INVESTIGATIONS OF ENVIRONMENTAL AND GENETIC INFLUENCES ON EAST AFRICAN DISTANCE RUNNING SUCCESS

Robert A. Scott BSc (Hons)

Submitted for the degree of Doctor of Philosophy (PhD) in the Faculty of Science, University of Glasgow.

March, 2006

Abstract

The primary objective of these experiments was to assess the environmental and genetic contributors to east African running success. Elite Ethiopian and Kenyan athletes, as well as matched controls, consented to participate in the current series of experiments by completing a questionnaire on their ethnic and environmental background, and by providing a buccal swab for analysis of novel and existing candidate genes for human performance.

The aim of the first two studies presented in this thesis (Chapters three and four) was to compare the demographic characteristics of elite Ethiopian and Kenyan athletes with those of their source populations. This was to assess the validity of reports linking their success to altitude inhabitation and having ran long distances to school each day during childhood. In both experiments, it was found that the elite athletes were clustered in regions of high altitude, proximal to the Rift Valley. However, in Ethiopia, non-endurance athletes were also found to cluster in these areas, so the link between altitude and elite athlete status may not simply be as a result of haematological adaptations conferring benefit in endurance performance. It was also found that elite athletes had travelled farther to school each day as children, and more of them had done so by running. In addition, the athletes displayed a distinct ethnicity relative to the general population. The associative nature of the experiments did not allow causality to be assigned to any one factor, but the results highlight the importance of environmental factors in the determination of elite east African athlete status.

The aim of Chapter five was to compare the frequency of mtDNA haplogroups between elite Ethiopian athletes with the general Ethiopian population. mtDNA is inherited in a matrilineal fashion, making it useful in population genetics. A number of studies have shown a maternal influence on the heritability of aerobic performance phenotypes, and more recent studies have suggested that mtDNA polymorphisms are associated with inter-individual variation in aerobic performance. Given the likely origin of modern humans in east Africa, Ethiopians display a wide degree of mtDNA haplogroup variation. Many of the haplogroups found in east Africa are rarely found elsewhere, while others are similar to those found commonly outwith east Africa. Given that mtDNA codes for components of mitochondria, mtDNA is a potential candidate gene for

physical performance. However, elite Ethiopian endurance athletes did not differ in their mtDNA haplogroup distribution from the control group. Elite Ethiopian athletes were not a genetically distinct group, as defined by their mtDNA. This does not support hypotheses that east African athletes arise from a limited genetic isolate.

Chapter six aimed to compare the mtDNA haplogroup frequency between elite Kenyan athletes and the Kenyan control group. Kenya also displays a wide degree of mtDNA variation, but displays a different haplogroup distribution from Ethiopia, with a higher frequency of typically African 'L' haplogroups (Ethiopia = 54 %, Kenya = 90 %). Kenyan athletes displayed a distinct distribution from controls. National standard athletes displayed an excess of M haplogroups relative to controls, and a tendency toward a lower frequency of L2 haplogroups, while international Kenyan athletes displayed an excess of L0 haplogroups, a lower frequency of L2 haplogroups and a tendency toward an excess of M haplogroups. Despite the associations of mtDNA haplogroups with elite Kenyan athlete status, the elite athletes retained a similar level of diversity to the general population, which attests to the complexity of the elite athlete phenotype. Findings of ethnic variation in mtDNA haplogroup frequency meant that influences of population stratification could not be ruled out.

The experiment in Chapter seven compared the frequency of Y chromosome haplogroups in elite Ethiopian athletes with that of the general Ethiopian population. Y chromosomal DNA is inherited solely down the paternal line and is confined solely to males. Y chromosome variants have been associated with a diverse range of phenotypes in health and disease, making the Y chromosome a viable candidate gene for physical performance in males. Similar to mtDNA, the non-recombinant pattern of inheritance makes the Y chromosome useful in population genetics to determine the ancestry of individuals or populations. Certain Y chromosome haplogroups were associated with elite Ethiopian athlete status. Athletes showed an excess of E3*, a lower frequency of E3b, and a tendency toward an excess of K(xP) haplogroups relative to controls. Although no influence of region or ethnicity was found on haplogroup distribution, effects of population stratification could not be ruled out. These results suggest that Y chromosome variation may influence elite Ethiopian athlete status.

Perhaps the most studied variant of the candidate genes for human performance is the ACE I/D polymorphism. A number of studies have associated the ACE I allele with endurance performance, while the D allele has been associated with power type performance. The experiment in Chapter eight aimed to compare ACE I/D genotype frequencies between elite Ethiopian and Kenyan athletes and their source populations. It was found that male Ethiopian marathon runners showed an excess of II genotypes relative to controls. However, this association was not replicated in 5-10 km Ethiopian runners, or in Kenyan distance runners. It is unlikely that the I/D polymorphism is functional in influencing variation in the ACE phenotype. Other variants in the ACE gene are more strongly associated with serum ACE levels in African populations, where levels of linkage disequilibrium differ from Caucasian populations. For this reason, A22982G was genotyped in elite athletes and controls. This variant showed a stronger association with serum ACE levels than the I/D polymorphism in the Kenyan control group. However, there was no association between A22982G genotype and elite athlete status in either Ethiopian or Kenyan athletes. These results do not support a role for ACE gene variation in the determination of elite East African athlete status.

Table of Contents

List of Tables

List of Figures

Acknowledgements

The first thanks are to Dr Yannis Pitsiladis. I would not have dreamed four years ago when I first started work on the fascinating topic of why east African runners dominate distance running that I would still be doing so today. I feel privileged to be investigating such interesting questions and to call it "work" to do so. Dr Pitsiladis has delivered on all promises to make this possible, and in making this a rewarding, and, despite some stressful times, extremely enjoyable period in my life. He has ensured that I had nothing but the best provision during my PhD, and ensured that I gained much experience, presenting at a number of national and international meetings, and undertaking fieldwork in Africa. These experiences have been extremely valuable to me, both from a working, and a personal perspective. I will be forever grateful to Dr Pitsiladis for being such an enthusiastic and committed supervisor.

My thanks to Dr Niall MacFarlane for assisting me to begin this PhD, and to Professor Ian McGrath for his intervention and allowing me to continue my studies.

My thanks also go to Dr Richard Wilson for excellent advice, interesting discussions and much-needed tutorials on all aspects of genetics. Similarly, Dr Will Goodwin has always provided excellent advice and remained patient when I began my training in molecular biology. Both have provided excellent guidance and friendship throughout the course of my PhD.

Thanks to Dr Colin Moran, who has also provided ecxellent training in molecular biology, and has always made time to listen when I have encountered problems.

Thanks to everyone in the lab: Chris, Barry, John, Steve, Andy, Rona, Heather, Fran, Nick and Vaso for making it a great place to work, and for helping create some legendary NABS Christmas parties! Fewer thanks go to Chris and John for their tips on football scores!

Dr Tony Turner also deserves thanks for understanding when I was doing too many things at once, and allowing me the time to complete this thesis when I was working at Edinburgh University. The opportunity to work with elite athletes in such an excellent setting has been extremely valuable to me.

Asante sana to Vincent Onywera for his tireless work in Kenya in organising the unbelievable feat of collecting hundreds of DNA samples from athletes all over Kenya. Also for his excellent friendship, tutelage in how to make the perfect ugali, and for teaching me the finer points of Swahili (Nataka….).

Thanks to Mum and wee brother John for their support, belief and encouragement throughout the course of my PhD, and for putting up with my increasingly infrequent visits in the latter stages of my studies. John, I look forward to resuming our football matches. I expect that you will soon be winning in the back garden as you invariably do on the Playstation.

For everything, my special thanks to Laura. Your unwavering support in all areas of my life means more than you can know.

Final, sincerest thanks to the subjects of the studies who made it all possible by participating enthusiastically. I am extremely grateful for your participation.

Author's Declaration

I hereby declare that this thesis has been composed by myself, and that the work of which it is a record has been done by myself, except where specifically acknowledged. I also confirm that it has not been submitted in any previous application for a higher degree and that all sources of information have been specifically acknowledged by means of references.

Some of the results contained in this thesis have been published in peer-reviewed journals as follows:

i. **Scott, R. A.**, Georgiades, E., Wilson, R. H., Goodwin, W. H., Wolde, B., & Pitsiladis, Y. P. (2003). Demographic characteristics of elite Ethiopian endurance runners. *Med.Sci.Sports Exerc.* **35**, 1727-1732.

ii. Moran, C. N., **Scott, R. A.**, Adams, S. M., Warrington, S. J., Jobling, M. A., Wilson, R. H., Goodwin, W. H., Georgiades, E., Wolde, B., & Pitsiladis, Y. P. (2004). Y chromosome haplogroups of elite Ethiopian endurance runners. *Hum.Genet.* **115**, 492-497.

iii. **Scott, R. A.**, Moran, C., Wilson, R. H., Goodwin, W. H., & Pitsiladis, Y. P. Genetic influences on east African running success. Equine and Comparative Exercise Physiology 1[4], 273-280. 2004.

iv. **Scott, R. A.**, Moran, C., Wilson, R. H., Onywera, V., Boit, M. K., Goodwin, W. H., Gohlke, P., Payne, J., Montgomery, H., & Pitsiladis, Y. P. (2005). No association between Angiotensin Converting Enzyme (ACE) gene variation and endurance athlete status in Kenyans. *Comp Biochem.Physiol A Mol.Integr.Physiol* **141**, 169-175.

v. **Scott, R. A.**, Wilson, R. H., Goodwin, W. H., Moran, C. N., Georgiades, E., Wolde, B., & Pitsiladis, Y. P. (2005). Mitochondrial DNA lineages of elite Ethiopian athletes. *Comp Biochem.Physiol B Biochem.Mol.Biol.* **140**, 497-503.

vi. Onywera, V. O., **Scott, R. A.**, Boit, M. K., & Pitsiladis, Y. P. (2006). Demographic characteristics of elite Kenyan endurance runners. *J.Sports Sci.* **24**, 415-422.

List of Abbreviations

ACE	Angiotensin Converting Enzyme
ANOVA	Analysis of variance
bp	Base pair(s)
°C	Degrees Celsius
CI	Confidence interval(s)
COX	cytochrome c oxidase
CRS	Cambridge reference sequence
d.f.	Degrees of freedom
DNA	Deoxyribonucleic acid
EM	Expectation maximisation
HWE	Hardy-Weinberg equilibrium
h	Hour(s)
HVS-1	Hyper-variable sequence one
Kb	Kilobases
kg	Kilogram
µl	Microlitres
ml	millilitre(s)
$ml.kg^{-1}\ min^{-1}$	Millilitres per kilogram per minute

mM	Millimolar
min	Minute(s)
mtDNA	Mitochondrial DNA
mRNA	Messenger ribonucleic acid
NADH	Nicotinamide adenine dinucleotide (reduced form)
ng	Nanogram(s)
np	Nucleotide position
OR	Odds ratio(s)
OXPHOS	Oxidative phosphorylation
P	Probability
PCR	Polymerase chain reaction
pmol	Picomole(s)
RFLP	Restriction fragment length polymorphism
s	Second(s)
SD	Standard deviation
SNP	Single nucleotide polymorphism
UV	Ultraviolet
$\dot{V}o_{2\,max}$	Maximal oxygen uptake
χ^2	Chi-square

Chapter 1

General Introduction

1.1 East African Running

In sporting competition, there are many examples of individuals or populations who dominate their particular event. The vast majority of elite athletes will retire without an Olympic Gold medal yet others win many during illustrious careers. Likewise, some nations are continually successful in certain events and win many medals, at the expense of the others. Some examples, such as that of the USA and their domination of Basketball, are readily explained by participation rates, coaching, and training (Phillips, 1976). Others, however, are not so readily explained. Such domination is not isolated to team sports; certain nations enjoy continual success in individual events, continually producing athletes who specialise in specific athletic events. Scandinavian runners have, in the past, enjoyed periods of success, producing excellent distance runners such as Paavo Nurmi, and later, Lasse Viren. Britain has also had periods of domination, particularly in middle distance running. After Sir Roger Bannister completed the first sub 4-minute mile, Chris Chataway and Brian Hewson, two other British athletes, closely followed him. However, since their first Olympic medal in 1960, with the barefoot marathon win of Abebe Bikila, it can be argued that no group has ever dominated a branch of athletics to the extent that east African athletes now dominate distance-running events. Numerous, but not even close to exhaustive, examples of outstanding east African runners include Abebe Bikila, Kip Keino, Amos Biwott, Nyandika Maiyoro, Miruts Yifter, Fatuma Roba, Haile Gebresellasie, Derartu Tulu, Paul Tergat, and Kenenisa Bekele. As can be seen from the above list, Ethiopia and Kenya have enjoyed particular success at the highest level of competition: the Olympic Games. Table 1.1 shows the success that Ethiopia and Kenya have had in recent Olympic competition. As can be seen, these two countries have won half of all Olympic medals from 3,000m steeplechase to the Marathon since 1992. The domination of these countries is not isolated to Olympic competition. At the last four IAAF World championships, Ethiopia and Kenya have won 48 of the 84 available medals from 3000m to the marathon. At international cross-country and road-race events, again these two countries dominate. Not only do they win awards and medals, east African athletes produce world-beating times in the process, which athletes from other countries find difficult to match. In addition to dominating Olympic middle- and long-distance track events, east Africans dominate the world road racing and cross country circuit. Figure 1.1 shows the percentage of the top 50 all time performances at distances ranging from 100m to marathon by Ethiopian and Kenyan athletes relative to the rest of the World. It can be seen that they contribute over 70 % of the all-time top 50 performances in the 10,000m. This figure shows only male athletes, and although female east African athletes have not,

historically, shown the same level of success as their male counterparts, they are now becoming more successful. Table 1.1 shows that the disparity between male and female athletes is declining, with female athletes winning five out of nine possible medals at the Olympic Games in Athens, 2004. Furthermore, at the Helsinki World Championships (2005), Ethiopia won every female medal at the 5,000m and 10,000m and Kenya took silver in the Marathon.

The phenomenal success of east African running can be ascribed not only to Ethiopia and Kenya, but can be further localised within these countries to regions and tribes producing a disproportionately high number of successful runners. This is particularly the case in Kenya, where analysis of the geographical distribution of their successful athletes has shown that the Kalenjin tribe produce a disproportionate number of successful runners (Manners, 1997). The Kalenjin consist of a number of tribes, amongst whom are the Keiyo, Tugen and Nandi: a tribe known, even within the Kalenjin, to produce successful runners. The data on the relative success of the Nandi is astounding. In 1988, this tribe, accounting for only 1.8 % of the Kenyan population, produced over 40 per cent of Kenya's most successful runners (Bale & Sang, 1996). Despite the locality of east African success, due to the concomitant success of sprint athletes of west-African ancestry and the domination of some US sports by African-Americans, east African success has been combined with the above in the misguided inference of genetically advantaged Black athletes in the popular media (Entine, 2001b). As can be seen from the track world records shown in Table 1.2, this argument may seem compelling on first examination of the data and the stereotype of the natural black athlete is now one that is widely believed. However, such stereotypes are not without their effects, not only in augmenting disparities in sport, but also in society. Hoberman (1997) gives a detailed discussion of the origin of the stereotype and the effects on society. It is argued that such a construct has a detrimental effect on race-relations, and of course, history has shown that ideas of racial genetic superiority are dangerous.

Table 1.1 Olympic medal wins of Kenya and Ethiopia from 1992 to 2004. Events from 3000m Steeplechase to Marathon. Ethiopian flag denotes an Ethiopian medal win and Kenyan flag denotes a Kenyan medal win.

