A Walk in the Garden of Eden

Genetic Trails into our African Past

Himla Soodyall

Social Cohesion and Integration Research Programme, Africa Human Genome Initiative Occasional Paper Series No. 2

Series Editor: Prof Wilmot James, Executive Director: Social Cohesion and Integration, Human Sciences Research Council (HSRC)

Published by HSRC Publishers
Private Bag X9182, Cape Town, 8000, South Africa
www.hsrc.ac.za/publishing

First published 2003

ISBN 0-7969-2021-4

Production by comPress

Printed by Paarl Print, Oosterland Street Paarl, South Africa

Distributed in South Africa by Blue Weaver Marketing and Distribution, P.O. Box 30370, Tokai, Cape Town, South Africa, 7966. Tel/Fax: (021) 701–7302, email: booksales@hsrc.ac.za

Preface

The Human Sciences Research Council (HSRC) publishes a number of occasional paper series. These are designed to be quick, convenient vehicles for making timely contributions to debates, disseminating interim research findings or they may be finished, publication-ready works. Authors invite comments and suggestions from readers.

This paper was originally presented as the first in the Sol Plaatje Lecture Series on Africa, jointly hosted by the Ministry of Education and the Africa Human Genome Initiative at the Iziko South African Museum in November 2002.

ACKOWLEDGEMENTS

This research was supported by the Medical Research Council (MRC) of South Africa, the National Health Laboratory Service, the University of the Witwatersrand and the National Research Foundation.

The author also wishes to acknowledge all subjects who participated in this research by donating a sample of blood for genetic studies and thanks Prof P van Helden and Dr E Hoal (University of Stellenbosch) for DNA samples from the Cape coloured and Cape Malay populations.

Foreword

By Trefor Jenkins

I feel a little bit like I imagine Jeremy Bentham might feel when, on auspicious occasions, at University College, London, he is wheeled out in his chair to preside over august gatherings. Jeremy Bentham, the great philosopher and reformer, one of the founders of utilitarianism, who died in 1832, made a generous bequest to University College, London. The bequest included his body, which was to be dissected by the medical students of that college and, stipulated that afterwards, it should be sent to a taxidermist who would prepare the body and dress him in his favourite suit and hat, and then install him in a chair with wheels. Jeremy Bentham still sits in that chair in the cupboard under the stairs at the entrance to University College, London. And if you are distinguished enough, you may succeed in your request to meet Mr Jeremy Bentham when you next visit London.

Now I'm not here under any duress. It's a great pleasure for me to be wheeled out to introduce to you a former student of mine, Himla Soodyall. In my enforced retirement (having reached the age of statutory senility) I say that I now work for Himla, and I am, indeed, privileged to be in that position. She is certainly teaching me much more than I ever taught her. But before introducing Dr Soodyall I should like to say a few words about the Human Genome Project (HGP) and the recently launched multidisciplinary Africa Human Genome Initiative (AHGI).

I have to confess that, in 1991, I published a paper in which I argued that we should probably not have a human genome project in South Africa. It was published in the *South African Medical Journal* (*SAMJ*),[1] and in it I reviewed the setting up of the project, which had been launched in 1990. I argued that perhaps the time was not ripe for South Africa to really make a significant

1 Jenkins T (1991) 'The Human Genome Project – does South Africa have a role to play in it?' *SAMJ* 80: 52–54.

contribution to this mammoth, mega-project that had just been launched, primarily by the Americans, but soon joined by the British, the French, the Germans, and the Australians. There were very few human geneticists in South Africa at that time and molecular biology was an emerging discipline. A few individual medical scientists in the country had, for a number of years, been contributing to the mapping of the human genome, with small-scale mapping of specific disease loci as well as the testing of DNA from families collected by CEPH (*Centre d'Etude du Polymorphisme Humaine*) in Paris. I argued in my *SAMJ* paper that we had more urgent and pressing uses for our limited research funds at that time. The total budget for the Medical Research Council (MRC) was, as I recall, about R40 million a year; the American Congress had allocated $200 million per year for the projected fifteen years of the HGP.

The term genome refers to the sum total of the DNA that exists in every nucleated cell of an organism. The human genome is all the DNA that exists in the nucleus of the cell of a human being together with the small amount of DNA that exists in the mitochondria the tiny organelles that are found in the cytoplasm of these cells. In terms of size, the DNA molecule is so thin that you couldn't possibly see it with the naked eye. You couldn't, in fact, see it with the most powerful light microscope. You would need an electron microscope to see it because it is so thin. But if the DNA in one cell – and this is true for all the cells with nuclei – were stretched out, that DNA molecule would be three metres long. And if you consider that we have three trillion cells in our bodies, if you were to unravel the DNA in every cell and lay it out end-to-end, it would stretch from the earth to the moon and back 20 or 30 times – I can't remember the exact number! But that is how much DNA exists in the human body. And it is this DNA which conforms to the famous shape of the double helix which was elucidated in 1953 by Watson and Crick, working in Cambridge, England, with some help from their friends, Maurice Wilkins and Rosalind Franklin. It is a truly remarkable molecule consisting of repeating sequences of a number of nitrogenous bases (as they are called), which number in total, along the full length of the DNA in one cell, three billion, that is, 3000 million. There are only four different bases,

each representing a letter in the genetic code: adenine (A), thymine (T), guanine (G) and cytosine (C). But these four letters are sufficient to write the long chemical message encoded in the DNA. There are 64 different ways in which four letters can be arranged in a specific sequence of three letters (and these three letter words are called triplets or codons) – more than enough to code for the specific 20 amino acids which make up the full repertoire of proteins – the main constituents of all living forms. In many cases, more than one triplet will code for one specific amino acid (as a result the code is said to be 'degenerate') and some of the triplets code for a stop signal. The four letters are joined to a backbone constituting a chain and there are two chains (one is complementary to the other), which are wound around one another to form the double helix. It is this DNA molecule which determines how the cell functions and also how the organism reproduces itself. Its information content is enormous and its design is ideally suited for carrying out all these functions.

