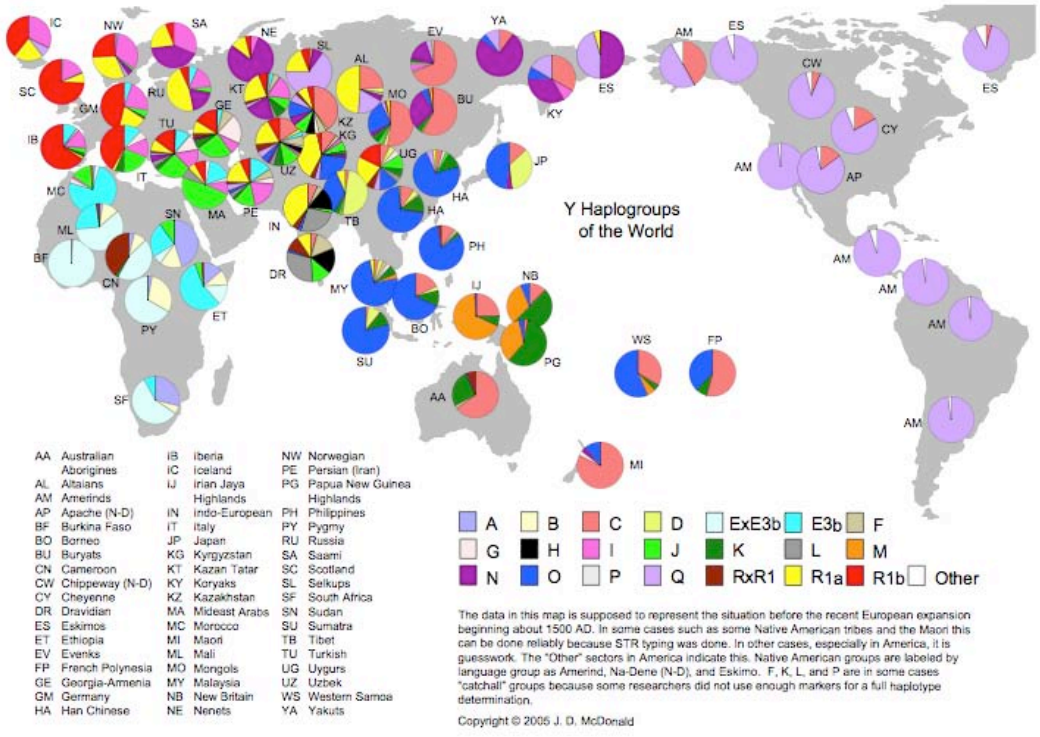
The other sub-clade, R1b is the one we will follow most closely in this course for the simple reason that the majority of men whose ancestry is European belong to one or other of its several sub-clades because earlier mutations of supra-HgR did not find their way into Europe. R1b — or more precisely in current terminology R1b1b2 or R-M269 — entered Europe after the LGM probably carried there by Neolithic immigrants during the warmer days of the Holocene.

These labels on their own don't mean much but we need to remember that each of these mutations occurred in a baby boy born to parents who lived according to the life-style of their time and place in the world. That baby boy grew up to father more men who, like him, carried that mutation and so on down the ages to all those men alive in the world today who belong to Haplogroup R-M269 and its many sub-clades. Of course we will never know the names of those men, but they are nonetheless part of our family history and as essential to it as our fathers and grandfathers whose names we engrave in our genealogies.

Of course there are many other y-haplogroups among the world's male population but what they are and how they got to wherever they are found today is the subject of future sessions. However, some idea of just how many and how wide-spread

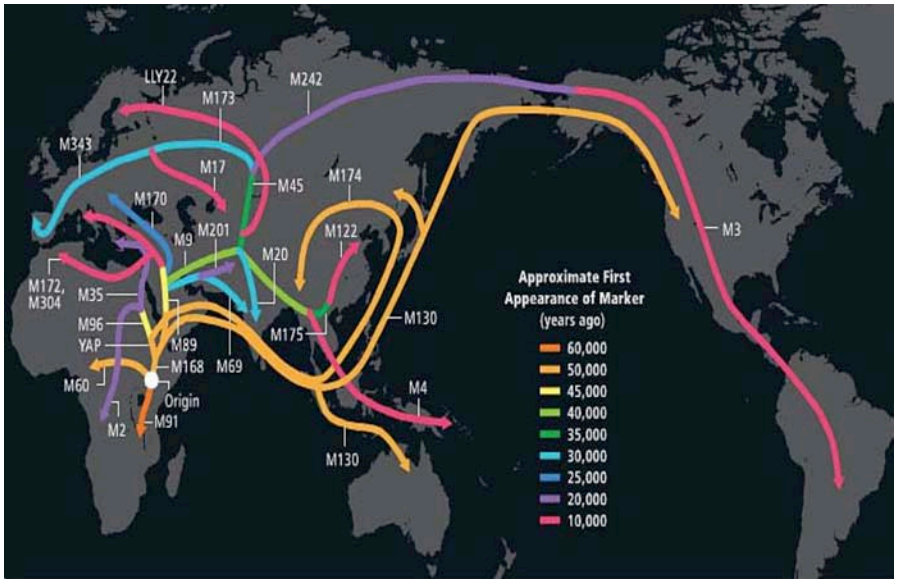
[27] Note that new SNPs are constantly being discovered — see http://isogg.org/tree/ (from which this is taken) for current updates.

they are can be seen in the map below. Note that the use of little pie-graphs is the standard method of demonstration haplogroup distributions.



World-wide distribution of y-Haplogroups

From beginning to end…



Once upon a time and — as I have already said — that was long ago, our ancestors all lived in Africa, although not necessarily on the banks of the great, green, greasy Limpopo River… their children set out walking, through the Gates of Despair, from the Horn of Africa to the super-continent Eurasia and from there, some went east, along the beaches to Australia and South-east Asia, others headed north into the great steppes of Central Asia. From there too, they took their different directions to people the world, a few ending up in the cold wastes of Western

Europe. Ice ages alternated with warm spells, new inventions were made, great paintings were made in caves…

It was a long way from Africa, in distance and time, but our ancestors eventually got there. Even so, such is our human wanderlust and the awful predictions of climate change, can we really be sure we can say "yes" when the kids ask, "Are we there yet?"……