Olympic Year	Medal	3000m S/C (M)	5000m (M)	5000m (F)	10000m (M)	10000m (F)	Marathon (M)	Marathon (F)
2004	Gold							
	Silver							
	Bronze							
2000	Gold							
	Silver							
	Bronze							
1996	Gold							
	Silver							
	Bronze							
1992	Gold							
	Silver							
	Bronze							

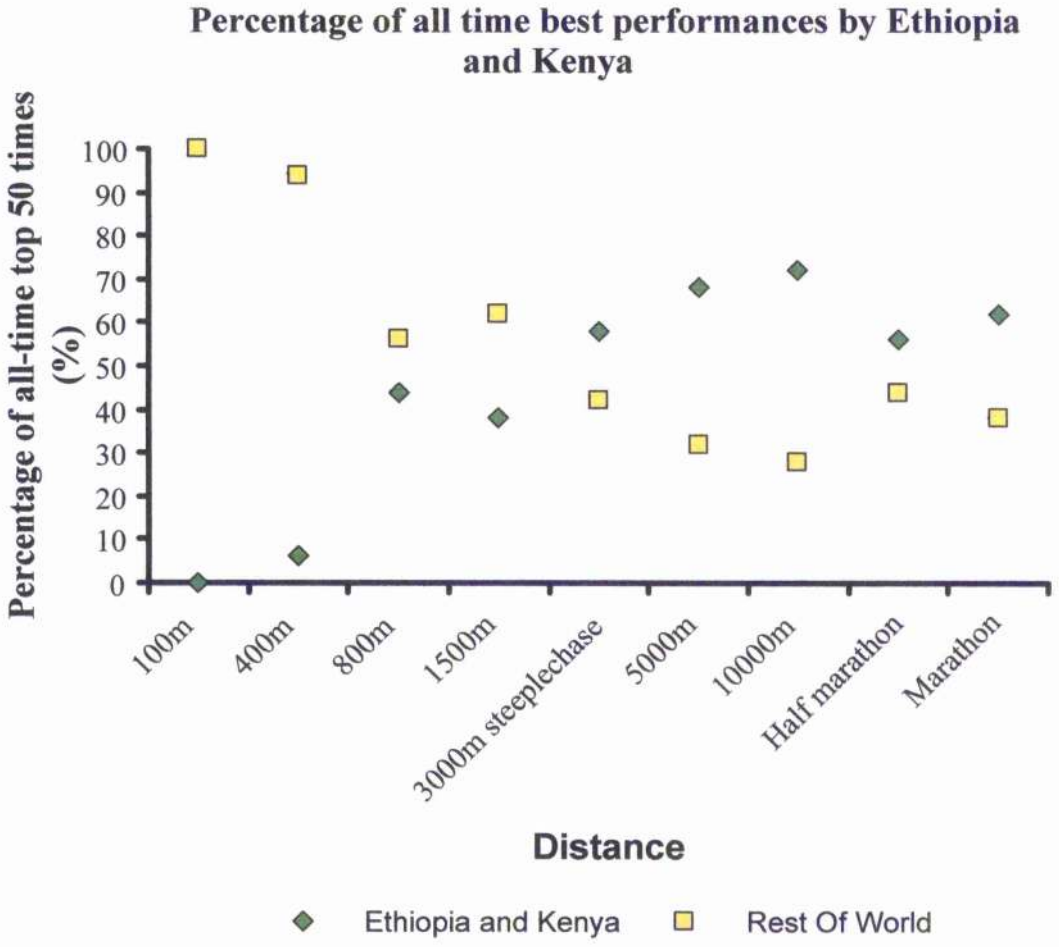


Figure 1.1 Percentage of all time top 50 times by Ethiopia and Kenya compared to the Rest of the World. Data taken from IAAF outdoor all time male times. Data correct 08/2005.

Table 1.2 Male World Records from 100m to marathon. Ancestral origin is derived from geographical and ethnic status

Distance	Athlete	Time	Ancestral Origin
100m	Asafa Powell (JAM)	9.77s	West Africa
110m Hurdles	Colin Jackson (GBR)	12.91s	West Africa
200m	Michael Johnson (USA)	19.32s	West Africa
400m	Michael Johnson (USA)	43.18s	West Africa
400m Hurdles	Kevin Young (USA)	46.78s	West Africa
800m	Wilson Kipketer (KEN)	1:41.11	East Africa
1000m	Noah Ngeny (KEN)	2:11.96	East Africa
1500m	Hicham El Guerrouj (MOR)	3:26.00	North Africa
Mile	Hicham El Guerrouj (MOR)	3:43.13	North Africa
3,000m	Daniel Komen(KEN)	7:20.67	East Africa
5,000m	Kenenisa Bekele (ETH)	12:37.35	East Africa
10,000m	Kenenisa Bekele (ETH)	26:17.53	East Africa
Marathon	Paul Tergat (KEN)	2:04:55	North Africa

1.2 Explanations for east African domination

The domination of east African runners has generated a great deal of interest in the reasons for their success. A number of explanations have been given to account for their success: socio-cultural, psychological, physiological, environmental and genetic. Each of these areas is addressed in sequence, although it is recognised that no one area alone can explain the phenomenon.

1.2.1 Psychological and socio-cultural explanations

Although optimal physiological preparation is paramount in athletic success, the success of east African runners has also been attributed to a number of socio-cultural and psychological factors. When stereotypes of black athletic superiority pervade to the level of fellow athletes, they ultimately give African athletes a psychological advantage, which will serve to perpetuate their current domination (Hamilton, 2000). This concept, known as "stereotype threat" has been proposed to account for some of the inter-population differences in sporting performance (Baker & Horton, 2003). Some studies have also suggested that success in a task is often based more on the perception of ability rather than actual ability. For example, Stone et al. (1999) nicely illustrated this concept. Black and white students were presented with a standardised golf task with the outcome varying under which category the test was presented: "sport psychology", "natural athletic ability", or "sport intelligence". Both groups of students performed equally well under the "sport psychology" heading. However, the black athletes outperformed the white students when

the task was presented as "natural athletic ability"; and the opposite was true when it was presented as a test of "sport intelligence". The expectation that east African athletes will win most major competitions may well be a self-fulfilling prophecy when it is held strong by athletes competing to be the first non-African runner rather than the gold medal winner.

It has also been suggested that there are certain cultural practices in Kenya, such as male circumcision, which confer upon the Kenyan child an emotional toughness and pain tolerance which proves useful in the rigours of distance running (Manners, 1997). Manners describes a procedure which tests the character traits of courage, endurance and determination, and contends that if a boy can stand up to the physical and emotional pressure during adolescence that the pain of a tough race is comparatively minor. An investigation into the cultural influences acting upon Kenyan running is given by Bale and Sang (1996), who examine the global forces acting on African culture to create the phenomenon of Kenyan running. They argue that it is not simply a case of the "natural African athlete" relying on "raw talent" but that such athletes are the result of cultural pressures. It is argued that to simplify the success of Kenyan athletes to their struggle against hardship is patronising and disguises the fact that the struggle against adversity is far from unique to Kenya, or any ethnic group within Kenya. It is also suggested that the channelling of athletes presumed to have athletic potential based upon existing stereotypes is a more likely explanation for their success. These cultural influences included, but were not confined to, the influence of colonialisation in directing African energies toward organised sporting competition as a means of aiding the administration of these countries. However, although this may not have been the primary goal, the ideology of sporting competition instilled during this period has undoubtedly contributed to east African running success (Bale & Sang, 1996).

1.2.2 Physiological explanations

In attempt to understand the success of east African distance runners, or more generally, black African distance runners, a number of studies have compared physiological characteristics important to distance running between groups of black and white athletes. Studies have focused on comparing characteristics such as $\dot{V}o_{2\,max}$, lactate accumulation and running economy, between groups of black and white athletes in attempt to understand the success of African runners. In South African athletes, it has been found that black athletes have lower levels of lactate accumulation than white athletes for given exercise intensities (Bosch *et al.*, 1990;Coetzer *et al.*, 1993;Weston *et al.*, 1999). It was also shown

that black athletes had better running economy (Weston *et al.*, 2000), and had higher fractional utilisation of $\dot{V}o_{2\,max}$ at race pace (Bosch *et al.*, 1990;Coetzer *et al.*, 1993;Weston *et al.*, 2000). It has been suggested that if the physiological characteristics of sub-elite South African runners are present in elite African runners, this may help to explain the success of this racial group in distance running (Weston *et al.*, 2000). However, this assertion is difficult to reconcile with earlier studies concluding that their findings were compatible with the notion that "Black male individuals are well endowed to perform in sport events of short duration" (Ama *et al.*, 1986). This conclusion by Ama et al. was based on a study comparing skeletal muscle characteristics between 23 sedentary Black males from five countries of West and Central Africa and 23 Canadian Caucasian males. The finding that the sedentary black males had lower levels of type I muscle fibres than white subjects (Ama *et al.*, 1986), was not replicated in a further study comparing fibre type in untrained US black and white subjects (Duey *et al.*, 1997), where it was concluded that racial differences in fibre type are small relative to the variability of the measure. A study into the anaerobic performance of sedentary west and central African subjects compared to sedentary Canadian Caucasian subjects showed that although both groups displayed similar peak force, the Black subjects had a larger decrement in force during the last 60s of a 90s anaerobic test (Ama *et al.*, 1990). Again, the suggestion that black subjects are less resistant to fatigue (Ama *et al.*, 1990) is in contrast to the studies of black South African runners where laboratory measures of fatigue resistance, such as treadmill tests to exhaustion (Weston *et al.*, 1999) and repeated isometric contractions (Coetzer *et al.*, 1993), showed that the black subjects had greater fatigue resistance. The contrasting findings of the above studies highlight the inability to extrapolate results from one ethnic group to another.

Saltin et al. (1995a;1995b) undertook the first studies of elite Kenyan athletes and factors which influenced their performance. They compared physiological characteristics, such as those in the above studies of South African runners, between Scandinavian athletes and elite Kenyan athletes, and found that the Kenyan athletes had lower levels of lactate accumulation and markedly lower ammonia accumulation during graded exercise. It was also found that the Kenyan runners had better running economy but that they had $\dot{V}o_{2\,max}$ levels comparable to those of the Scandinavian runners. Another study by Saltin et al. (1995a) comparing skeletal muscle characteristics found that elite Kenyan runners (senior and junior) did not differ in fibre type distribution from their Scandinavian counterparts, although the senior Kenyan runners had a (statistically non-significant) tendency toward higher muscle capillarity than their Scandinavian counterparts.

It is clear that good physiological studies are required to understand the factors influencing elite performance. However, the approach of comparing physiological characteristics between groups defined solely by skin colour does not offer much insight into why some populations are more successful than others. The inadequate classification of subjects into groups based on superficial characteristics, such as skin colour, will undoubtedly lead to equivocal results without identifying any underlying genetic mechanisms and will serve only to augment existing stereotypes of genetically advantaged black athletes. As shall be discussed later, the classification "black" or "white" does not necessarily correlate with a homogenous genetic background.

1.2.3 Demographic explanations

The studies listed above offer insight into the determinants of elite performance and differences between groups of African and Caucasian runners, but offer little insight into the causes of African running success. As mentioned there is a belief in some sections of the popular media, and among the general public, that east African runners or black athletes in general have some form of genetic advantage (Entine, 2001b). There are also suggestions in the scientific press that genetic endowment may have a role to play in African running success (Saltin, 1996;Larsen, 2003). At present, however, there is no evidence to suggest that east African running success is a genetically mediated phenomenon. The quoted studies of elite Kenyan athletes by Saltin et al. (1995a;1995b) concluded that factors such as childhood physical activity and hard training are the major contributors to their success. It was also concluded that geographical and ethnic disparities in athlete production within Kenya were more likely to be attributable to cultural rather than genetic differences (Saltin *et al.*, 1995b). Therein lie other commonly proposed explanations for the success of east African athletes: that their altitudinous habitat and the distances that they cover to and from school each day as children are responsible for their endurance performance. A key finding by Saltin et al. (1995b) was that Kenyan boys who used running as a form of transport since childhood, but did not undertake any structured training, had $\dot{V}o_{2\,max}$ levels 30 % higher than sedentary Kenyan boys. This finding highlights the importance of childhood physical activity in developing aerobic capacity in east Africans, although the sample size was admittedly small: 6 sedentary boys and 4 active boys. It seems, therefore, that the anecdotal explanation that the distances travelled to school by east African children favours their performance may have some grounding. The conclusion by Saltin et al.(1995b) was that the success of Kenyan runners, and the Nandi tribe in particular, was probably due to cultural factors, such as the success of

previous Nandi people and by the culture of organised training and competition instilled in Nandi schools.

Another explanation often given for the success of east African runners is that the altitude at which they reside and train is influential in their performance. It is true that many areas of Ethiopia and Kenya are altitudinous, both countries being dissected by the Rift Valley. Altitude training is undertaken by many athletes in the hope that the adaptations to altitude result in haematological adaptations resulting in improved performance. A number of trends exist in altitude training, and there is debate over which is the most effective in improving sea level running performance (Wilber, 2001). Some propose that the "live high, train low" method (Levine & Stray-Gundersen, 1997) is most effective by allowing the positive adaptations associated with altitude, such as the increase in red cell mass, without incurring the deficit in training intensities associated with altitude. The method involves sleeping at altitude (or simulated altitude) while completing training at lower (<1000m) altitude to allow maintenance of sea level training intensities. This method has been shown to improve sea level running performance in collegiate level (Levine & Stray-Gundersen, 1997), and elite level (Stray-Gundersen *et al.*, 2001) runners. Others infer a beneficial effect of the opposite concept: "training high, living low", from results showing improvements in oxygen transport capacity in skeletal muscle following such a protocol (Hoppeler & Vogt, 2001), although there is little evidence to support an actual improvement in endurance performance resulting from such a regime (Wilber, 2001). However, it is difficult to extrapolate any research findings to account for the success of east African athletes. Kalenjin country in Kenya has many regions of altitude over 2000m and regions of Ethiopia close to the Rift Valley have many areas over this elevation. What differs in the east African situation is that the athletes have been born and raised at such altitudes and complete most of their training at altitude where the climate is cooler. It is unclear how this may influence their performance, and if their status as indigenous high altitude natives affords them any benefit that allows them to overcome the training deficits normally associated with altitude. A study by Schmidt et al. (2002) proposed that the chronic altitude exposure and endurance training of natives to moderate altitude combine synergistically to enhance haemoglobin mass and blood volume to influence their performance. The comparison between cyclists from sea level and altitude found that their $\dot{V}o_{2\,max}$ levels were similar when measured at their respective altitudes. It followed that the $\dot{V}o_{2\,max}$ of the altitude cyclists would increase at sea level and offer an advantage (Schmidt *et al.*, 2002), although evidence to this effect is equivocal (Favier *et al.*, 1995;Saltin *et al.*, 1995b) . The fact that geographically isolated populations indigenous to high altitude, such

as east Africans and Tibetans, appear to have differing strategies for maintaining oxygen delivery (Beall, 2003), may suggest that the adaptations of east Africans have some influence on their endurance success. Although no direct evidence is available, there are indications that genetic variation may influence the response to high altitude (Mortimer *et al.*, 2004). The possibility exists, therefore, that environmental pressure in the form of hypobaric hypoxia has caused selection for genetic variants conferring advantages in oxygen transport. Although there are a number of high altitude populations, it may be that the pattern of adaptation in Ethiopians may have the potential to concurrently influence the endurance phenotype. It is intriguing, therefore, that there may have been selection for genetic variants that influence oxygen transport in high altitude Ethiopians, which confers a benefit in exercise performance. It is recognised that the populations from which the athletes arise are at moderate rather than high altitude, but there is a possibility of gene flow between these geographically proximal populations to confer similar genetic variants on the populations from which the athletes arise.

1.2.4 Genetic explanations

Although it is an intriguing possibility that there is a genetic component to the success of east African runners, and there is certainly a perception of one in the popular media (Entine, 2001a;Entine, 2001b), there is currently no evidence to that effect. Findings of physiological differences between groups of black and white subjects (Ama *et al.*, 1986), have been suggested to display evidence for a genetic predisposition to higher levels of type II muscle fibres in black Africans (Saltin *et al.*, 1995a), despite the failure to replicate the findings on which this assertion is based (Duey *et al.*, 1997). Saltin et al. (1995a) found that the muscle fibre type distributions were similar in Kenyan senior runners, Kenyan juniors and Scandinavian runners. However, it was noticed that the Kenyan senior runners had a tendency toward a higher level of capillarity than the Scandinavian runners (Saltin *et al.*, 1995a). A finding that may be overlooked, however, is that the Kenyan junior runners had a lower capillarity than both the Scandinavian and Kenyan seniors (Saltin *et al.*, 1995a), suggesting that intense training, and not genetics, was responsible for the high capillarity of the Kenyan senior runners. As mentioned, the domination of athletes of west African origin in sprint-based events sustains the myth of the genetically superior black athlete. The clustering of successful athletes in certain areas of Kenya further adds to the perception of a genetically mediated phenomenon, with the suggestion that there may have been selection in the Nandi tribe for those exhibiting good endurance performance (Manners, 1997). It has been suggested that east Africans have a, genetically endowed

body type, advantageous for distance running (Entine, 2001a), although evidence to this effect does not exist. Suggestions also exist in the scientific literature that certain ethnic groups in Kenya have the "proper genes" for elite athletic performance (Larsen, 2003). It is clear, therefore, that this is a pervasive idea, which may have some physiological grounding, but requires further investigation.

1.3 Genetics of elite performance

1.3.1 Unmeasured genotype approaches

Despite the paucity of evidence that the success of east African runners is genetically mediated, the belief that genetic variation influences the inter-individual variation in physical performance is growing. The capacity to become an elite athlete is not simply a result of physiological conditioning but a rare, yet advantageous combination of factors: genetic, physiological, biochemical and environmental (Myburgh, 2003). The question of "nature" Vs. "nurture" is often at the fore in discussions of the determinants of elite performance. Although no evidence exists for a genetic component in the inter-population differences in performance, there is a growing body of evidence to suggest that genetic variation does influence the inter-individual variation in athletic performance. It has long been considered that some individuals may have an unquantified innate advantage in sporting competition (Pitsiladis & Scott, 2005). Initial attempts to establish the basis of any genetic component on physical performance were through studies of the heritability of selected fitness parameters such as $\dot{V}o_{2\,max}$. Such studies make comparisons of the phenotypic intra-pair difference in monozygotic and dizygotic twins, or measure the extent to which phenotypes aggregate within families relative to between different families (For review of methodologies, see (Bouchard *et al.*, 1997)). Early heritability estimates of $\dot{V}o_{2\,max}$ have been as high as 93 % (Klissouras, 1971), although subsequent measures have often been lower (Fagard *et al.*, 1991;Lesage *et al.*, 1985). Despite the equivocal results arising from such studies, which are likely to be the result of a number of confounding factors (Klissouras, 1997), it seems that "nature" does have a role to play in the determination of exercise performance. A number of studies have suggested significant genetic effects for a number of parameters of fitness in the sedentary state, such as $\dot{V}o_{2\,max}$ (Bouchard *et al.*, 1998), ventilatory threshold (Gaskill *et al.*, 2001), and muscle fibre type proportion (Simoneau & Bouchard, 1995). In elite athletes, it is likely that the phenotypic adaptation to training is of more importance than baseline values. To that effect, in addition to baseline values, studies have also shown that there may be a genetic component to

"trainability" of a variety of phenotypes important to exercise performance. The finding that there was 2.5 times more variance between families than within in the $\dot{V}o_{2\,max}$ response to a standardised training program (Bouchard *et al.*, 1999), supports the idea that there is also a genetic component to the adaptive response to training. These findings have prompted the study of the genome to identify the particular gene variants that are responsible for this unquantified genetic effect.