The goal of the HGP was to sequence the three billion nucleotides, a mammoth task, which many people said could not be completed in the span of 15 years that the scientists had considered to be adequate. Due to the efforts of very distinguished scientists, particularly James Watson (the co-discoverer, with Francis Crick, of the DNA molecule), the Congress of the United States voted $200 million per year for 15 years (at the 1989 value of the dollar). And so the project was launched. Britain was soon to join with, initially, the support of its Medical Research Council and then followed an enormous grant from the Wellcome Trust, totalling many hundreds of millions of pounds. Other countries set up their own human genome projects, but the US and the UK were the major players. An unexpected contribution – and this is significant – came from the pharmaceutical and biotechnology industries which contributed even more funds than the statutory bodies and trusts had together contributed. And thereby hangs a cautionary tale. Pharmaceutical companies and the biotechnology industry do not give money for altruistic reasons. There are shareholders who demand their dividends. So, we are going to have to pay for the benefits that are anticipated to come from the Human Genome Project.

Well, the project began. The pace of sequencing these three billion nucleotides accelerated. It was projected that there would be 80 000 to 100 000 genes to be found. It was already known that about 97 per cent of the genome was what is called 'junk' DNA, i.e. DNA that does not code for anything as far as we can tell. 'Junk' DNA is a term coined by South African-born, and trained, molecular geneticist, and Nobel laureate, Sydney Brenner, to refer to the DNA that, apparently, does not do anything. And when challenged by someone, with the argument that God would not have created us with 97 per cent of redundant or useless DNA, Sydney is said to have retorted: 'I said it was "junk" DNA, not "trash". Everyone knows that you throw away trash. But junk we keep in the attic until there may be some need for it.'[2]

We still don't know what function the junk DNA might have, but, if Sydney is right on this one, as he has been on so many other issues, we will, eventually, learn that it does have some purpose. The other three per cent of the genome constitutes the genes. The HGP was completed in February, 2001, and we now know that the estimate of the number of genes was rather high; it might, in fact, be only 30–35 000 genes that go to make a human being. Now there's a tendency by some people, especially scientists perhaps, to think that we are our genes, that is, that we are only our genes. So let me make my caveat straight away and say that I believe that we are more than our genes. Many people are somewhat nervous of genes – and I believe most of us are to some extent – so they should be reassured that the geneticists are not all committed to what is called genetic determinism. We believe Watson was guilty of hyperbole when, writing about the HGP, he said: 'How can we not do it? We used to think our fate was in our stars. Now we know, in large measure, our fate is in our genes.'[3] I do not believe that everything that we do (our behaviour, our preferences, our dislikes and prejudices) are determined by our genes; neither do I believe that most ill health is due to faulty genes. Unlike other animals, we possess consciousness and an awareness that transcends the strictly biological. We know that we are human beings because of

2 Brenner S (1990) 'The human genome: the nature of the enterprise' *Human Genetic Information: Science, Law and Ethics* (Ciba Foundation Symposium, 149), pp. 6–17. Wiley: Chichester.

3 Jaroff L (1989) 'The gene hunt' *Time Magazine*, 20 March, pp. 62–67.

other human beings (I knew that before I had heard of *ubuntu*, although that's a very good term to describe this concept).

James Watson, who was one of the major protagonists of the HGP, realised very early on that there would be tremendous public opposition to setting up such a project. He feared that the senators and members of congress would not approve the money that was needed. He argued from the beginning that, because of its social implications, the project would allocate three to four per cent of its total budget to a programme called ELSI (ethical, legal and social implications), which would study these implications. And that has in fact happened. There have been more books and papers written on the ethical and social and legal issues raised by the HGP than ethicists have ever written before on a medically related subject. This has stimulated the public debate which has reassured Americans and others in the developed world, that these are not mad scientists simply following their crazy ideas, but are responsible human beings guided by a deepening awareness of the possible abuses to which their discoveries may be put.

If advances in molecular medicine were to lead to a dramatic increase in predictive and preventative approaches to disease management, then individuals, whilst still apparently healthy, will be screened for large numbers of genes, some of which will predispose them to ill health. They will then be counseled to modify life-styles and they may also be offered medication to minimize the risk of developing the particular disease for which they are at risk. Such genetic screening will obviously be voluntary and will only be carried out with the individual's informed consent. The results of the tests will be kept confidential, even though these results may have implications for other family members. Or will the 'at risk' relatives have the right to be alerted to the risk they may run? The doctor-patient relationship may need to be scrutinized anew, with respect to issues of privacy and confidentiality. Such screening-test results will, of course, also be of interest to present, and future, employers, as well as to life insurance and health insurance companies. The state may claim that it, too, has an interest in this information – if it might result in reducing the escalating health care budget, for example. Forensic DNA databases are being set up in many countries, including

South Africa, because of their potential in helping to reduce crime. There is no law in place in South Africa that requires the police service to destroy DNA fingerprint data on the individual who has been acquitted of a serious crime. In the UK it is a legal requirement that such data be destroyed.

The appointment of Dr Malegapuru Makgoba to the presidency of the MRC in 1999 led to a reconsideration by the Council of its attitude to genomics. The completion of the HGP was in sight (it occurred in February 2001 with the public sector publishing the human genome in *Nature*[4] on 15 February and the private sector, represented by Celera Genomics, publishing its version of the genome a day later in *Science*[5]) and Dr Makgoba announced that genomics was to be one of the six priority areas for research, which also included AIDS, TB and malaria. The MRC set up three units to research genomics and bioinformatics, including one headed by Dr Himla Soodyall, and in 2002 the AHGI was launched by the HSRC in partnership with the Academy of Science of South Africa and the Sustainability Institute. The AHGI seeks to ensure that South Africans will keep up with, contribute to and benefit from revolutionary advances in genetic knowledge. Prof Wilmot James has been the driving force behind the creation of this initiative and I wish it every success.