1.3.2 Measured genotype approaches

Recent advances in molecular technology have allowed the identification of a number of genetic variants associated with fitness related phenotypes. Rather than infer a genetic component through the assessment of phenotypic aggregation between and within families or twins (unmeasured genotype approaches), these studies identify particular genetic loci at the DNA level and their potential influence on performance phenotypes (measured genotype approach). Exercise performance is a complex, multi-factorial trait, which is the result of many contributors: physiological, anatomical, and biochemical. This makes the identification of genetic contributors difficult, given that that many small gene effects are likely to be more influential than major single gene effects in accounting for the inter-individual variation in the general (non-pathological) population. Nevertheless, the human gene map for performance and health related fitness phenotypes (Rankinen *et al.*, 2001;Rankinen *et al.*, 2002;Perusse *et al.*, 2003;Rankinen *et al.*, 2004;Wolfarth *et al.*, 2005), a yearly publication summarising genes and gene variants contributing to human performance, now includes over 150 entries. These include 140 autosomal, four on the X-chromosome and 16 mitochondrial DNA entries (Wolfarth *et al.*, 2005). Many of these gene variants influence some of the inter-individual variation in performance; although few have yet been identified as important in attaining elite level performance. Studies to identify gene variants and their influence fall into different categories. These include case-control association studies, cross-sectional association studies and linkage studies. Association studies compare genotype frequencies between groups of elite athlete and controls (case-control) or assess genotypic variation between individuals of different genotype (cross-sectional). Linkage studies, however, use family pedigrees to assess how alleles cosegregate with a given phenotype. Linkage studies can identify areas of the genome that may be influential in phenotypic variation. These areas can then be investigated further to identify resident genes, which may serve as potential candidate genes for human performance.

One of the first, and the most studied of the putative candidate genes for human performance is the Angiotensin converting enzyme (ACE) gene, where an insertion polymorphism (I) in intron 16 of the gene is associated with lower levels of circulating (Rigat *et al.*, 1990) and tissue ACE than the deletion (D) (Danser *et al.*, 1995). The ACE I/D polymorphism has been associated with performance in mountaineers, where climbers were found to differ significantly from healthy controls in their genotype distribution, showing an increased frequency of the I allele (Montgomery *et al.*, 1998). It was also shown that homozygotes for the I allele had an eleven fold greater improvement in a bicep curl movement after training (Montgomery *et al.*, 1998). The belief that the I allele and, therefore, lower ACE levels improve endurance performance was sustained by the finding that the frequency of the I allele increased linearly with event duration in a study of runners competing in events of <200m, 400-3000m, or >5000m (Myerson *et al.*, 1999). An excess of the I allele has also been found in elite Australian rowers relative to Australian controls (Gayagay *et al.*, 1998), and in elite Italian aerobic athletes relative to controls (Scanavini *et al.*, 2002). Similar findings have been made in Spain, where elite athletes had a significantly higher frequency of the I allele than controls (Alvarez *et al.*, 2000). However, studies assessing the association between the ACE gene and fitness related phenotypes are far from conclusive. Findings of an association between the ACE I allele and $\dot{V}o_{2\,max}$ in postmenopausal women (Hagberg *et al.*, 1998) have not been replicated in later studies, albeit not in the same subject group (Woods *et al.*, 2002;Rankinen *et al.*, 2000a). Furthermore, a study of elite endurance athletes of mixed discipline relative to sedentary controls found no association between the ACE gene and elite athlete status (Rankinen *et al.*, 2000b).

Other gene variants have been identified as influential in attaining elite level performance. α-actinin-3 is a major component of the skeletal muscle Z-line in type II muscle fibres, where they anchor together adjacent actin filaments and maintain sarcomeric integrity (Blanchard *et al.*, 1989). A variant of this gene codes for a premature stop codon (R577X) and subsequent α-actinin-3 deficiency (North *et al.*, 1999). Homozygotes for the null variant (577XX) have been found to be less frequent in power athletes relative to endurance athletes or non-athlete controls, and were not present at all in Australian Olympic level power athletes (Yang *et al.*, 2003). Conversely, female endurance athletes had a higher frequency of the 577XX genotype relative to controls (Yang *et al.*, 2003). The gene map of health and fitness related phenotypes gives a detailed review of established gene associations (Wolfarth *et al.*, 2005).

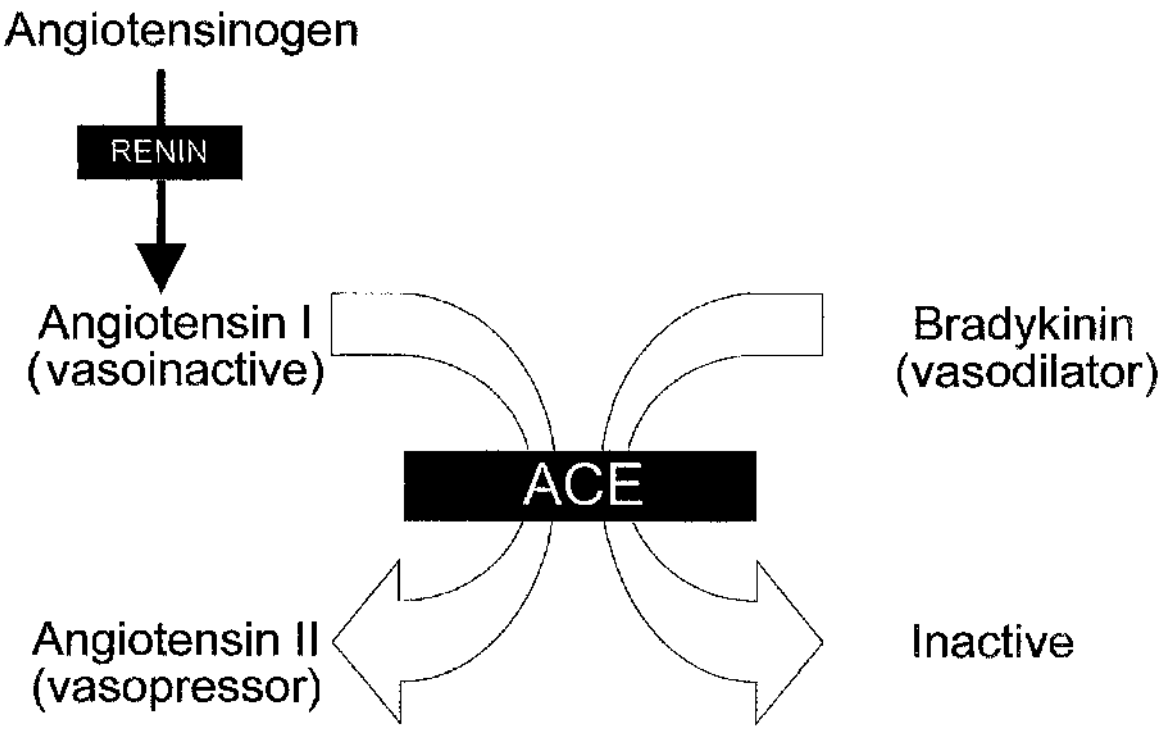


Figure 1.2 The primary role of ACE in converting Angiotensin I to Angiotensin II is shown, along with the degradation of the vasodilator bradykinin to inactive forms.

As well as variants in the nuclear genome, polymorphisms in mitochondrial DNA (mtDNA) have also been suggested to influence the variation in human performance. mtDNA is a small (16,569 bp), cytoplasmic genome which encodes 13 subunits of a number of enzyme complexes of oxidative phosphorylation, as well as components of the mitochondrial protein synthesis system (Anderson *et al.*, 1981). It is subject to a matrilineal pattern of inheritance, which results in the accumulation of linked complexes of mutations down different lines of descent from a single ancestor molecule. This non-mendelian pattern of inheritance can be used to trace the ancestry of individuals or populations (Richards *et al.*, 2000). mtDNA is also a potential candidate to contain variants influential in the interindividual differences in human performance, given that it codes for a number of components of oxidative phosphorylation. Family based studies have often shown a strong maternal influence in the heritability of aerobic performance phenotypes such as $\dot{V}o_{2\,max}$ (Lesage *et al.*, 1985;Bouchard *et al.*, 1999;Bouchard *et al.*, 1998), hinting at a role for mtDNA variation in the determination of aerobic capacity. It has been suggested that polymorphisms in mtDNA are associated with $\dot{V}o_{2\,max}$ in the untrained state and with

variation in the trainability of $\dot{V}o_{2\,max}$ (Dionne *et al.*, 1993). However, a further study found no frequency difference in mtDNA polymorphisms between sedentary controls and elite endurance athletes (Rivera *et al.*, 1998). Although no differences in mtDNA encoded oxidative capacity were seen between elite athletes and sedentary controls (Soma *et al.*, 2001) or were found to account for differences in endurance capacity and its trainability (Murakami *et al.*, 2001); it has been suggested that polymorphisms in the control region of mtDNA may account for some of the inter-individual variation in aerobic performance (Murakami *et al.*, 2002).

1.3.2.1 The genetics of race and East African running dominance

Evidence for a genetic component to the multi factorial trait of exercise performance is accumulating (Wolfarth *et al.*, 2005), and there is also a small body of evidence to suggest that certain genes are influential in elite performance (Gayagay *et al.*, 1998;Yang *et al.*, 2003). However, there is currently no evidence to suggest that the success of east African athletes is a genetically conferred phenomenon, despite the perpetuation of the idea that black athletes are genetically adapted for athletic performance (Hamilton & Weston, 2000). This notion is based on a preconception that each race constitutes a genetically homogeneous group, with race being defined simply by skin colour. This belief is contrary to the assertion that there is more genetic variation within Africa than between Africa and Eurasia (Yu *et al.*, 2002). The genetics of race is an understandably controversial topic and has produced a number of contrasting viewpoints. Some argue that there is a role for race, and that the potential benefits to be gained in terms of diagnosis and research of disease outweigh the potential social costs of linking race or ethnicity with genetics (Burchard *et al.*, 2003). Others, however, advocate that race should be abandoned as a tool for assessing the prevalence of disease genotypes in certain populations, and that race is not an acceptable surrogate for genetics in assessing the risk of disease in human populations (Cooper *et al.*, 2003). Arguments for the inclusion of race in biomedical research often focus on the use of the term to identify single gene disorders and their medical outcome, and it is acknowledged that the genetic basis of complex phenotypes, such as athletic performance, are poorly understood and far more difficult to study. It is estimated that most human genetic variation is shared by all humans and that a marginal proportion (normally less than 10 %) is specific to major continental groups (Cavalli-Sforza & Feldman, 2003). Estimates from the human genome project and analysis of haplotype frequencies show that most haplotypes (linked segments of genetic variation, rarely subject to reassortment by recombination) are shared between two of the three major geographic

populations: Europe, Asia and Africa (Gabriel *et al.*, 2002). Alleles found in one population are usually common to all human populations, particularly when they occur at a frequency of over 20 % in one population (Burchard *et al.*, 2003). Indeed it is currently estimated that the level of genetic diversity between human populations is not large enough to justify the use of the term race (see Jobling et al (2004) for a review). Consequently, any differences in physiology, biochemistry and/or anatomy between groups defined solely by skin colour (e.g. black versus white) are inapplicable to East African runners, even if the differences found are indeed genetically determined. This highlights the shortcomings of approaches comparing groups classified by phenotypes such as skin colour and particularly in extrapolating results from such studies to other groups of similar skin colour.

1.4 Aims and objectives

Given the wealth of opinion and debate on the factors influencing the success of east African running and the accumulating evidence for a significant genetic component underlying the inter-individual variation in human performance, the main objectives of the following research were:

i. To investigate the environmental characteristics associated with elite east African status, and in doing so, better understand the reasons for the phenomenal success of east Africa in distance running. This was achieved by investigating the regional and ethnic origins of east African athletes, in response to evidence that they are not equally distributed throughout east Africa, but are confined mainly to distinct regions and ethnic groups. The suggestions that altitude inhabitation and the distance that east Africans travel to school as children are influential in their success were also investigated.

ii. To establish if the distinct ethnic characteristics of elite athletes are indicative that the athletes arise from a limited genetic isolate, selected for endurance performance, through analysis of non-recombining molecules which allow the ancestry of populations to be traced. This was achieved by analysis of mtDNA and Y chromosome variation in elite east African distance runners relative to controls and also by comparing regions producing a large number of athletes to other regions to establish the level of population stratification throughout Ethiopia and Kenya.

iii. To investigate any genetic influences on the performance of east Africa in distance running through the analysis of existing candidate genes for human performance, and also by using the novel technique of mtDNA and Y chromosome haplogroup analysis of east African athletes relative to the general east African population. In doing so, this would provide the first information on the association between physical performance and genetic variation in east African subjects.

iv. To validate existing associations of genes for human performance by establishing their frequency in a cohort of the world's most successful athletes. To date, all genetic associations with human physical performance have been in Caucasian subjects. The following experiments aimed to replicate existing associations in African subjects. If the associations could be replicated in a cohort of the world's most successful athletes, of different ethnic background from those in the existing literature, it would add credibility to these gene variants having a functional influence on endurance performance.

Chapter 2

General Methods

2.1 Subject Populations

2.1.1 Ethiopia

For the studies in Chapters three, five, seven and eight, 225 Ethiopian subjects provided written informed consent to participate in the study, which was approved by the University Ethics committee and local Ethiopian authorities (Appendix 1). Subjects comprised 114 athletes, all members of the Ethiopian national senior and junior athletic teams, and 109 control subjects (C). The control group comprised 109 students at Kotebe Teaching College, Addis Ababa, and were representative of the general population. Ethiopian colleges and Universities have a rationing system where places are allocated per region, so individuals from all regions and ethnicities of Ethiopia were represented. Consent was obtained from subjects through a consent form (Appendix 1), which was translated into Amharic (the national language of Ethiopia) to ensure comprehension. For the studies in chapter six and eight, it was possible to return to Ethiopia to collect an additional control group (N = 92) from the region of Arsi (A), from which a large proportion of athletes originate (as discussed in Chapter 3).

The athletes were classified into one of three groups, according to their specific athletic discipline. Athletes specializing in 5,000- to 10,000-m distances (5–10 km; N = 42), marathon runners (M; N = 34), and a third group, comprised of track and field athletes, track athletes being runners up to 1500 m (TF; N = 38). Athletes in the marathon and 5- to 10 km groups were truly elite athletes, regularly successful in international distance running, and both groups included World and Olympic champions. All subjects responded to a questionnaire detailing ethnic and environmental background (Appendix 1), and also provided a buccal swab from which DNA could be extracted for analysis of key genes presented in the following chapters.

2.1.2 Kenya

404 elite Kenyan athletes and 87 Kenyan control subjects provided written informed consent to participate. Again, the study was approved by the University ethics committee, and by local Kenyan authorities (Appendix 2). Kenyan control subjects were students at Kenyatta University, Nairobi and were representative of the general Kenyan population in their geographical distribution throughout Kenya.

Athletes were classified into one of two groups, dependent on their athletic success. Athletes who had represented Kenya at international standard competition were classified as international athletes (I), and those competing within Kenya at the National level were classified as national athletes (N). The international athletes comprised many Olympic and World champions, as well as World Record holders. These athletes were not limited to those currently in competition but also to those athletes who had previously achieved athletic success on behalf of Kenya, in competitions from the 1960s to the present day. DNA samples were obtained from 70 such athletes, of whom, 42 had won Olympic, World or Commonwealth medals, had a top 3 finish in an international marathon or equivalent road race, or had been ranked in the top 50 runners in the world at their event. All subjects responded to a questionnaire detailing ethnic and environmental background (Appendix 2), and also provided a buccal swab from which DNA could be extracted for analysis of key genes, presented in the following chapters. In addition, all Kenyan controls also provided a supine 10 ml venous blood sample.

2.2 Demographic information

Questionnaires obtained information on the demographic and environmental characteristics of all subjects. This included information on their place of birth, ethnicity, and distance and method of travel to school.

2.2.1 Place of birth

Information collected on the subjects' place of birth allowed identification of particular regions with a disproportionately high number of athletes, in response to reports that the majority of the most successful east African runners originate from distinct regions of East Africa (Bale & Sang, 1996;Manners, 1997).

2.2.2 Language

A common language is often indicative of common origin, and a related language (i.e., a language of the same family), suggests a common origin, but one that dates farther back in time. This was to establish if elite athlete originate from distinct ethnic groups within east Africa.

2.2.3 Distance and Method of travel to school

Subjects were asked to provide information on the distance that they travelled to school as children and the means of transport that they used. This was to assess the validity of reports linking distance travelled to school to East African running success.

2.3 Molecular Methods

2.3.1 DNA collection and extraction

DNA was collected by buccal swabs using cell cytology brushes (Medical Packaging Corporation, CA, USA). Swabs were stored in 1ml of cell lysis solution (0.1M EDTA, 1 % SDS, 0.1M Tris-HCl. PH = 7.6) to ensure safe transit of DNA. On return to Glasgow, buccal swabs were stored at –20°C, and DNA was subsequently extracted using either Puregene buccal swab extraction protocol (PureGene, Gentra Systems, MN, USA), or the Qiagen buccal cell spin protocol (Qiagen Ltd., Crawley, UK).

2.3.2 Genotype determination

All PCR reactions were performed in a 96 well thermocycler (Px2, Thermo Electron Corporation, UK) in 96 well plates (Omega Scientific Inc, CA, USA). Reactions were in a 25µl total volume (unless specified otherwise) with 12.5 µl of 2X Reddymix™ mastermix, containing 1.5mM $MgCl_2$ (ABgene, Epsom, UK), 0.4µM of each primer, 1µl of total DNA (estimated 10 ng) and H_2O. DNA fragments were visualised under UV transillumination after separation on 2 % agarose gels in comparison to 100bp ladder (Promega, WI, USA). All restriction endonucleases (HpaI, DdeI, AluI and BsrBI) were supplied by New England Biolabs (NEB UK Ltd, Herts, UK).

2.3.2.1 Mitochondrial DNA (mtDNA)

The HVS-I was amplified and sequenced to identify patterns of polymorphisms and known HVS-I motifs. Primers and PCR conditions are shown in Table 2.1. Actual sequencing reactions were carried out as follows at the Sir Henry Wellcome Functional Genomics Facility, University of Glasgow. Reverse strands were sequenced using fluorescent dideoxynucleotides in BigDye™ terminator cycle sequencing ready reaction (Applied Biosystems, Foster City, CA, USA). In sequences with a T189C polymorphism, a C-string occurred, which was followed by incoherent sequence. In such cases, both forward and

reverse strands were sequenced to characterise the entire range of HVS-I. Sequencing products were separated by 5 % denatured Long Ranger™ gel (FMC BioProducts, Rockland, Maine, USA) and detected using a PE Applied Biosystems 377 DNA sequencer. Sequences were read using Chromas (version 1.45, Technelysium Pty Ltd, Australia), and differences from the Cambridge Reference Sequence (CRS) (Anderson *et al.*, 1981) were recorded. Three known restriction fragment length polymorphisms (RFLPs) of mtDNA were amplified by PCR and digested (Table 2.1). HVS-I sequencing was performed from a 50 µl total volume, containing 46 µl of 1.1X ReddyMix (ABGene, Epsom, UK), 0.2 µM of each primer and 2 µl of total DNA (estimated 10 ng). RFLP PCRs contained a total volume of 20 µl with 16 µl of 1.1X ReddyMix, 0.4 µM each primer and 2 µl of total DNA (estimated 10 ng). DNA was denatured for 5 min at 94 °C followed by 30 cycles of denaturation (30 s, 94 °C), annealing (45 s, annealing temperature shown in Table 2.1), and elongation (60 s, 72 °C). A final elongation step was performed (10 min, 72 °C) before samples were stored at 4 °C. Negative controls with no DNA were included in each sample run to test for cross contamination during reaction preparation. RFLP digestions were performed under the manufacturers recommended conditions (NEB UK Ltd, Herts, UK). Samples were incubated at 37 °C for at least 10 h before the fragments were separated on a 2 % agarose gel, stained with ethidium bromide and visualised on a UV transilluminator.

All subjects were screened for known polymorphisms C3594T, A10398G, and C10400T, which cause restriction site changes -3592 *Hpa* I, +10394 *Dde* I, and -10397 *Alu* I respectively. The size pattern of each mtDNA reaction after digestion is shown in Figure 2.1.

	Range	Site	Enzyme		Primers	Annealing temperature	Digested PCR product (bp)	
					5' - 3'		+	-
HVS-I	15975-16450			Forward	ctccaccattagcacccaaagc	51 °C		
				Reverse	cgaggagagtagcactcttg			
RFLP								
	3389 - 3717	3594	*Hpa* I	Forward	taggctatatacaactacgc	51 °C	206, 123	329
				Reverse	ggctactgctcgcagtg			
	10269-10579	10398	*Dde* I	Forward	tcctttaccctaccatgag	53 °C	87, 38, 185	87, 223
				Reverse	attattccttctaggcatagtag			
	10270-10569	10400	*Alu* I	Forward	tcctttaccctaccatgag	53 °C	129, 181	310
				Reverse	attattccttctaggcatagtag			

Table 2.1 mtDNA genotyping conditions.

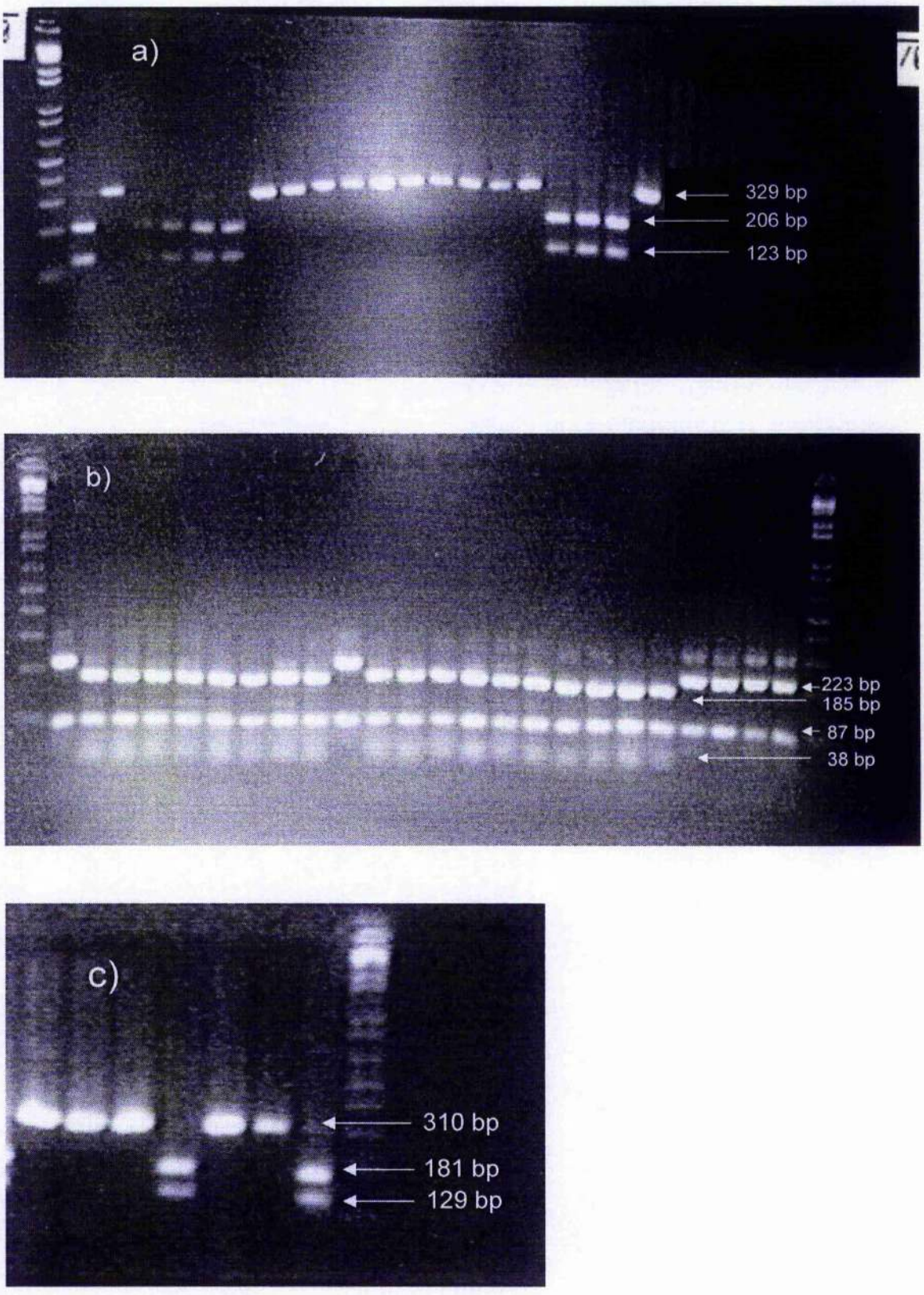


Figure 2.1 Size fragments of mtDNA digests. a) Shows size fragments of *Hpa* I digest. b) Size fragments of *Dde* I digest. c) Size fragments of *Alu* I digest.

2.3.2.2 ACE Gene

New primers were designed for all genotyping reactions. I/D genotypes were determined by PCR, using a three-primer system; a forward primer which specifically recognised the deletion (D) sequence, a forward primer which specifically recognized the insertion (I) sequence, and a common reverse primer. The forward deletion primer was (5' – CTCTAGACCTGCTGCCTATTACAGTC - 3'), the forward insertion primer was (5' – CGGGATGGTCTCGATCTC – 3') and the common reverse primer was (5' - CCCTCCCATGCCCATAAC - 3'). PCR conditions for the I/D polymorphism reaction consisted of 35 cycles of denaturation at 94 °C, annealing at 56.5 °C, and extension at 72 °C. Cycling was preceded by 3 min of denaturation at 94 °C and followed by 10 min of extension at 72 °C. If present in the template DNA, the D allele yielded a product of 197 bp and the I allele a product of 252 bp. The forward D primer also imperfectly recognised a second site only present in the I allele, which could result in the production of a 485 bp band, however, this was non-essential for genotyping and inconsistently produced with poorer quality DNA samples. A 495 bp section was amplified around np 22982 and restriction enzyme digest assays were used to identify genotype at A22982G. The forward primer was (5' - AGTACAACTGGACGCCGAAC – 3') and the reverse primer was (5' – GTGTGGGGTGTGTCCTAG – 3'). PCR conditions for the A22982G polymorphism consisted of 35 cycles of denaturation at 94 °C, annealing at 61 °C, and extension at 72 °C. Cycling was preceded by 5 min of denaturation at 94 °C and followed by 10 min of extension at 72 °C. Unlike the 22982 A allele, the PCR product with the 22982 G allele was cleaved into a 325 bp and 170 bp product after digestion with restriction enzyme *BsrB* I, as shown by Figure 2.2.

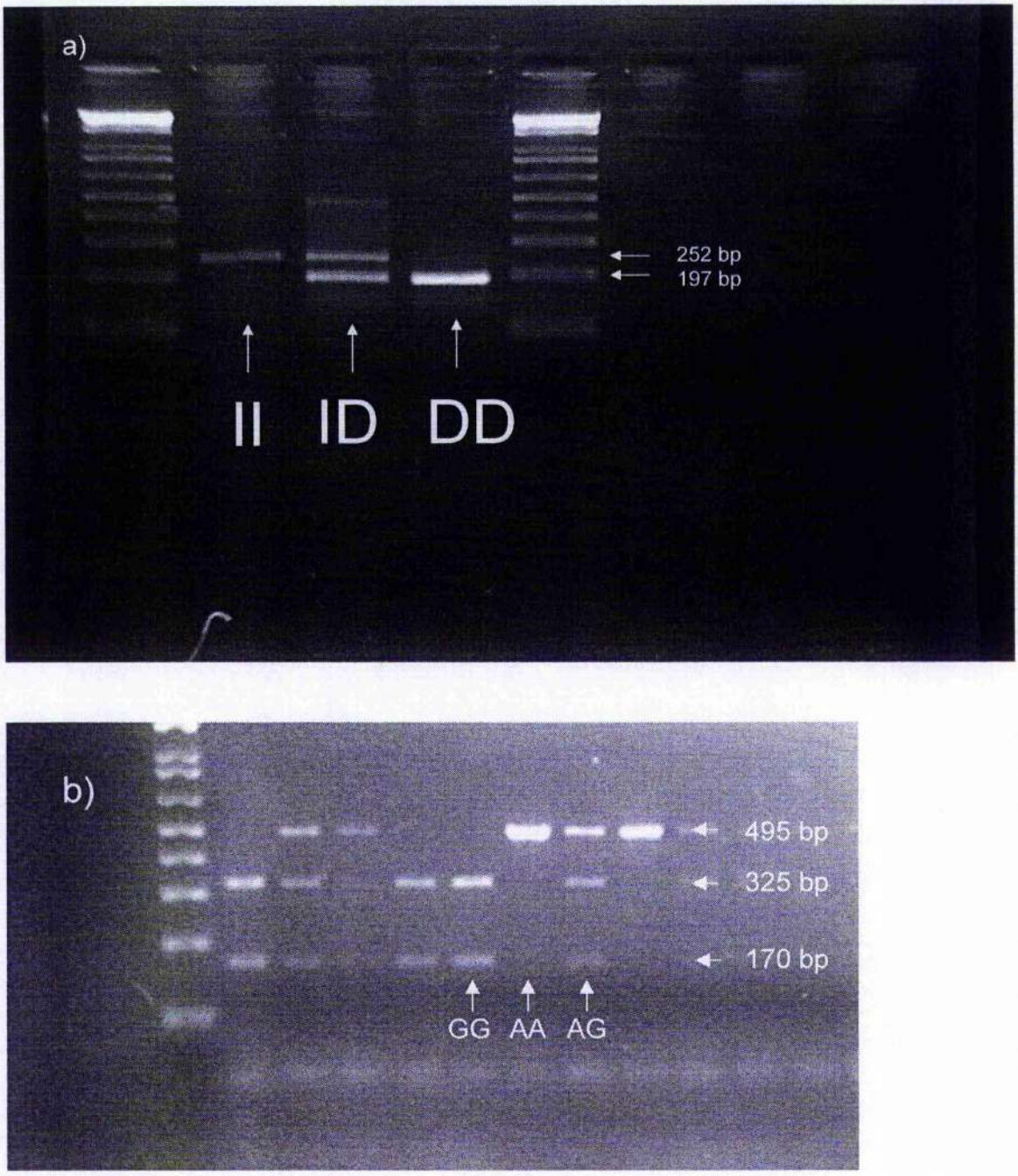


Figure 2.2 ACE genotype reactions. a) shows size fragments of I/D PCR reactions. b) shows size fragments of A22982G digest with BsrBI.

2.3.2.3 Y-Chromosome

Y chromosome SnaPshot reactions were carried out in the Department of Genetics, University of Leicester, as follows. A set of up to 30 Y-chromosomal binary markers (Figure 7.1) was typed in a hierarchical fashion by using the SNaPshot (Applied Biosystems, CA, USA) minisequencing method according to manufacturer's instructions on an ABI3100 capillary electrophoresis apparatus (Applied Biosystems). Sequences of amplification and minisequencing primers were as described by Paracchini et al. (2002) except for markers M32, M69, M75, M182, M201, M207, P2, P25 and 12f2.1 (Y Chromosome Consortium, 2002) for which new primers were designed. Samples were typed blind with respect to athletic status. Haplogroups were defined according to the Y chromosome consortium nomenclature (Y Chromosome Consortium, 2002).

2.4 Data analysis

All questionnaire data was collated, stored and analysed using Microsoft Excel (Microsoft, CA, USA). Statistical analyses were performed, unless specified otherwise, in Minitab (version 13.30, Minitab Ltd, Coventry, UK). Inter-group differences in genotype frequency were tested by chi-square (χ^2) tests, with statistical significance declared at $P < 0.05$. The specific methods of analysis are detailed in each chapter as appropriate to each study. Due to the higher resolution analysis of Y chromosome haplogroups, and subsequently lower subject numbers in each field, exact tests of population differentiation were carried out to establish frequency differences between groups.

Subjects were tested at each locus for HWE using χ^2 tests with one degree of freedom. They were tested as an entire subject group, comprising athletes and controls, and then again when separated by gender. HWE was further tested in sub-groups of athletes and controls. Kenyan and Ethiopian female athletes were tested as an entire group due to low subject numbers. Ethiopian controls were also tested as an entire group due to low numbers of females. To allow statistical analysis of regional distribution of subjects and to maintain subject numbers, it was necessary to collapse subject groups containing small numbers of subjects into one larger group: 'other'. The circumstances in which this method was used are detailed in Chapters three and four.