Himla Soodyall is a great all-round scientist, with a passion for her subject, human genetics. She comes from humble beginnings, which I say with some pride, because I think I did myself. Her mother is a schoolteacher and her late father was a clerk at a bakery. She received her early education in Durban and her BSc and Honours degrees were obtained at the University of Durban-Westville. She then had an inspired move to Wits University, and after doing a Master's degree in biotechnology, she came into my orbit and I'm glad to think that my gravity drew her in and may have helped to keep her in human genetics. It's a great pleasure and a source of joy to retired professors to have students continue to work in their disciplines and to take them to greater heights.

4 Lander ES et al. (2001. 'Initial sequencing and analysis of the human genome' *Nature* 409: 860–921. Nature Publishing Group, Macmillan Publisher Ltd: Hampshire.

5 Venter JC et al. (2001 'The Sequence of the Human Genome' *Science* 291: 1304–1351. The American Association for the Advancement of Science.

Himla has done that. After completing her PhD on an early study into mitochondrial DNA variation in southern African peoples, she then did a post-doctoral fellowship in the United States working with Mark Stoneking, a leading researcher in mitochondrial DNA variation. And then, unlike so many of our graduates from Wits and UCT, she returned to South Africa where she has carried on – not just where she left off – but much further along the road of discovery; and she has taught all of us a great deal about population genetics and its relevance to the distribution of disease. She's a great teacher, as you will see. She's a caring mentor. She is committed to helping disadvantaged students, and gives an enormous amount of time to that difficult task. And, in addition to all that, she is an efficient organiser who is not afraid of hard work. She is playing an important role in furthering the aims of the AHGI. Himla Soodyall is an enthusiast; a great human being, a credit to our species.

I hope I've given you the message that you're in for a treat and that you're going to learn about the relevance of genetics, not strictly to health, although there is a relevance there, too, but to human origins and the evolution of our species, *Homo sapiens sapiens*. Himla is going to try, I think, to answer the important question: Where do we come from? If we know where we've come from, we may better understand who we are – this assemblage of different populations who are in the process of being blended into our rainbow nation. And if we know where we have come from, we might more clearly know where we are going.

A Walk in the Garden of Eden
Genetic Trails into our African Past

Humans have pondered their origins for as long as they have existed. This is reflected in the many myths and creation stories. We need only think about the Judeo-Christian Garden of Eden for example. Indeed, such stories seem to be a nearly universal feature of human cultures. I have borrowed the biblical meaning of the 'Garden of Eden' in my title to make reference to the geographic origins of modern humans in Africa.

We can reconstruct human history using a number of different methods. In the absence of written records, scholars have made use of information from disciplines as diverse as linguistics, archaeology, physical anthropology, cultural anthropology, history and paleo-anthropology to reconstruct their prehistory. The most direct account of our past is inferred from the fossil record. Skeletal

remains have been instrumental in establishing the evolution of human ancestors in Africa, and they have also provided important information about the evolution of modern *Homo sapiens*.

The genetic variation among living peoples offers another way of studying human evolution. Before proceeding to the discussion of how the genes are used to identify patterns of genetic similarity and difference, which in turn are used to reconstruct human history, let us understand a few concepts that we are familiar with concerning heritability. We all identify with the family unit – our siblings, parents, grandparents, great-grandparents and so on. We are quick to recognise certain physical traits like hair colour, nose shape, etc., as well as behavioural traits, like temperament, voice, and temper, that we consider to be inherited from one or other parent.

The concept of ancestry is deeply rooted in our different cultures. Sol Plaatje, who is being honoured by this lecture hosted by the Ministry of Education and the Africa Human Genome Initiative, was particularly proud of his Barolong ancestry, and took the time to reconstruct his genealogical history, believing that he was the first in his family 'to put memory to paper'.[6] He traced his paternal ancestry to King Morolong who is believed to have lived around the twelfth or thirteenth century. He also traced his maternal ancestry to Tau, the founder of the four royal branches of the Baralong. Sol Plaatje deduced from the genealogical data that his 'father and mother shared a common ancestry but 27 degrees apart'. Former president Nelson Mandela also acknowledged his ancestry in his book *Long walk to freedom*. He refers to his father Gadla Henry Mphakanyiswa, as a chief 'by both blood and custom' who belonged to the Thembu tribe.[7]

Paying respect to our ancestors is part of our cultural evolution. The thread that connects us biologically with our ancestry is stored in the human genome. The genome that carries the biochemical instructions that determine inherited traits contains an indelible record of our evolutionary past. Ridley[8] describes the human genome as a book in which there are 23 chapters, called

6 Willan B (1984) *Sol Plaatje: a biography*, p.4. Ravan, Johannesburg.

7 Mandela N (1994) *Long walk to freedom: The autobiography of Nelson Mandela*, pp.3–7. Little, Brown and Company. Boston, New York, Toronto, London.

8 Ridley M (1999) *Genome: The autobiography of a species in 23 chapters*. HarperCollins Publishers: New York.

chromosomes; each chapter contains several thousand stories, called *genes*; each story is made up of paragraphs, called *exons*, which are interrupted by advertisements called *introns*; each paragraph is made up of words, called *codons* and each word is written in letters called *bases*. Whereas English books are written in words of variable length using 26 letters, genomes are written entirely in three-letter words, using only four letters A, G, T and C (which stand for adenine, guanine, thymine and cytosine, respectively). Instead of being written on flat pages, the bases are written on long chains of sugar and phosphate and all these are chemically found together in a molecule referred to as deoxyribonucleic acid (DNA).

The genome is a very clever book, because under the right conditions it can both photocopy (*replicate*) itself and be read (*translated*). The total genetic complement of humans contains some three billion bases in different combinations controlling the development of the organism from conception to birth, to death, and producing the genetic variation that distinguishes one individual from everyone else. Humans have 46 chromosomes, half of which are inherited from our mothers and the other half from our fathers. Chromosomal DNA is found in the nucleus of the cell and is referred to as nuclear DNA (see figure 1). In addition to nuclear DNA, the mitochondria, the energy-producing organelles in the cytoplasm of all cells, also contain DNA that is referred to as mitochondrial DNA (mtDNA). MtDNA is inherited only from our mothers and only females can pass it on to their children (see figure 2). The Y chromosome is found in the nucleus and is transmitted exclusively from father to son (see figure 3).