Necessarily small numbers of elite athletes placed a limit on statistical power. Although the unavoidably low numbers of elite athletes is a limitation of the following work, certain steps were taken to maximize the impact of the work. The athletes are homogenous in their

athletic discipline: all being elite distance runners. All of the elite Ethiopian athletes were members of the Ethiopian national athletics squad, and Kenyan international athletes had represented Kenya in international distance running competition. These criteria ensure that the athletes are of the highest calibre, which, although not raising statistical power, does limit the chance of a type II error as much as possible. In attempt to avoid making type I errors, a statistical correction can be applied to account for multiple testing. In this instance, no correction was made for multiple testing. However, positive results occur more frequently than would be expected by chance, as detailed in the following chapters. Caution is taken in the interpretation of results

Chapter 3

Demographic Characteristics of Elite Ethiopian Athletes

3.1 Introduction

The unanswered questions surrounding the success of east African distance runners have generated a plethora of studies attempting to elucidate putative contributory mechanisms. Proposed explanations have included favourable physiological characteristics (Coetzer *et al.*, 1993;Bosch *et al.*, 1990;Saltin *et al.*, 1995a;Saltin *et al.*, 1995b;Weston *et al.*, 2000;Weston *et al.*, 1999), which may include favourable genetic endowment (Larsen, 2003), advantageous environmental conditions, such as being born and raised at altitude (Saltin, 1996;Schmidt *et al.*, 2002), running a long way to school each day (Saltin, 1996;Saltin *et al.*, 1995b), and psychological advantage (Hamilton, 2000).

In studies comparing groups of black and white athletes, findings have included lower lactate levels at a given exercise intensity (Coetzer *et al.*, 1993;Saltin *et al.*, 1995b;Weston *et al.*, 1999), better running economy (Saltin *et al.*, 1995b;Weston *et al.*, 2000) and higher fractional utilization of $\dot{V}o_{2\,max}$ in the black athletes (Coetzer *et al.*, 1993;Bosch *et al.*, 1990;Weston *et al.*, 2000). However, these studies have classified groups based primarily on skin colour, not accounting for the fact that there is more genetic difference within race groups than between (Cavalli-Sforza & Feldman, 2003). The validity of extrapolating such results to account for East African running dominance is therefore questionable.

A study by Schmidt et al. (2002) proposed that chronic altitude exposure and endurance training combine synergistically to induce positive haematological adaptations, accounting for the success of East African endurance athletes. An earlier study by Saltin et al. (1995b) comparing various physiological variables between Kenyan and Scandinavian distance runners, concluded that the superior performance of the Kenyan athletes was not likely to be directly attributable to altitude. However, the finding that Kenyan boys who travelled a long distance to school by foot had a $\dot{V}o_{2\,max}$ 30 % higher than those who did not (Saltin *et al.*, 1995b) supported the belief that the distance travelled to school by African athletes gives them some advantage in endurance athletics.

Ethiopia was selected as the model for the present study, taking into account the fact that Ethiopian athletes boast a recent success record in international distance running second only to Kenya. With a population of 65 million, the third highest in Africa, the Ethiopian population displays a high degree of heterogeneity. Since 2,000 B.C., the indigenous Ethiopian population has been classified into three clusters of people (Omotic speakers, Cushites and Semites), the distribution of which has remained largely unchanged to the

present day (Passarino *et al.*, 1998). To our knowledge, no investigation to date has attempted to trace the ethnic origins of Ethiopian distance runners and, by doing so, examine the possibility that they may share a common ethnic background and possibly some form of homogeneity relative to the general Ethiopian population.

The present study, therefore, aimed to determine the ethnic and environmental background of elite Ethiopian distance runners. The findings were then compared to those of the general 'non-athlete' Ethiopian population, to assess whether the athletes were of a distinct ethnic or environmental background. Reports that have linked living at altitude and running long distances to school with athletic success of African athletes, particularly in endurance events, were also examined.

3.2 Methods

3.2.1 Subjects

Subjects and subject groups are as detailed in Chapter two. Questionnaires, translated into Amharic, the national language of Ethiopia, to ensure comprehension, were administered to the subjects participating in this study. The questions included in these were designed to obtain the following information:

1. Place of birth. This was classified according to the 14 provinces of Ethiopia from 1946-1980 (Henze PB, 2000). We wished to identify any particular regions with a disproportionately high number of elite athletes and possibly, therefore, elucidate any link between distance running success and altitude.

2. Spoken Language (and that of their grandparents). This was to provide further information on ethnicity. A common language is often indicative of common origin, and a related language (i.e., a language of the same family), suggests a common origin, but one that dates farther back in time. At present, over 70 languages are in every day usage in Ethiopia, most of which belong to the Afro-Asiatic family. The majority of these languages belong to either the Semitic or Cushitic subset of languages. To simplify the analysis in the present study, three separate language categories were used: Semitic, Cushitic and Other (which includes the less common subsets within the Afro-Asiatic language family such as Omotic).

3. Mode of travel to school (walk; run; other), and distance travelled (less than 5 km; 5-20 km; other). This was to assess the validity of reports linking distance travelled to school to East African running success.

3.2.2 Data analysis

For each section of the questionnaire, chi-squared tests were initially performed to establish the presence of significant differences between the subject groups (Place of birth: d.f. = 12, Language, distance and method of travel to school: d.f. = 6). Pairwise chi-squares were then performed to identify between which groups the differences lay. Statistical significance was declared at $P < 0.05$. The fourteen regions appearing in the questionnaire responses were collapsed into five categories according to the most common responses. Regions accounting for less than seven percent of the total subject number were collapsed into the category 'Other'. In Figure 3.1, the 'Other' category also represents certain subjects whose birthplace could not be assigned into one of the fourteen regions used for classification purposes. In Figure 3.2, the language of both sets of grandparents was used to minimise the influence on spoken language of the recent shift towards Amharic as the national language of Ethiopia.

3.3 Results

3.3.1 Place of birth

The regional distribution of controls differed significantly from all athlete groups (C Vs. TF: $\chi^2 = 12.2$, $P = 0.016$, C Vs. 5-10 km: $\chi^2 = 24.7$, $P < 0.001$, C Vs. M: $\chi^2 = 48.2$, $P < 0.001$). The marathon athletes differed significantly from track and field athletes ($\chi^2 = 16.3$, $P = 0.002$) but not to those of the 5-10 km team ($\chi^2 = 8.9$, $P = 0.064$). Additionally, the place of birth of the track and field athletes did not differ significantly to the 5-10 km athletes ($\chi^2 = 5.2$, $P = 0.27$). As can be seen in Figure 3.1, there is an excess of athletes, particularly marathon athletes (71 %), from the two regions of Arsi and Shewa, which, together, account for only 15 % of control subjects. The majority of controls originate from the ten regions collapsed into the 'Other' category (55 %). All groups of athletes showed an excess of subjects from the Arsi province, the association being strongest in the marathon athletes (38 %), compared to controls (3 %). The excess of athlete production in Arsi is shown in Figure 3.5.

3.3.2 Language Family

The origin of language of control subjects differed significantly from all athlete groups (C Vs. TF: $\chi^2 = 19.1$, $P < 0.001$, C Vs. 5-10 km: $\chi^2 = 21$, $P < 0.001$, C Vs. M: $\chi^2 = 107$, $P < 0.001$). The marathon athletes also differed significantly from all groups (M Vs. TF: $\chi^2 = 34.4$, $P < 0.001$, M Vs. 5-10 km: $\chi^2 = 36$, $P < 0.001$) with a predominance of languages of Cushitic origin (76 %) relative to the other groups (5-10 km: 52 %, TF: 46 %, C: 32 %), as can be seen in Figure 3.2. There was no significant difference between track and field athletes and the 5-10 km athletes ($\chi^2 = 4.1$, $P = 0.13$). Each of the athlete groups showed an excess of languages of Cushitic origin relative to controls, with the effect most pronounced in the marathon athletes, where 75 % speak languages of Cushitic origin compared to 32 % of controls.

3.3.3 Distance and method of travel to school

In distance travelled to school, the marathon athletes differed significantly from all other groups (M Vs. C: $\chi^2 = 15.1$, $P = 0.001$, M Vs. TF: $\chi^2 = 16.2$, $P < 0.001$, M Vs. 5-10 km: $\chi^2 = 13.6$, $P = 0.001$), with no significant differences present between the other groups ($P > 0.36$). As can be seen in Figure 3.3, a higher proportion of the marathon athletes travelled over 5 km to school each day (71 %) than in the other groups (5-10 km: 41 %, TF: 34 %, C: 37 %). In their method of travel to school, again, the marathon athletes differed significantly from all other groups (M Vs. C: $\chi^2 = 18.7$, $P < 0.001$, M Vs. TF: $\chi^2 = 16.4$, $P < 0.001$, M Vs. 5-10 km: $\chi^2 = 11.6$, $P < 0.001$). Controls also differed significantly from all other groups (C Vs. TF: $\chi^2 = 9.2$, $P = 0.01$, C Vs. 5-10 km: $\chi^2 = 8.4$, $P = 0.015$). Track and field athletes did not differ significantly from the 5-10 km athletes ($P = 0.37$). Figure 3.4 shows that a higher percentage of the marathon athletes ran to school each day, compared to other groups (M: 59 %, 5-10 km: 24 %, TF: 16 %, C: 21 %).

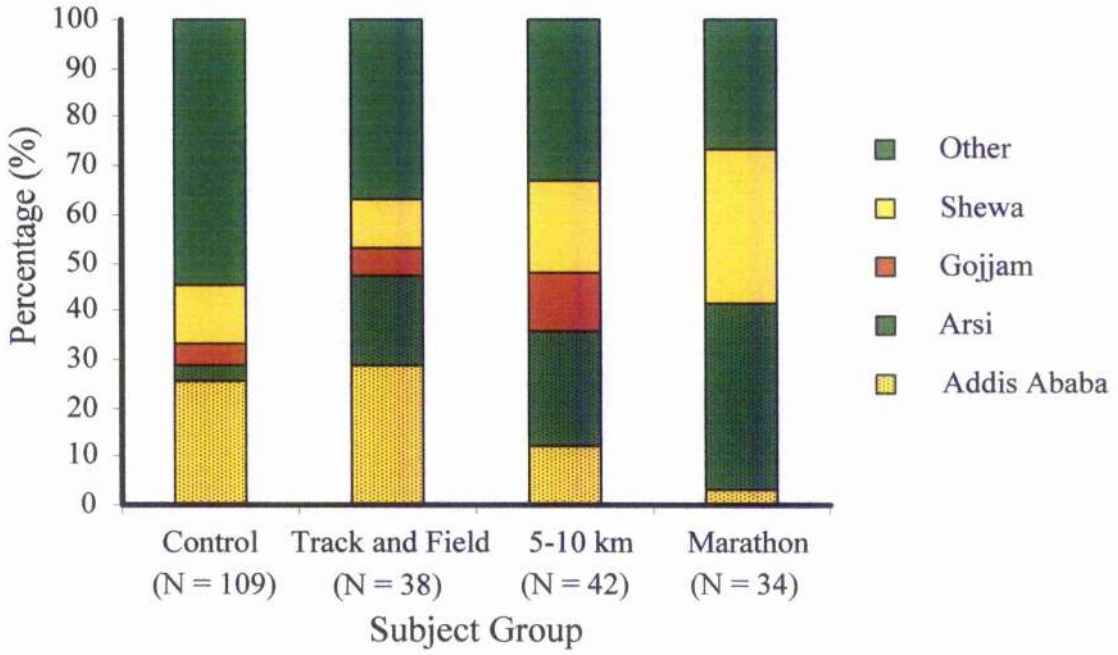


Figure 3.1: Place of Birth distributions. Figure shows percentage of subjects from each province of Ethiopia. Regional distribution of controls differs significantly from all groups. Distribution of marathon athletes differed significantly from track and field athletes (P = 0.002) and controls ($P < 0.001$) but not to the 5-10 km athletes ($P < 0.001$). Track and field athletes and the 5-10 km athletes were not significantly different ($P = 0.27$).

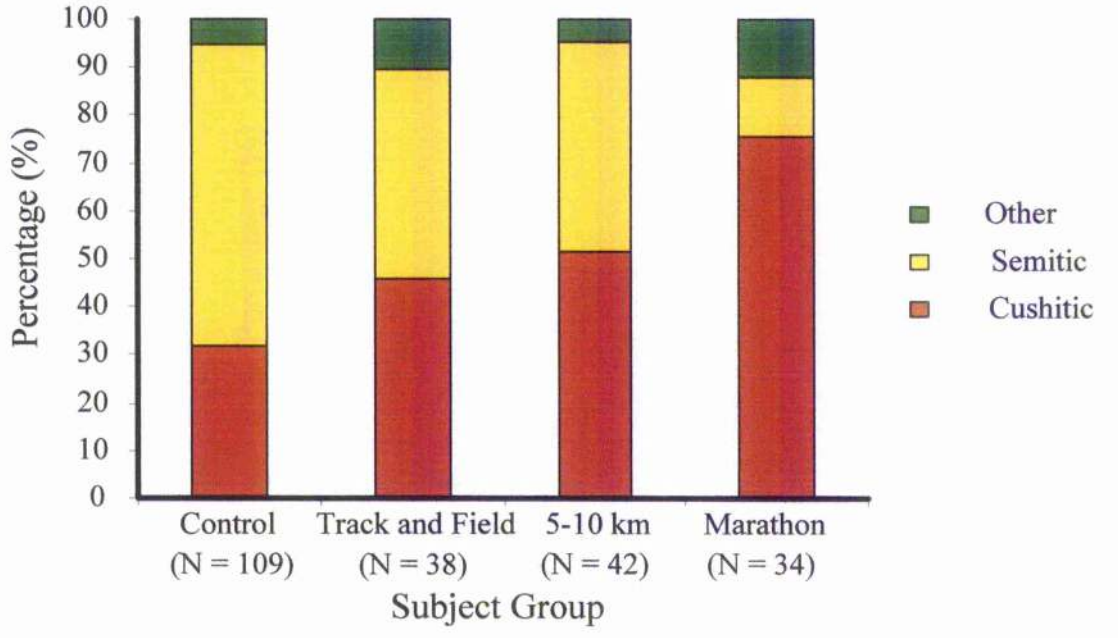


Figure 3.2: Language family distributions. Percentage of subjects speaking languages of each origin is shown. Language family distribution of marathon athletes differs significantly to all groups ($P < 0.001$), as does the distribution of controls ($P < 0.001$). Track and field athletes did not differ significantly from the 5-10 km athletes ($P = 0.13$).

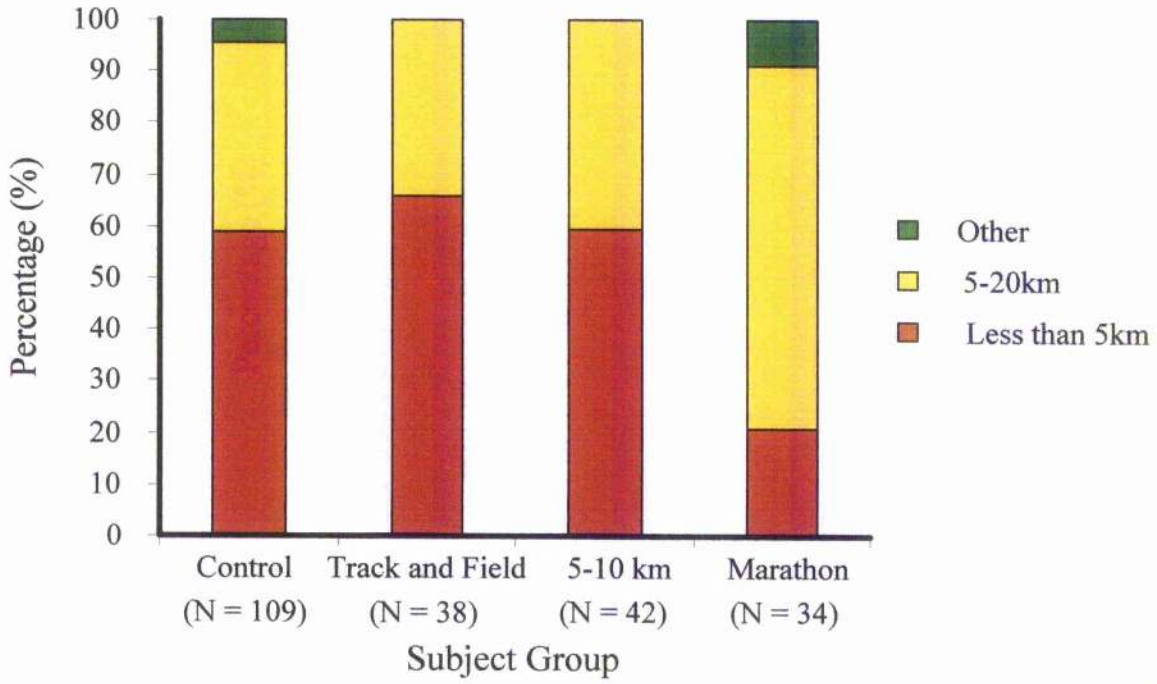


Figure 3.3: Distance travelled to School. Percentage of subjects travelling each distance to school each day is shown. The distances that the marathon athletes travelled to school each day differed significantly from all other groups ($P < 0.001$). There were no significant differences between the other groups ($P > 0.36$).