The A, B and O blood groups, discovered in 1801, were first used to study genetic variation in humans. Biologists began to use the data to assess the affinities and origins of the various populations that make up humankind. The underlying principle of this approach was to reconstruct the history of mutations found in the DNA of contemporary individuals, and to trace their origins to a common ancestor who would have lived at some point in the past (see figure 4). Certain demographic events such as population migrations, a dramatic reduction in numbers in a population (a so-called bottleneck), and an increase in population numbers

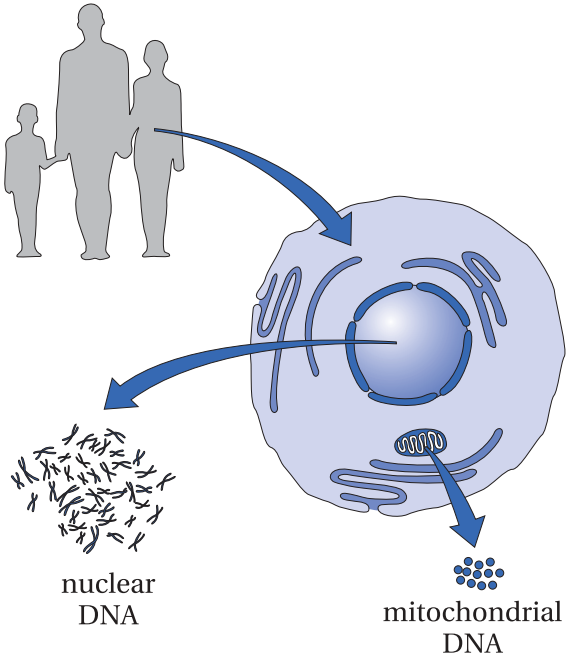

Figure 1. Schematic diagram of a cell showing the biparental inheritance of nuclear DNA and the maternal inheritance of mtDNA both found in the mitochondria of the cell.

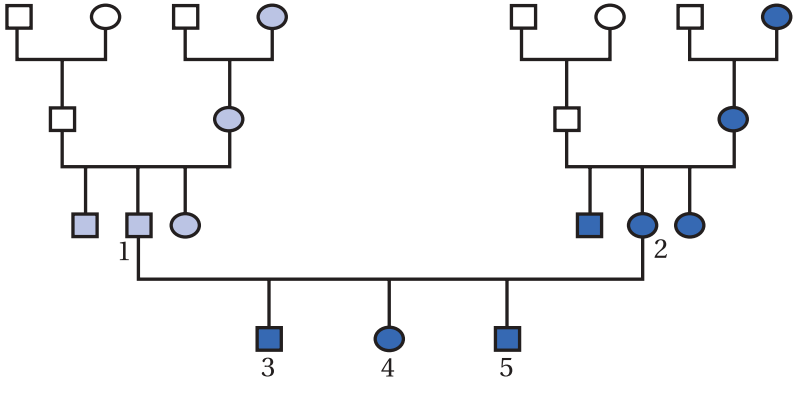

Figure 2. Maternal transmission of mtDNA inferred from family pedigrees. (Squares are males; circles are females). Individual 4 and her brothers (3 & 5) have inherited their mtDNA pattern (denoted in dark blue) from their mother (2), and not from their father (1) whose mtDNA pattern is shown in light blue. Only females pass on their mtDNA to both their sons and daughters, but only the daughters pass it on to their offspring.

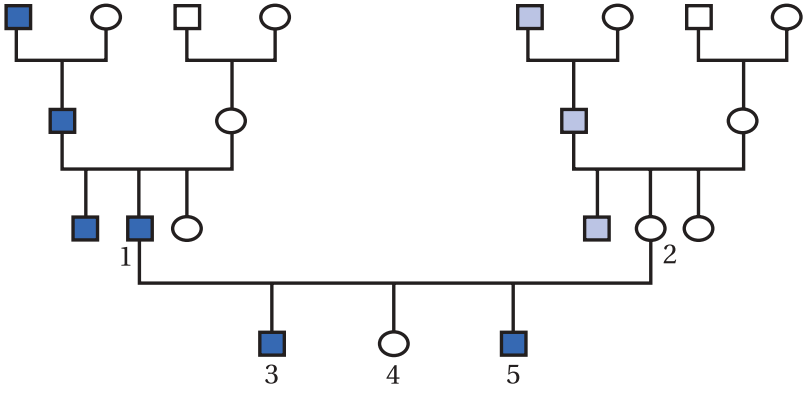

Figure 3. Paternal transmission of Y chromosome DNA. Using the same family pedigree as before, we note that only the male children (3 & 5) have inherited their father's (1) Y chromosome pattern (denoted in dark blue). On the mother's (2) side of the family, her brothers have inherited their father's Y chromosome pattern shown in light blue.

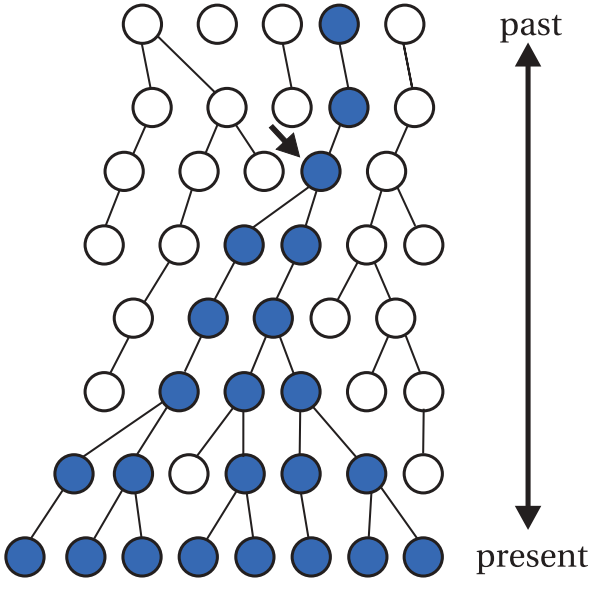

Figure 4. An illustration of the principle that all contemporary mtDNA types must trace back to a single ancestor. The filled circles indicate the path of descent from the ancestor (arrow) to the present generation; empty circles represent mtDNA types that went extinct. While the contemporary mtDNA types ultimately trace back to a single ancestor, note that other individuals co-existed with the mtDNA ancestor, and that the mtDNA ancestor had ancestors. (Adapted from Stoneking, 1993)[9].