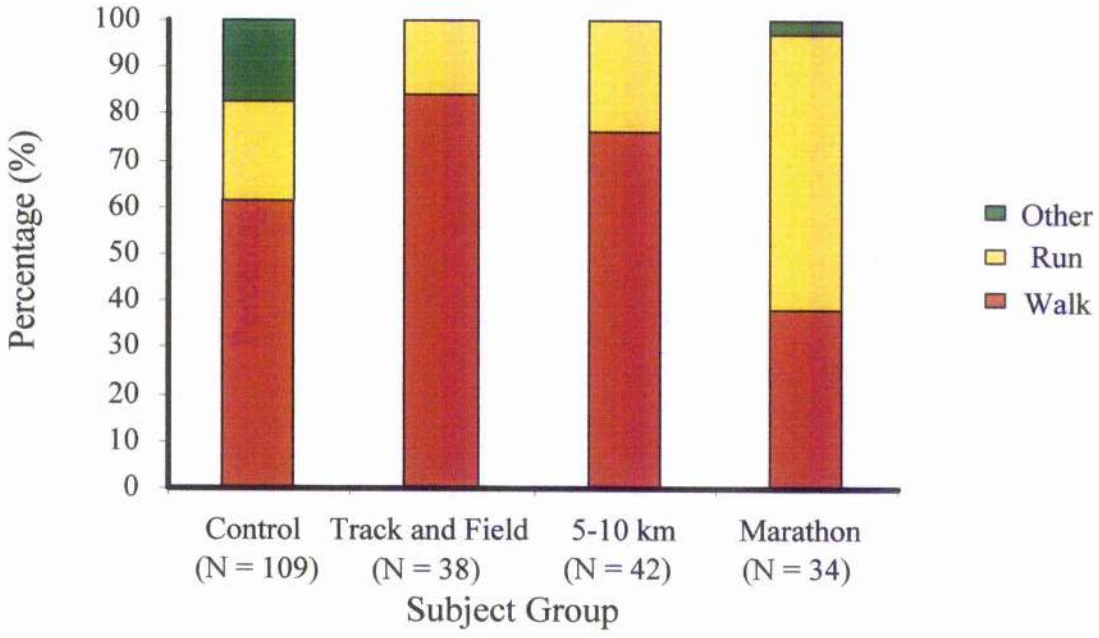


Figure 3.4: Figure shows the method of travel employed to school by each group. The marathon athletes differed significantly from all groups ($P < 0.001$), as did controls ($P < 0.015$). Track and field athletes did not differ significantly from the 5-10 km athletes ($P = 0.37$).

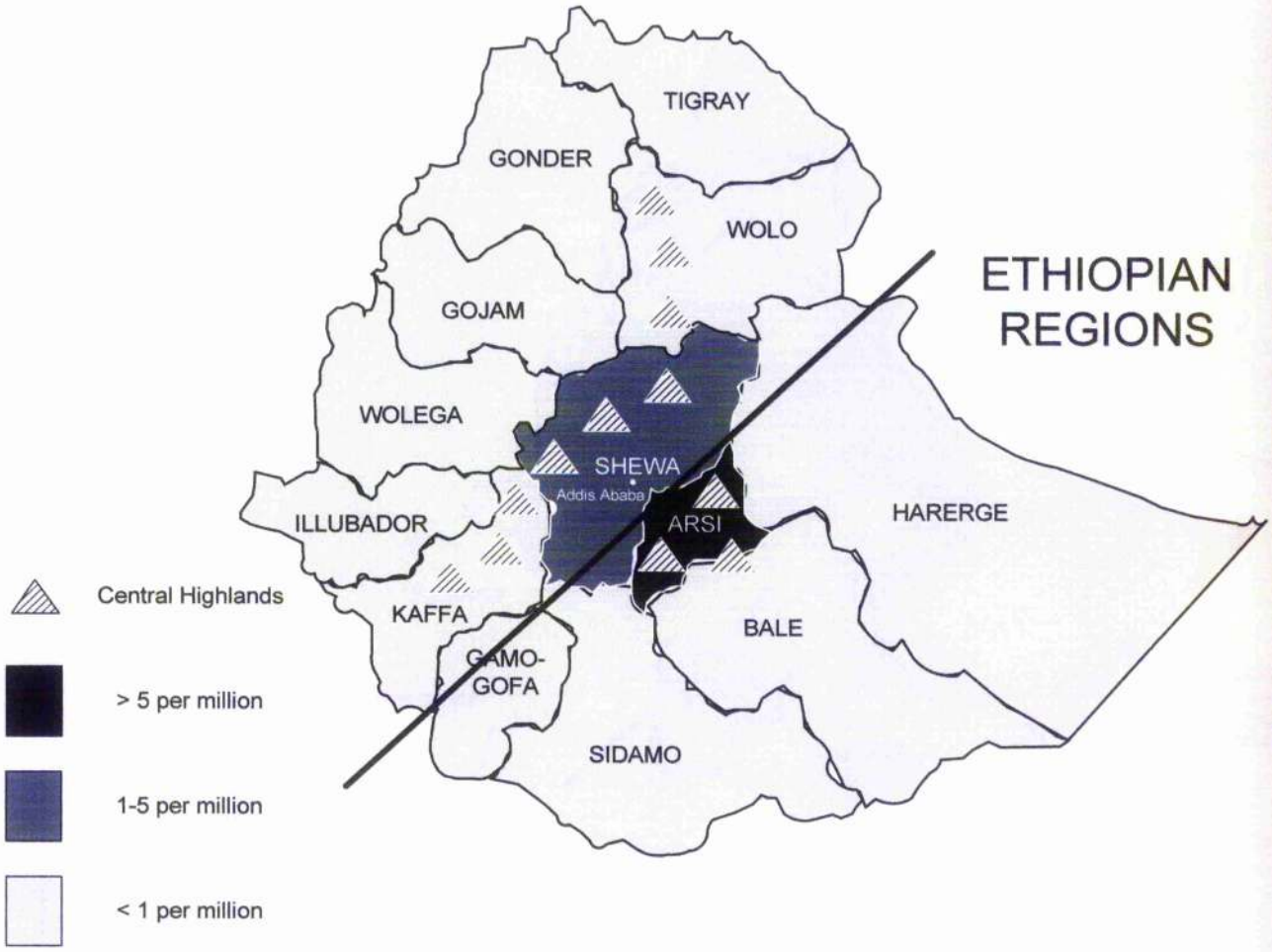


Figure 3.5: The fourteen regions of Ethiopia used in regional classification are shown. Darkly shaded regions produce a high number of athletes, light shading shows otherwise. The line represents the Rift Valley, and the Central Ethiopian highlands are denoted by triangles

3.4 Discussion

The findings of the present study are that elite Ethiopian athletes, particularly the marathon athletes, are of a distinct background, both ethnically and environmentally. The marathon athletes display a distinct environmental background relative to the other groups in all categories. Seventy-one percent of the marathon athletes are from the two regions of Arsi and Shewa, compared to controls, of whom only 15 % are from these two regions. Similar patterns are found in the language results with the marathon athletes exhibiting an excess of languages of Cushitic origin in comparison to control subjects (M: 76 %, 5-10 km: 52 %, TF: 46 %, C: 32 %). A higher percentage of marathon athletes travelled farther to school than the other groups (M: 71 %, 5-10 km: 40 %, TF: 34 %, C: 37 %), and had a higher prevalence of running to school (M: 59 %, 5-10 km: 24 %, TF: 16 %, C: 21 %). These results indicate that a number of factors, other than genetic endowment, may be influencing the athletic potential of elite Ethiopian athletes.

Although intended to be representative of the general population, the control group is comprised of students at an Addis Ababa teaching college, which may account for the excess of control subjects from Addis Ababa relative to early census data, which shows only 4 % of subjects from Addis Ababa (Tesfaghiorghis, 1986) compared to 26 % in the control group. The significant differences between controls and all groups of athletes suggest some link between place of birth and athletic ability. As can be seen in Figure 3.1, the marathon athletes have a distinct appearance relative to all other groups. Thirty-eight percent of marathon athletes are from the region of Arsi, which accounts for less than 5 % of the Ethiopian population and comprises only 3 % of control subjects, 18 % of track and field athletes, and 24 % of the 5-10 km athletes. It is interesting to note that of the marathon athletes, 26 % are from the other 10 regions of Ethiopia with a total of 71 % from the regions of Arsi and Shewa. A similar phenomenon of regional over-representation has been documented in elite Kenyan athletes. Up until 1989, 45 % of Kenyan world elite results were made by runners belonging to the Nandi tribe, who account for less than 3 % of the Kenyan population (Saltin *et al.*, 1995b). Reasons for the geographical imbalance of elite Kenyan athletes have been primarily attributed to cultural factors within the Kalenjin tribe. Kipchoge Keino, the first famous Kenyan runner was part of the Kalenjin people, and his success is thought to have spurred a tradition of distance running, sustained by talent recruitment to high school with formal training and competition (Saltin *et al.*, 1995b). It may be the case that this is mediating the Ethiopian phenomenon, as Haile Gebrselassie, the most famous Ethiopian athlete is from Arsi.

There have been suggestions that the success of Kenyan runners is in some way linked to their proximity to the Rift Valley, as the Nandi people inhabit an area adjacent to it (Entine, 2001a). It seems, from Figure 3.5, that parallels may be drawn to Ethiopia, as the Rift Valley intersects the regions of Arsi and Shewa along their common border. Although much of Ethiopia is altitudinous, the regions of Arsi and Shewa are particularly so: both in the central highlands of Ethiopia, reaching altitudes of over 4,000 m. Up to 80 % of the population in these regions live around and above 2,000 m. The finding that the marathon athletes seem to be clustered in particularly altitudinous regions may support the suggestion of a link between altitude and endurance success. However, an alternative hypothesis is presented when it is considered that Arsi is also considerably over-represented in track and field athletes (18 %). There is no perceived benefit to throwers or jumpers from altitude inhabitation for sea level performance, yet it is found that there is an excess of track and field athletes from Arsi, which may suggest that altitude is not the main influence on the prevalence of elite Ethiopian athletes from Arsi. One of the senior Ethiopian athletic coaches informed the investigators that most of the marathon athletes would be found to be from Arsi. If those in charge of athletic development believe this, it may cause a self-fulfilling prophecy through talent scouts focusing more attention to this region, or through increased regional development of athletics.

Figure 3.3 shows that a higher proportion of the marathon athletes travelled greater distances to school each day than those of the other three groups. This is comparable to the method of travel to school, where it was found that 59 % of them ran to school each day compared to only 16-25 % of the other groups. When considered alongside the findings that Kenyan boys who walked and ran long distances to school each day but did not formally train for athletics had $\dot{V}o_{2\,max}$ values some 30 % higher than those who did not (Saltin *et al.*, 1995b), the results implicate childhood endurance activity as a key determinant of the ability to become an elite Ethiopian distance runner. With the prevalence of childhood obesity in the USA and UK at an all time high (Ebbeling *et al.*, 2002), and physical activity levels among such populations in stark contrast to the daily aerobic activity of Ethiopian children, these factors may offer an explanation for the success of east African athletes on the international stage.

The marathon athletes exhibit an excess of people speaking languages of Cushitic origin (Figure 3.2), most of these speaking Oromigna, which is the language of Oromo people. The official language of Ethiopia is now Amharic, which is of Semitic origin, so the language distribution of the marathon athletes appears to be strongly influenced by some

other factor. Studies into the frequency of certain gene polymorphisms in Ethiopia have, in some cases, shown significant differences in allelic frequencies and haplotypes between the two main ethnic groups of Ethiopia: Amhara and Oromo who are of Semitic and Cushitic origin, respectively (Gennarelli *et al.*, 1999). Given the identification and association of particular gene polymorphisms with endurance performance (Wolfarth *et al.*, 2005), the excess of elite Ethiopian endurance athletes originating from one ethnic group may reflect a high frequency of potential 'performance genes' within this particular group. However, it is perhaps more likely that the distinctive ethnic origin of the marathon athletes is a reflection of their geographical distribution, as primarily Oromo people populate the region of Arsi.

3.4.1 Conclusions

The results of this study suggest that elite Ethiopian distance runners are of a distinct ethnic and environmental background, relative to the general Ethiopian population. The marathon athletes were found to have a distinct profile for all variables investigated in comparison to the general Ethiopian population, in that they tended to speak languages of Cushitic origin, and ran a long way to school each day in Arsi or Shewa. This profile is associated with endurance success, and it is likely that these environmental conditions along with the cultural and motivational factors inherent in Ethiopia, are contributors to east African distance running success. Although not excluding any genetic influence, the results of the present study highlight the importance of environment in the determination of elite endurance athlete status.

Chapter 4

Demographic Characteristics of Elite Kenyan Athletes

4.1 Introduction

Several studies have attempted to explain the success of East African runners. Proposed explanations have included environmental factors (Saltin, 1996), psychological advantage (Baker & Horton, 2003), and favourable physiological characteristics which could be genetically conferred, or environmentally determined (for review see (Larsen, 2003)). Studies have also compared physiological characteristics of black and white runners and have reported for the former lower blood and muscle lactate concentrations at a given exercise intensity(Coetzer *et al.*, 1993;Saltin *et al.*, 1995b;Weston *et al.*, 2000), better running economy (Saltin *et al.*, 1995b;Weston *et al.*, 2000) and an ability to tolerate higher fractional utilisation of $\dot{V}o_{2\,max}$ (Bosch *et al.*, 1990;Coetzer *et al.*, 1993;Weston *et al.*, 2000). However, the validity of extrapolating such findings to account for the success of east African athletes is questionable when participants have been classified into groups based primarily on skin colour, given that there are more genetic differences within "race" groups than between (Cavalli-Sforza & Feldman, 2003). Although "performance genes" have been identified as being important in elite sporting performance (Wolfarth *et al.*, 2005), no direct genetic evidence has been found to account for the success of east African runners.

Factors such as altitude and distance travelled to school each day have been proposed to account partially for the success of east African runners (Saltin, 1996). Although equivocal, some studies have concluded that chronic altitude exposure and endurance training, as experienced by many east African runners, combine synergistically to induce haematological adaptations which partially account for their success (Schmidt *et al.*, 2002). A high proportion of elite Ethiopian runners originate from particularly altitudinous regions of Ethiopia (Chapter three), although this was not necessarily causal to their success. In the study of Ethiopian athletes in Chapter three, a higher proportion of the Ethiopian runners travelled long distances to school each day by running compared with controls. It has previously been shown that Kenyan boys who walked and ran to school each day had a 30 % higher $\dot{V}o_{2\,max}$ than those who did not (Saltin *et al.*, 1995b). Such findings support suggestions that running long distances to school each day influences the success of east African runners, by developing their aerobic capacity.

Kenya is the model for the current study as it has an unparalleled record of success in international distance running competition (Larsen, 2003). Kenya has a population of approximately 30 million, distributed amongst eight provinces, and is peopled by three

main language groups: Bantu, Nilotic and Cushitic. The largest of these groups is the Bantu who account for approximately 65 % of the Kenyan population. Nilotic speakers account for approximately 30 % of the population and include the Kalenjin, who produce a disproportionately large number of elite Kenyan runners (Manners, 1997). The Kalenjin have a population of approximately three million, about 10 % of the Kenyan population, yet have won about 75 % of all major distance running races in Kenya (Manners, 1997). Internationally, Kalenjin runners have won 73 % of all Kenyan gold medals and a similar percentage of silver medals at major international running competitions. The Cushites account for part of the remaining 5 % of the total population (Pkalya & Aden, 2003). To our knowledge, no study has attempted to trace the ethnic or environmental background of elite Kenyan runners and, by doing so, examine the possibility that they might share a common ethnic or environmental origin and possibly some form of homogeneity in comparison with the general population. This study, therefore, aimed to determine the ethnic and environmental background of elite Kenyan runners. The findings were then compared to the general "non-athlete" Kenyan population to identify whether the athletes were of a distinct ethnic or environmental background.

4.2 Methods

4.2.1 Subjects

Subjects are as detailed in Chapter two. The athletes were classified into one of two groups according to athletic success: international athletes (I; $N = 97$), who had represented Kenya at international level in events such as the Olympic games, Commonwealth games, All-African games or in international marathon events, and national athletes (N; $N = 307$) who were active in national level competition in Kenya but without any major international distinctions. Athletes specialised in distances ranging from the 800 m to the marathon.

Questionnaires. Questionnaires were administered to all individuals participating in this study. The questions included in these were designed to obtain the following information:

1. Place of birth. This was classified according to the eight provinces of Kenya (Kenyan Central Bureau of Statistics, 2003). Our intention was to identify particular regions with a disproportionately high number of athletes in response to reports that the majority of the most successful Kenyan runners are from one altitudinous region of Kenya (i.e. Nandi District, Figure 4.1)(Manners, 1997). The

regional distribution of the control group was also compared with that of the Kenyan population (Kenyan Central Bureau of Statistics, 2003).

2. Spoken language (and that of their grandparents). This was to provide further information on ethnicity. A common language can be indicative of common origin, and a related language (i.e., a language of the same family) can also suggest a common origin, but one that is older. Currently, Kenya has 42 ethnic groups, who speak languages belonging to one of the following three language subsets: Bantu, Nilotic, or Cushitic. It is estimated that these language categories are spoken by approximately 65 %, 30 % and 2 % of the Kenyan population, respectively; the remaining small proportion being languages of European or other origin (Pkalya & Aden, 2003).

3. Distance travelled (< 5 km, 5-10 km, > 10 km) and mode of travel (walk, run, other) to school. We sought to find out how far the athletes travelled to school and the mode of transport they used. This was to assess the influence of running long distances to school each day on East African running success.

4.2.2 Data analysis

Contingency chi-square tests were performed to identify frequency differences between groups (Place of birth: d.f. = 10, language group: d.f. = 2, ethnicity: d.f. = 16, distance and method of travel to school: d.f. = 4). Pairwise chi-squares were then used to identify between which groups the differences lay (Place of birth: d.f. = 5, language: d.f. = 1, ethnicity: d.f. = 8, distance and method of travel to school: d.f. = 2). Statistical significance was declared at $P < 0.05$. In Figure 4.2, Nairobi, Coast, Western and North Eastern provinces that collectively accounted for less than 9 % of all participants, were collapsed into "Other" to allow statistical analysis. In Figure 4.4, the ethnic communities that accounted for less than 5 % of all participants were also collapsed into "Other".

4.3 Results

4.3.1 *Place of birth*

The regional distribution of the controls did not differ from that of the Kenyan population ($\chi^2 = 5.7$, $P = 0.23$) based on recent Census data (Kenyan Central Bureau of Statistics, 2003), but differed both from national ($\chi^2 = 86$, $P < 0.001$) and international ($\chi^2 = 84$, $P < 0.001$) runners (Figure 4.2). Both national and international groups displayed a marked over-representation of athletes from the Rift Valley province (C: 20 %, N: 65 %, I: 81 %). National athletes also differed from international athletes ($\chi^2 = 11$, $P = 0.022$) (Figure 4.2).

4.3.2 *Language*

The origin of language of control participants differed from national ($\chi^2 = 42$, $P < 0.001$) and international ($\chi^2 = 63$, $P < 0.001$) athletes with both athlete groups showing a predominance of languages of Nilotic origin compared with controls (C: 21 %, N: 60 %, I: 79 %) (Figure 4.3). National athletes also differed from international athletes ($\chi^2 = 12$, $P < 0.001$) (Figure 4.3).

4.3.3 Ethnicity

The language spoken by controls differed from national ($\chi^2 = 73$, $P < 0.001$) and international ($\chi^2 = 112$, $P < 0.001$) athletes (Figure 4.4). National athletes also differed in spoken language from international athletes ($\chi^2 = 34$, $P < 0.001$). Both athlete groups showed an over-representation of Kalenjin athletes (C: 8 %, N: 49 %, I: 75 %) especially from the Nandi sub-tribe (C: 5 %, N: 26 %, I: 44 %). Nandi is a district in Kenya accounting for less than 3 % of the Kenyan population (Kenyan Central Bureau of Statistics, 2003).

4.3.4 *Distance travelled to school*

In distance travelled to school, control participants differed from both national ($\chi^2 = 9$, $P = 0.01$) and international ($\chi^2 = 13$, $P = 0.002$) athletes (Figure 4.5). National athletes also differed from international athletes in distance travelled to school ($\chi^2 = 8$, $P = 0.02$). 75 % of controls travelled less than 5 km to school each day compared with 58 % of national athletes and 50 % of international athletes (Figure 4.5). Of the international athletes, 23 %

travelled further than 10 km to school each day, compared with 11 % of the national athletes and 9 % of the controls (Figure 4.5).

4.3.5 *Method of travel to school*

In method of travel to school, controls differed from national ($\chi^2 = 75$, $P < 0.001$) and international ($\chi^2 = 65$, $P < 0.001$) athletes; a higher proportion of athletes ran to school each day (C: 22 %, N: 73 %, I: 81 %) (Figure 4.6). National athletes did not differ from international athletes in method of travel to school ($\chi^2 = 3.1$, P = 0.22). Groups also differed in their method of travel to school when travelling equivalent distances. Of the subjects who travelled further than 5 km each day to school (Figure 4.5), only 23 % of the 22 controls ran to school, 23 % walked and 54 % used another form of transport. However, amongst the athletes who travelled further than 5 km each day, the overwhelming majority ran to school (77 % of the 130 national athletes and 86 % of the 49 international athletes; C Vs. N: $\chi^2 = 35.2, P < 0.001$, C Vs. I: $\chi^2 = 28.5, P < 0.001$). However, it is worth noting that 54 % of the controls who travelled further than 5 km to school each day used means of transport to school other than walking or running, compared with only 8 % and 6 % of national and international athletes, respectively. When participants using other means of transport such as a bicycle or motorised transport were excluded from all three groups, the differences in running to school between groups remained, with 50 % of controls ($N = 10$), 84 % of national ($N = 119$) and 91 % of international ($N = 46$) athletes running to school (C Vs. N: $\chi^2 = 7.10$, $P = 0.008$, C Vs. I: $\chi^2 = 10.39$, $P = 0.001$).

4.3.6 *Motivation to become a competitive athlete*

The motivation of Kenyan national and international athletes to become a competitive athlete was similar ($\chi^2 = 6.6$, $P = 0.36$) (Figure 4.7). Both national (39 %) and international (33 %) athletes declared economic reasons as their primary motivation to undertake athletic competition. Typically, Kenyan athletes see athletics as a means of making money to help their families, parents and siblings.

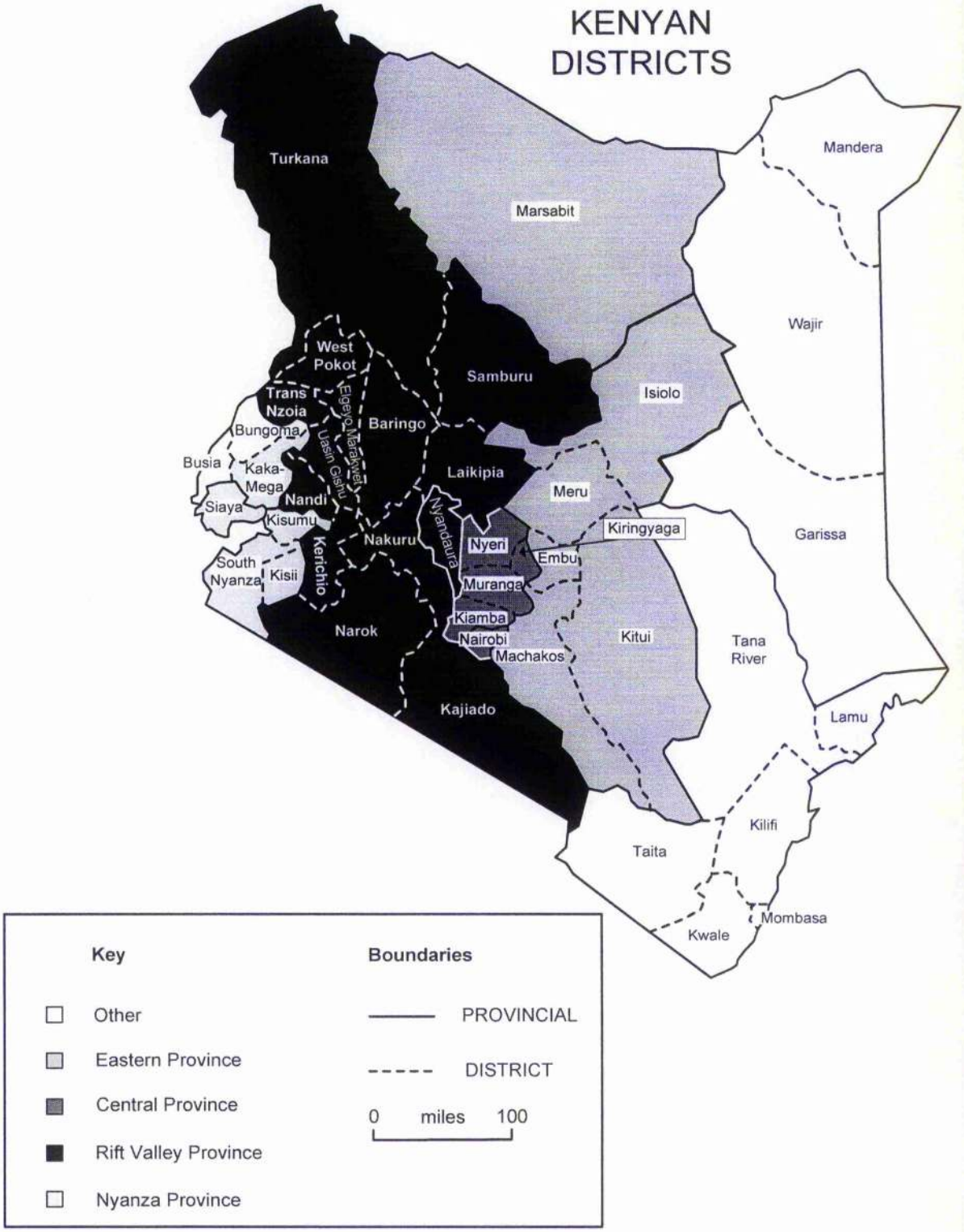


Figure 4.1: Districts of Kenya reflecting distribution of participants in the present study. "Other" region covers Coast, North-Eastern and Western Provinces.

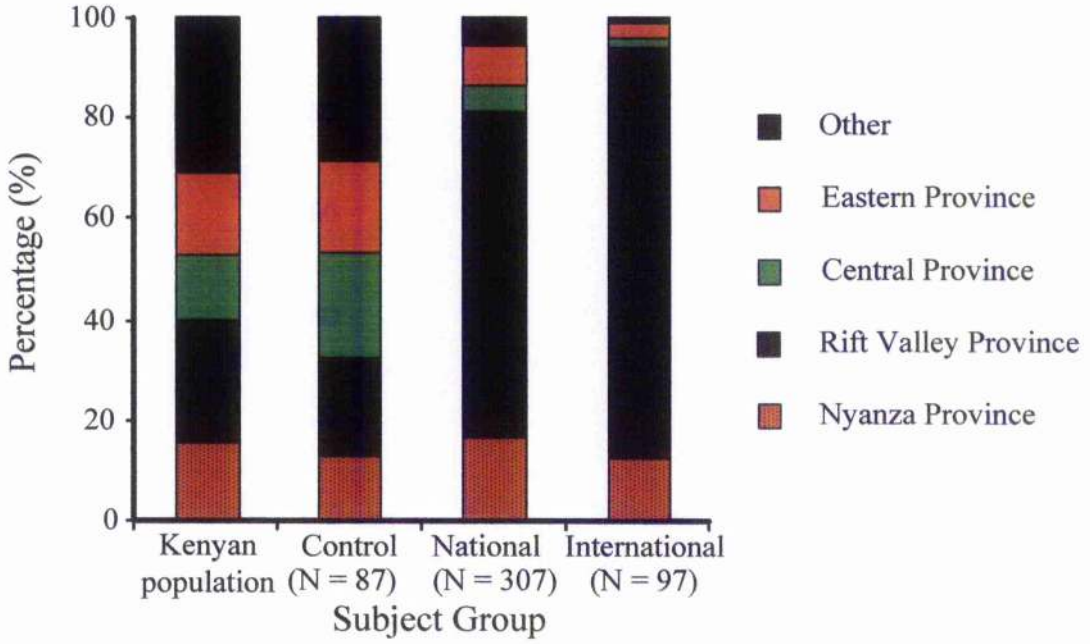


Figure 4.2: Regional distribution of subject groups and Kenyan population (K). Regional distribution of controls did not differ from the Kenyan population ($P = 0.23$), but differed from both national ($P < 0.001$) and international athletes ($P < 0.001$). National athletes also differed from international athletes ($P = 0.022$).

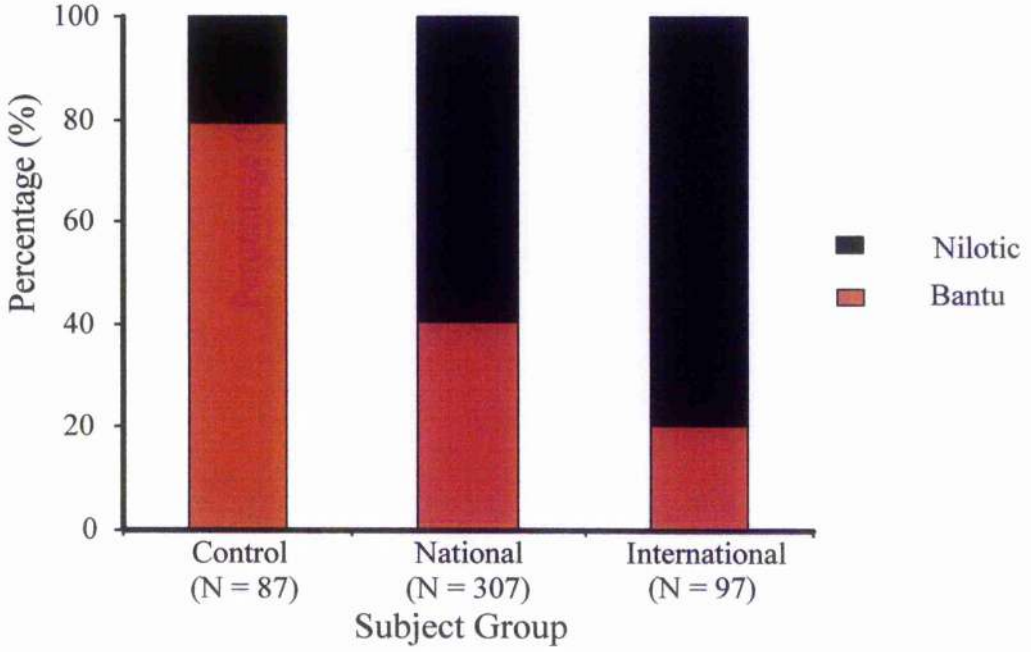


Figure 4.3: Language group distributions. Percentage of participants' grandparents speaking languages of each origin is shown. Language group distributions of all groups differed from each other (C Vs. N: $P < 0.001$, C Vs. I: $P < 0.001$, N Vs. I: $P < 0.001$).

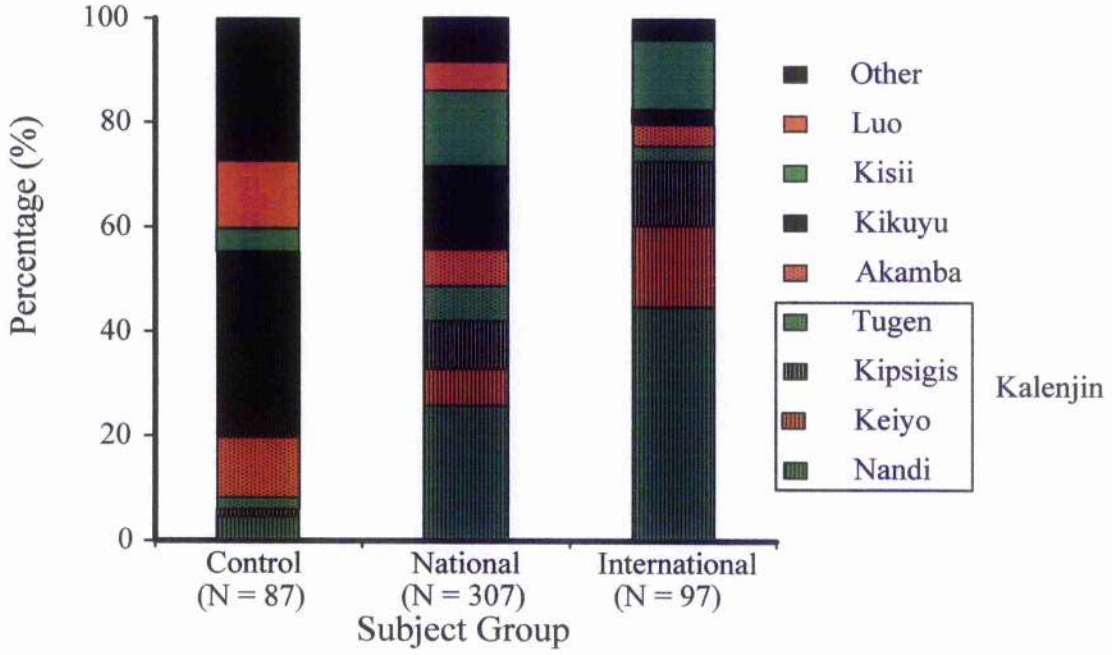


Figure 4.4: Ethnicity of participant groups. The proportion of participants' ethnicity is shown. Ethnicity distributions of all the groups differed from each other (C Vs. N: $P < 0.001$, C Vs. I: $P < 0.001$, N Vs. I: $P < 0.001$).