9 Stoneking M (1993) 'DNA and recent human evolution' *Evolutionary Anthropology* 2: 60–73.

(population expansions), leave imprints, in the form of altered allele frequencies, on the collective human genome. Since these imprints are transmitted to succeeding generations, the genomes of living peoples are packaged with 'stories' depicting events in our evolutionary past. Thus, by studying human variation at the molecular or gene level, not only can we learn more about our evolutionary history, the focus of this talk, but we can also better understand the genetic contribution to health and disease.

Several lines of evidence have independently suggested that Africa is the birthplace of humankind. MtDNA is particularly useful when studying human evolution because of its unique pattern of inheritance. Unlike nuclear DNA, it is strictly maternally inherited and does not undergo recombination, a process of shuffling genes between paired chromosomes during meiosis, that is, when the ovum or sperm is being produced. Generally, differences in mtDNA are the direct result of mutations, and the 'history' of these mutational events can be reconstructed from contemporary divergent lineages. Also, mtDNA evolves about ten to 15 times faster than nuclear DNA, thus facilitating the discrimination between closely related populations. For these reasons, mtDNA has been exploited as a genetic marker to study the transmission of inherited traits passed on exclusively by females.

An earlier mtDNA study conducted by Rebecca Cann, Mark Stoneking and Alan Wilson[10] claimed that the mtDNA found in living peoples could be traced to a most recent common ancestor (MRCA) who lived in Africa approximately 200 000 years ago. When comparing mtDNA obtained from about 150 individuals throughout the world, these researchers observed that mtDNA from African populations were more diverse compared with mtDNA from non-African populations (see figure 5). This study advanced the 'Out of Africa' theory (also referred to as The Recent African Origin or Replacement Model) concerning modern human origins. This theory or model claims that there was only one geographic region where there was a complete evolutionary sequence from *Homo erectus* to modern humans, and that region

10 Cann RL, Stoneking M & Wilson AC (1987) 'Mitochondrial DNA and human evolution' *Nature* 325: pp. 31–36. Nature Publishing Group, Macmillan Publishers Ltd: Hampshire.

was Africa.[11] The multiregional theory, on the other hand, claims that over the last one to two million years, anatomically modern humans have evolved gradually from their archaic *Homo erectus* ancestors after these ancestors had left Africa and had spread to other parts of the Old World, and that there was gene admixture between archaic and modern humans.[12]

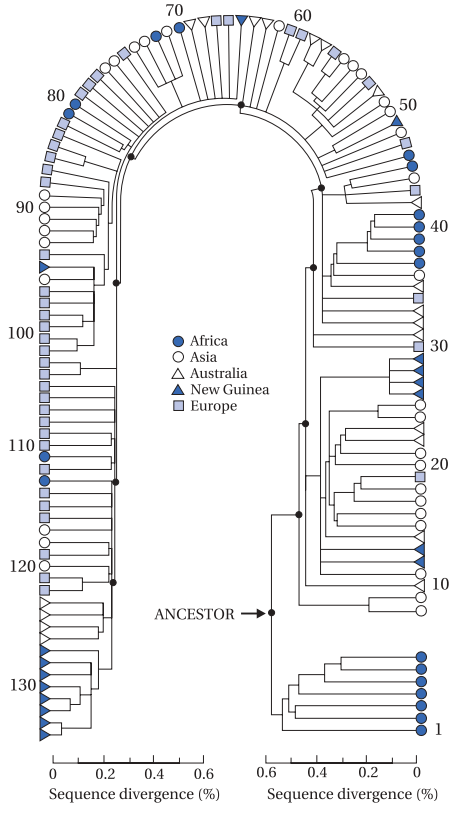

Figure 5. The ancestral mtDNA is inferred to have existed in Africa because of the split in the tree between the seven modern African mtDNA genomes placed below the ancestral sequence and all the other mtDNA sequences above it. Because the lower branch is purely African it is deduced that the ancestor was also African. The scale bars at the bottom indicate sequence divergence from which, using the mtDNA molecular clock, it is possible to assign dates to the branch points in the tree. The clock suggests that the ancestral sequence existed between 140 000 and 290 000 years ago.[13]

11 Stringer C (2001) 'The evolution of modern humans: where are we now?' General Anthropology. 7:1–5.

12 Wolpoff MH, Hawks J & Caspari R (2000) 'Multiregional, not multiple origins' *American Journal of Physical Anthropology*. 112: 129–36.

13 Brown TA (1999) *Genomes*, BIOS Scientific Publishers Ltd, New York.

A more convincing argument in support of the Out of Africa Theory was made when it became possible to extract DNA from a Neanderthal fossil specimen found in what is today Germany, dated to about 40 000 years ago, and to derive a mtDNA sequence from it.[14] Comparison of the Neanderthal mtDNA with over 3 000 mtDNA sequences from globally derived modern humans revealed that there were on average 28 differences in the segment of about 400 base pairs of mtDNA examined, compared with a maximum of only eight differences between any two humans living today. Moreover, when the Neanderthal mtDNA sequence was compared with chimpanzee and modern human mtDNA sequences on a neighbor-joining (NJ) tree, the Neanderthal sequence was placed at a position that was between chimpanzees and modern humans (see figure 6). These data suggested that mtDNA in modern humans and Neanderthals diverged from a common ancestral type over 650 000 years ago. More recently, two additional Neanderthal specimens, the Mezmaiskaya specimen from the northern Caucasus[15] and a specimen from the Vindija Cave in Croatia, confirmed these

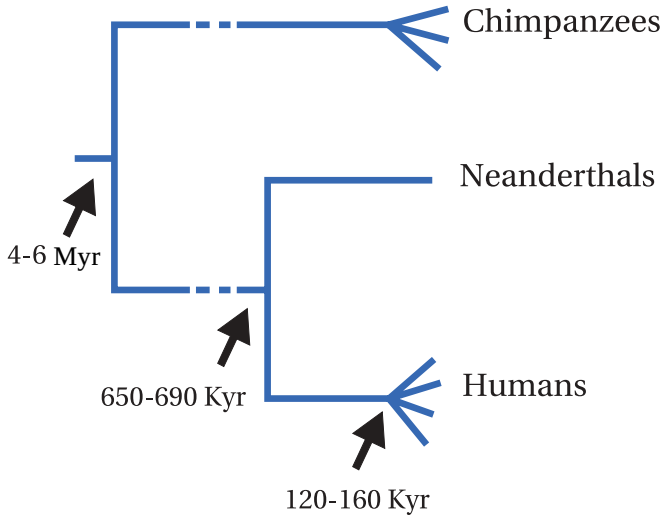

Figure 6. Schematic NJ-tree showing the evolutionary relationship of mtDNA in chimpanzees, Neanderthals and modern humans.

14 Krings M Stone A, Schmitz RW, Krainitzki H, Stoneking M & Pääbo S (1997) 'Neanderthal DNA sequences and the origin of modern humans' *Cell* 90: 19–30.

15 Ovchinnikov IV, Götherström A, Romanova P, Kharitonov VM, Lidén K & Goodwin, W (2000) 'Molecular analysis of Neanderthal DNA from the northern Caucasus' *Nature* 404: 490–493.

findings.[16] Interbreeding between Neanderthals and modern humans cannot be definitely excluded from these studies, but the findings to date suggest that Neanderthals did not contribute mtDNA to the contemporary gene pool.

One of the most significant findings to emerge from genetic studies is that non-African populations often show evidence of a severe reduction in diversity and population size, a 'bottleneck,' at some time in the past, followed by an expansion.[17] This bottleneck and expansion routine is presumed to have occurred when a branch of the early modern human population from Africa split off to form a small sub-population, which then expanded in size as it spread out to colonise Eurasia.[18] The distribution of mtDNA types among populations from different regions of the world is consistent with the Out of Africa theory, which is being increasingly accepted as the preferred theory of human origins. The mtDNA subhaplogroups (different mtDNA patterns) showing greatest antiquity are still retained in African populations (see figure 7). All other subhaplogroups found in non-African populations can be traced ultimately to subhaplogroup L3.

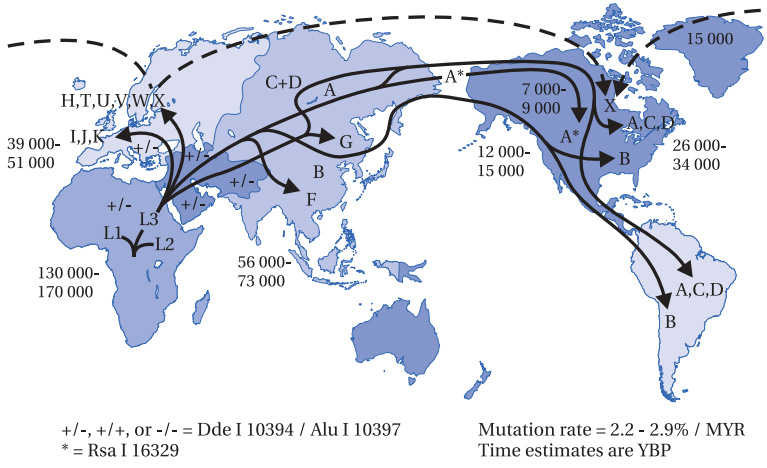

Figure 7. The global distribution of mtDNA types (denoted by different letters of the alphabet) found in contemporary populations. (Source: MITOMAP).[19]

16 Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grossschmidt K, Possnert G, Paunovic M & Pääbo S (2000) 'A view of Neandertal genetic diversity' *Nature Genetics* 26: 144–146.

17 Ingman M, Kaessmann H, Pääbo S & Gyllensten U (2000) 'Mitochondrial genome variation and the origins of modern humans' *Nature* 408: 708–713.

18 Brown SJ (2001) Genetic evidence (Present-Day DNA) http://www.neanderthal-modern.com/genetic3.htm.

19 MITOMAP: World Wide Web at http://www/gen.emory.edu/mitomap.html.

We have used mtDNA to examine the genetic affinities of populations in Africa. We find that the mtDNA pool of all populations is composed of L1, L2 and L3 lineages, albeit at different frequencies (see figure 8). Khoisan populations, for example, have a higher frequency of L1 lineages than other populations. More importantly, some of the oldest mtDNA lineages found in living peoples throughout the world are retained in some Khoisan populations. It is possible that other populations have lost these mtDNA lineages purely by chance or by drift effects including the bottleneck effect described above. These data strongly argue in favour of the origins of modern humans in southern Africa.

The Y chromosome is the paternally inherited equivalent to mtDNA. Most of the Y chromosome is non-recombining, and variation in its structure is brought about by mutation alone. Recent studies have identified a number of useful microsatellite markers,[20] as well as biallelic markers on the non-recombining region of the Y chromosome,[21] that have enhanced our understanding of Y chromosome variation. Using over 200 single nucleotide polymorphisms (SNPs), Underhill *et al.* (2001) have shown that the Y chromosome lineages found among contemporary humans could be assigned to ten (I–X) haplogroups (that is, groups of different Y chromosome haplotypes).

The deepest lineage in the human family tree (see figure 9) was found in African populations within haplogroup I, and was found in Khoisan populations from southern Africa. This lineage was also found in some Ethiopian and Sudanese populations, but at lower frequencies than in the Khoisan. Haplogroups I–III were found exclusively among African populations, with the remaining six haplogroups (IV–X) found at varying frequencies in Africans as well.

20 Seielstad M, Bekele E, Ibrahim M, Touré A & Touré M (1999) 'A view of modern human origins from Y chromosome microsatellite variation' *Genome Research*, 9: 558–567.

21 See: Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H & Zegura SL (2001) 'Hierarchical patterns of global human Y chromosome diversity' *Molecular Biology Evolution* 8: 1189–1203. See also: Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL & Oefner PJ (2000) 'Y chromosome sequence variation and the history of human populations' *Nature Genetics* 26: 358–361; and: Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) 'The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations' *Annals of Human Genetics* 65: 43–62.

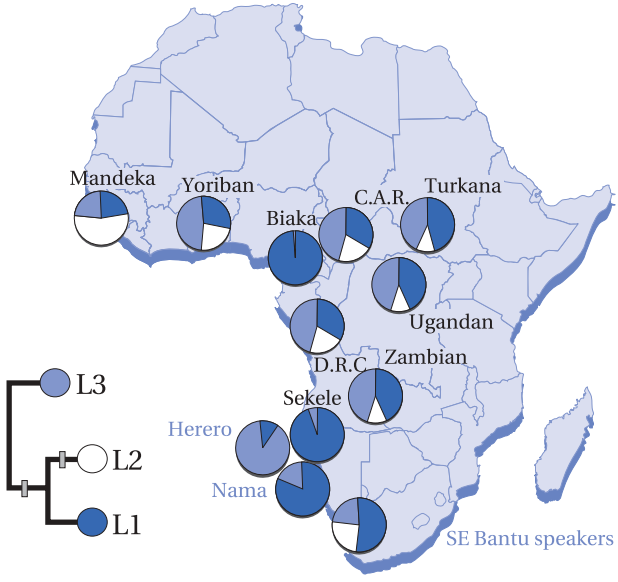

Figure 8. The distribution of the three common mtDNA subhaplogroups L1, L2 and L3 among different African populations. (CAR: Central African Republic; DRC: Democratic Republic of Congo) (Soodyall, unpublished).

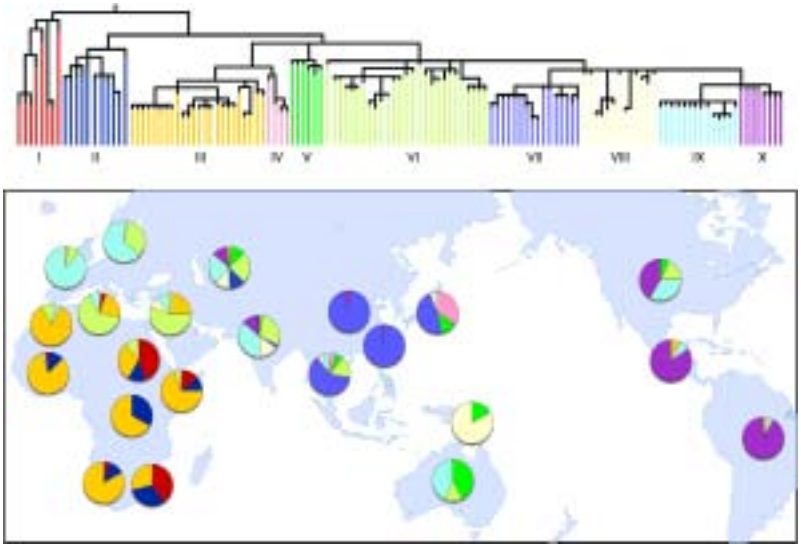

Figure 9. Global distribution of Y chromosome haplogroups. The frequencies and distribution of haplogroups I–X are indicated in the pie charts. Source: Underhill et al. 2001.[22]

22 Underhill PA. et al. (2001) 'The phylogeography of Y chromosome binary haplotypes'.

Thus, Y chromosome data are also consistent with the greater antiquity of Y chromosome lineages in Africa (80 000–150 000 years), and seem to confirm the Out of Africa theory of human origins.

We have used a combination of Y chromosome markers to assess the genetic affinities of African populations and to examine how males have contributed to shaping the gene pool of the continent. More than 70 per cent of Y chromosomes studied in sub-Saharan African populations were assigned to haplogroup III. Lane and colleagues[23] examined Y chromosome variation in seven South African Bantu-speaking groups and estimated the genetic variation among these groups to be insignificant (1.4 per cent). Another way of putting this is that the seven groups share roughly 98.6 per cent of the Y chromosome variation. These findings suggest that the groups studied are descended from a common ancestral population but have not been isolated from each other for long even though their languages have diverged sufficiently to become distinct from one another. It is estimated that is linguistic divergence has occurred over the past 2 000 years.

The history of the peoples of southern Africa can be reconstructed using a variety of methods, each having its own strengths and limitations. In trying to understand the complex patterns of genetic variation among the peoples of southern Africa, we have to use genetic data in conjunction with historical information gleaned from other disciplines. The written history of Africa is linked with the arrival of Europeans on the continent. Historical information, language, anthropological, and archaeological data confirm that the group of people often referred to collectively as the Khoisan are the aboriginal inhabitants of southern Africa. Southern Africa received three major immigrations in the last two millennia; the first from people speaking Bantu languages, perhaps in the last 2 000 years; the second from sea-borne European immigrants in the last 350 years; and the third from India and the Malay Archipelago in the past 100 to 150 years (see figure 10). There have been varying degrees of genetic admixture

23 Lane AB, Soodyall H, Arndt S , Ratshikhopha ME, Jonker E, Freeman C, Young L, Morar B & Toffie L (2002) 'Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y chromosome studies' A*merican Journal of Physical Anthropology* 119: 175–185.

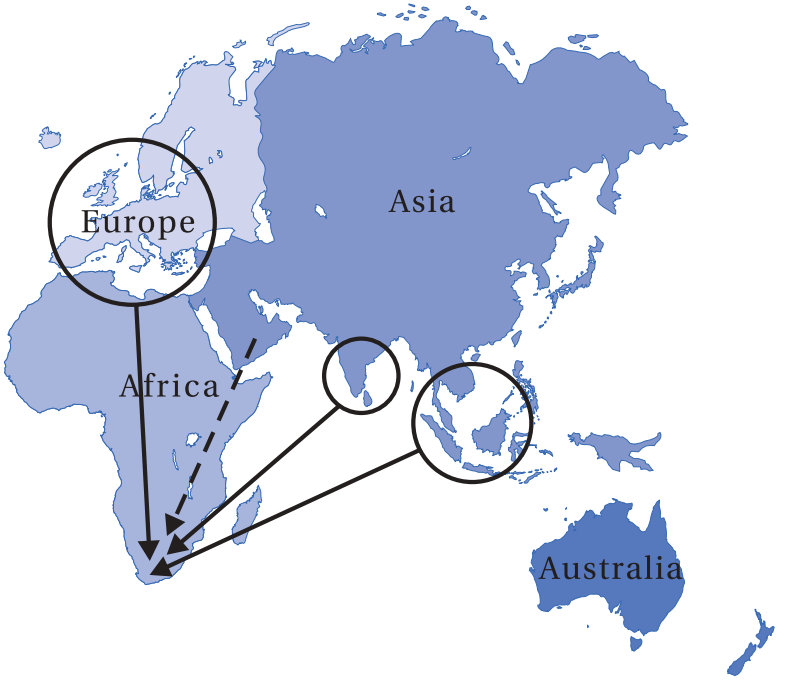

Figure 10. Recent contributions from outside of Africa to the gene pool of the South Africa.

between different southern African populations resulting in a complex pattern of diversity among the peoples of the region.

We can use Y chromosome data to examine how gene admixture and genetic trails have contributed to the Y chromosome composition of 'coloured' populations. The term 'coloured' has been used historically to refer to people of mixed ancestry in whom one parental contribution could be traced to European sources. However, various combinations of parental populations – European, Indian, Malay, Khoisan and Bantu-speaking Negroids – could have contributed to their gene pool.

We compared the Y chromosome lineages found in two groups of coloureds from the Cape (Cape coloured and Cape Malay) and one group of coloureds living in Johannesburg, to Khoisan (Nama, !Kung, Sekele and Kwengo), European (South African white and Ashkenazi Jews) and Bantu-speaking groups (pooled together and referred to as southeastern Bantu-speakers (SEB) from southern Africa (see figure 11). Using the global distribution of Y chromosome

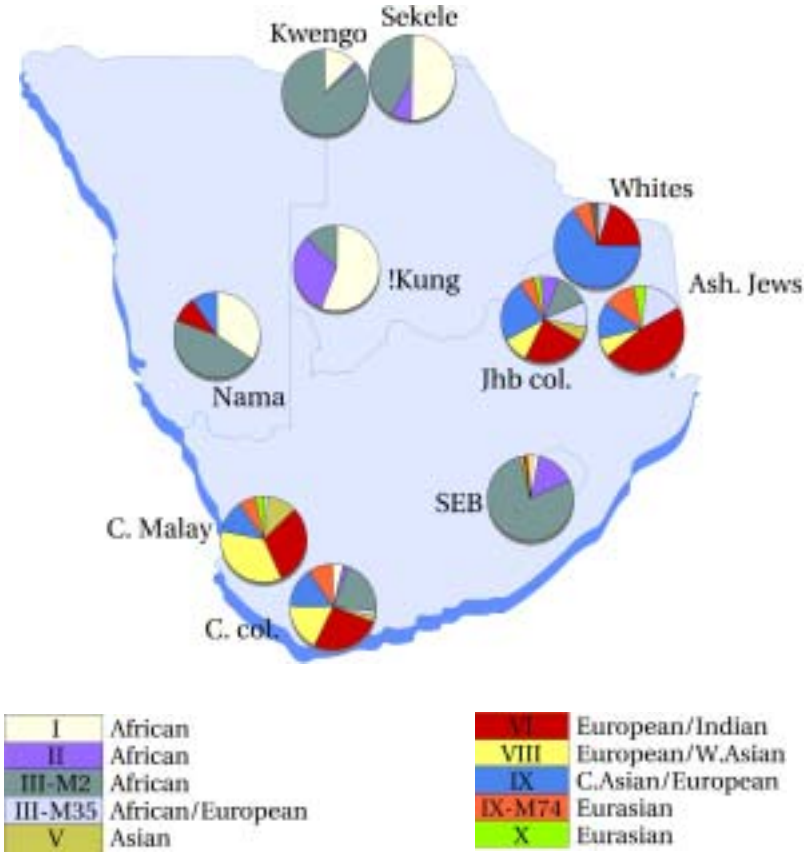

Figure 11. Map of southern Africa showing the proportion of Y chromosome haplogroups (I to X – refer to key) found in Khoisan Nama, !Kung, Sekele and Kwengo), European (South African whiles and Ashkenazi Jews), southeastern Bantu-speakers (SEB) and three coloured populations (Cape coloured, Cape Malay and Johannesburg coloured).

haplogroups (refer to figure 9), we estimated that the African contribution to the Y chromosome pool of the Cape coloured, Cape Malay and Johannesburg coloured populations to be 27, 2 and 19 per cent, respectively (see figure 11). The remaining Y chromosomes were derived from non-African sources – Europe and Asia. We could not unambiguously resolve these haplogroups to European or Asian origins given their widespread distribution in these regions. However, the inclusion of short tandem repeat markers that have a faster mutation rate would permit a better resolution of the Eurasian haplogroups. This work is currently in progress.

Genetic data should be considered as another 'tool' to study history. Unlike some of the other lines of evidence, the evidence in the genes is the direct result of our inheritance and is transmitted from one generation to the next in an unbiased way. The patterns of genetic variation found in living humans can be represented as branches on a tree; all branches (representing the various lineages found among all living peoples) are connected via the trunk with its roots deeply entrenched in Africa.

Genetic approaches to addressing questions of anthropological interest have made, and will continue to make, a significant contribution towards the generation of knowledge concerning the evolutionary history of our species. I am very optimistic that our research has a major role to play in contributing to the African Renaissance theme, and in restoring pride and a sense of identity to every South African.