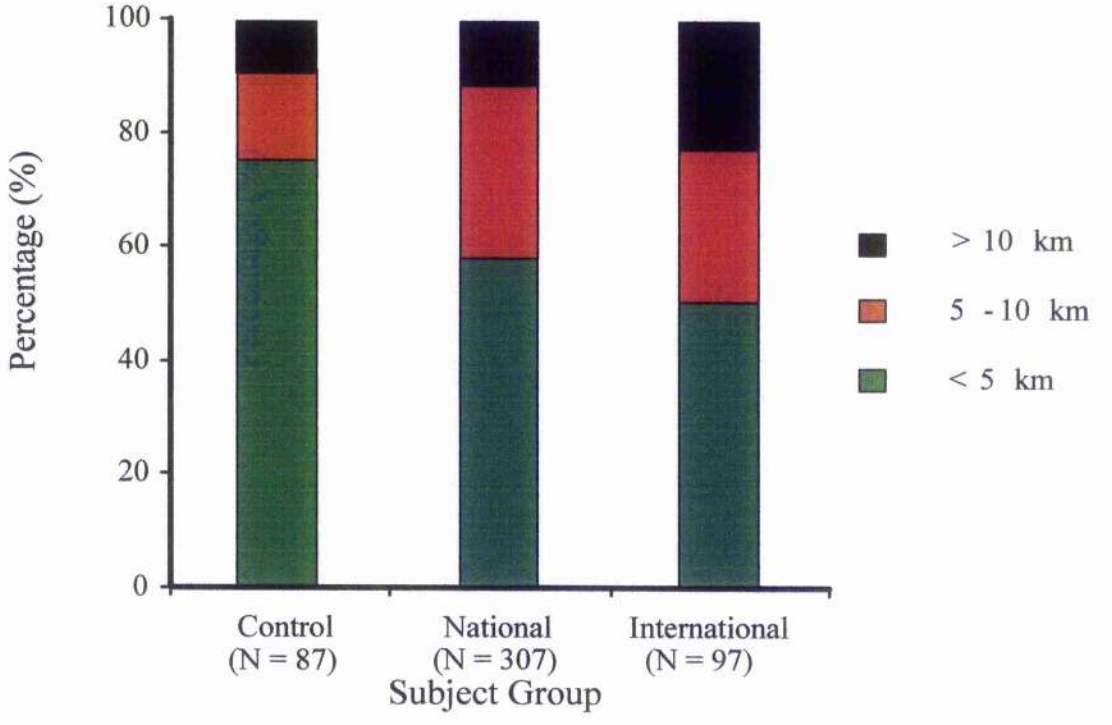


Figure 4.5: Distance travelled to school. The percentage of participants travelling each distance to school is shown. All groups differed from each other (C Vs. N: $P = 0.01$, C Vs. I: $P = 0.002$, N Vs. I: $P = 0.021$).

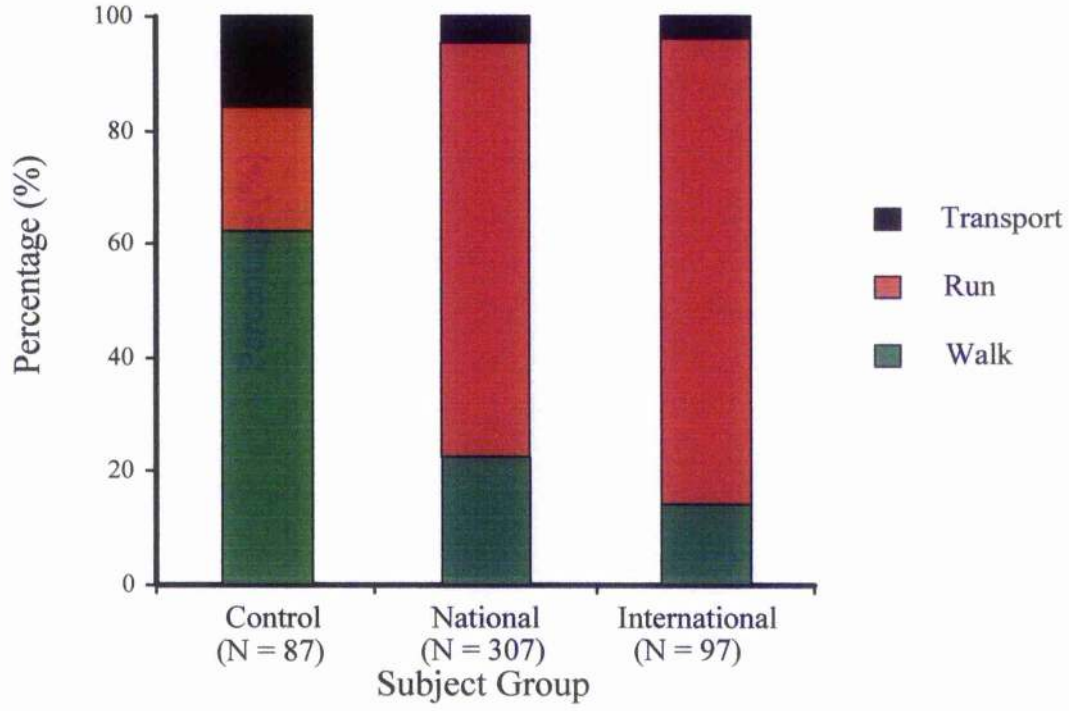


Figure 4.6: Method of travel to school. The percentage of participants using each method of travel to school is shown. Controls differed from both athlete groups (N: $P < 0.001$, I: $P < 0.001$). National and international athletes did not differ in their distribution ($P = 0.22$).

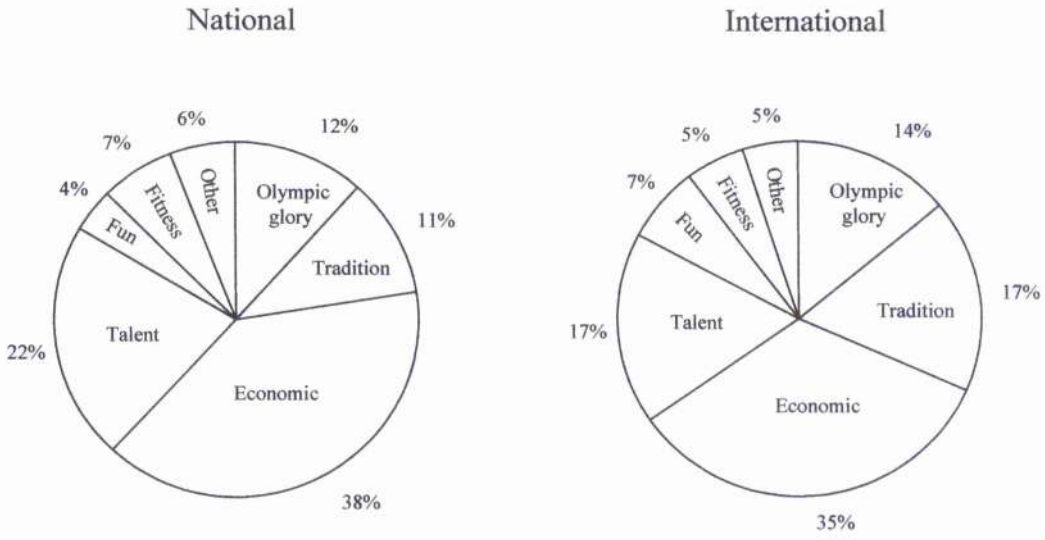


Figure 4.7: Motivation to become a competitive athlete. Both athlete groups showed similar reasons for becoming a competitive athlete ($P = 0.36$), with economic motivation most prevalent.

4.4 Discussion

The findings of this study are that Kenyan athletes, particularly the international athletes, are of a distinct environmental and ethnic background. Both the national and international athlete groups differed from the control group in all categories that were considered. Sixty-five percent of national athletes, and 81 % of International athletes, were from the Rift Valley province, in contrast to control participants, of whom only 20 % were from the Rift Valley province. Regarding the language group, there is a gradation of results from controls to national and international athletes: 79 % of international athletes speak languages of Nilotic origin compared with 60 % of national athletes and only 21 % of control participants. It was also found that a higher proportion of athletes travelled farther than 5 km to school than controls (C: 25 %, N: 42 %, $P = 0.01$; I: 51 %, $P = 0.002$) and a higher proportion of athletes travelled these distances by running (C: 22 %, N: 73 %, $P < 0.001$; I: 81 %, $P < 0.001$). These findings highlight the importance of environmental factors in the selection of elite Kenyan athletes.

As can be seen in Figure 4.2, the control group have a similar regional distribution to the Kenyan population. However, both athlete groups show a distinct distribution with an over-representation from the Rift Valley province. Of the international athletes, the most successful Kenyan athletes of past and present, 81 % were from this province, which accounts for less than a quarter of the Kenyan population (Kenyan Central Bureau of Statistics, 2003). The over-representation from the Rift Valley province mirrors previous findings in elite Ethiopian marathon athletes where 73 % were from Arsi or Shewa, regions proximal to the Rift Valley. Regions proximal to the Rift Valley often lie at altitudes of over 2,000 m above sea level, which might be haematologically beneficial in distance running (Schmidt *et al.*, 2002). These findings support suggestions that the success of Kenyan runners is in some way linked to their close proximity to the Rift Valley (Entine, 2001b).

A higher proportion of athletes, particularly international athletes, spoke languages of Nilotic origin compared with controls (C Vs. N: $P < 0.001$, C Vs. I: $P < 0.001$) (Figure 4.3). This is considered to be a reflection of the geographical distribution of the athletes and the peopling of the Rift Valley province, which is primarily populated by people speaking languages of Nilotic origin. In Ethiopian athletes, it was found that a higher proportion of marathon athletes spoke languages of Cushitic origin than controls ($P < 0.001$). In Kenya, languages of Cushitic origin account for less than 2 % of spoken

languages and, as such, no participants in this study spoke languages of Cushitic origin. When the ethnicity of athletes was compared with controls, there was an excess of athletes belonging to the Kalenjin tribe (Figure 4.4) (C: 8 %, N: 49 %, $P < 0.001$; I: 75 %, $P < 0.001$). This imbalance of elite Kenyan athletes has been attributed to cultural factors within the Kalenjin tribe, where the success of athletes such as Kipchoge Keino, the most famous Kenyan runner, is thought to have spurred on a tradition of distance running which is now augmented by talent recruitment to high schools with formal training and competition (Saltin *et al.*, 1995b). As can be seen in Figure 4.4, many of the national and international athletes belong to the Nandi sub-tribe (C: 5 %, N: 26 %, I: 44 %), yet the Nandi region accounts for only 3 % of the total Kenyan population. Again, cultural factors could be important here, as many of the most successful Kenyan athletes have belonged to the Nandi. The fame and fortune afforded to these runners could inspire young Nandi people to adopt distance running as a potential career.

In this study, a higher proportion of international (51 %, $P = 0.002$) and national athletes (42 %, $P = 0.01$) travelled farther than 5 km to school each day than controls (25 %) (Figure 4.5). Although mirrored in Ethiopia, these findings are in stark contrast to the sedentary lifestyles of young Scottish Children (Reilly *et al.*, 2004) and the increasing prevalence of childhood obesity in the United States and Great Britain (Ebbeling *et al.*, 2002). Athletes also differed from controls in their method of travel to school, with a higher proportion of athletes running to school each day (C: 22 %, N: 73 %, $P < 0.001$; I: 82 %, $P < 0.001$). When participants who had travelled equivalent distances (i.e. > 5 km) were compared between groups, 86 % of international athletes and 77 % of national athletes ran to school compared with only 23 % of controls (C Vs. N: $\chi^2 = 35.2$, $P < 0.001$, C Vs. I: $\chi^2 = 28.5$, $P < 0.001$). When participants using other means of transport, such as a bicycle or motorised transport, were excluded from all three groups, the differences in running to school between groups remained, with 50 % of controls ($N = 10$), 84 % of national and 91 % of international athletes running to school (C Vs. N: $\chi^2 = 7.10$, $P = 0.008$, C Vs. I: $\chi^2 = 10.39$, $P = 0.001$). Although based on only a few control participants with whom to compare the athletes, the finding that when travelling equivalent distances a higher proportion of athletes chose to run, raises the interesting question of whether these athletes were inherently more able to cover these large distances to and from school each day by running. Regardless of the reasons for running such long distances to school, this amount of exercise each day would undoubtedly lead to training adaptations for distance running.

From the results of the present study, a considerable proportion of national and international athletes were motivated to run for economic reasons (Figure 4.7). Thirty-nine percent of the national and 33 % of the international runners became athletes for economic empowerment. Given the levels of economic deprivation in Kenya, it is likely that this would act as an important contributing factor to the success of east African athletes in distance running, in strengthening their motivation.

4.4.1 Conclusions

The results of this study suggest that elite Kenyan athletes have a distinct ethnic and environmental background compared with the general Kenyan population. Athletes differed from controls in all variables considered. The majority of Kenyan runners were from the Rift Valley province, were Kalenjin and spoke languages of Nilotic origin. Athletes also travelled farther to school than controls and mainly did so by running. International athletes tended to show a pronounced version of this profile compared with national athletes. These results highlight the importance of such environmental factors in the success of elite Kenyan distance runners.

Chapter 5

Mitochondrial DNA Lineages of Elite Ethiopian Athletes

5.1 Introduction

Although no direct genetic effect has been found to account for the success of East African athletes, it is recognised that environmental factors alone may not influence athletic success. There is an increasing volume of support for the role of favourable genetics in the determination of athletic success (Wolfarth *et al.*, 2005). A number of nuclear genes have been proposed as being influential in the determination of athletic success. Polymorphisms of genes such as the Angiotensin Converting Enzyme (ACE) gene (Gayagay *et al.*, 1998;Myerson *et al.*, 1999), and the α-actinin-3 gene (ACTN3) (Yang *et al.*, 2003) have been shown, in some cases, to be over-represented in groups of elite athletes compared to non-athlete control populations. However, support for the role of such 'performance genes' in the determination of athletic performance has not been universal and some studies did not find a role for the ACE gene in elite endurance athlete status (Rankinen *et al.*, 2000a;Rankinen *et al.*, 2000b). Findings of a maternal effect in the inheritance of $\dot{V}o_{2\,max}$ (Lesage *et al.*, 1985;Bouchard *et al.*, 1999) hinted to a possible influence of mitochondrial DNA (mtDNA) in the determination of aerobic capacity. In addition to polymorphisms in the nuclear genome, some studies have suggested that polymorphisms in mtDNA may account for some of the inter-individual differences in endurance performance and response to endurance training (Dionne *et al.*, 1993;Murakami *et al.*, 2002) (for review see (Rupert, 2003)). Although the findings concerning the influence of mtDNA polymorphisms are equivocal, many studies have shown that mtDNA mutations are linked to various exercise intolerance pathologies (Wolfarth *et al.*, 2005), and evidence is growing of adaptive selection of particular mtDNA types in different geographic regions due to climate variation (Mishmar *et al.*, 2003;Ruiz-Pesini *et al.*, 2004).

mtDNA is a circular, double stranded DNA molecule of 16,569 bp which encodes 13 subunits of a number of enzyme complexes of oxidative phosphorylation, as well as components of the mitochondrial protein synthesis system (Anderson *et al.*, 1981). In addition, proteins that interact with the non-coding D-Loop region of the sequence regulate the replication of mtDNA and its transcription to mRNAs. mtDNA is highly mutable and is inherited in a matrilineal fashion, undergoing no recombination. This results in the accumulation of linked complexes of polymorphisms down different lines of descent from an ancestral mtDNA molecule. The branching pattern of descent can be used to trace the ancestry of individuals or populations (Richards *et al.*, 2000). Each of these branches, as shown in Figure 5.1, is referred to in the present study as a haplogroup. All present-day human mtDNA sequences can be traced back to an ancestral mtDNA that existed over

120,000 years ago (Ingman *et al.*, 2000). Since then, mtDNA sequences have accumulated between 40 and 70 mutations from the ancestral human mtDNA sequence (Maca-Meyer *et al.*, 2001), about 1 mutation per 2,500 years, or 100 human generations. The non-coding D-Loop region of mtDNA is exceptionally mutable, and the Hyper-Variable Sequence I (HVS-I) of the D-Loop is commonly used to assess mtDNA relatedness (Richards *et al.*, 2003).

It is known that there are a wide variety of mtDNA haplogroups in Ethiopia (Salas *et al.*, 2002), some of which are ancient African 'L' haplogroups that have remained in Africa (Chen *et al.*, 1995). There are also a proportion of sequences similar to those commonly found outside Africa (Richards *et al.*, 2003). These sequences are thought to be found in high numbers in Ethiopia probably as a result of substantial gene flow into Ethiopia from Eurasia (Richards *et al.*, 2003;Chen *et al.*, 2000). Regardless of the mechanism, the high divergence time of Ethiopian mtDNA means that there has been time for a number of haplogroup specific polymorphisms to occur, all of which have the potential to affect endurance performance through influences on oxidative phosphorylation (Dionne *et al.*, 1993;Murakami *et al.*, 2002). Given the lack of recombination, if mtDNA polymorphisms were important in the success of Ethiopian distance runners, selection for variants beneficial to exercise performance would lead to an increased frequency of the haplogroups on which the polymorphism occurred amongst elite athletes relative to the general population through selection. In addition, as some of the mtDNA haplogroups are more commonly found in individuals indigenous to Ethiopia, if any of these haplogroups contain beneficial variants, this may partially account for the success of Ethiopian athletes in international distance running. The present study, therefore, aimed to compare the mtDNA haplogroup distribution amongst elite Ethiopian athletes relative to the general Ethiopian population.

5.2 Methods

5.2.1 Subjects

Subjects are as described in chapter two, except that only elite endurance athletes were compared to control subjects. Subjects comprised 76 athletes (44 males, 32 females), all members (past and present) of the Ethiopian national athletics team, and 108 control subjects (C) (93 males, 15 females). The control group, students at Kotebe Teaching College, Addis Ababa, was intended to be representative of the general Ethiopian

population in their mtDNA distribution; none were regularly training for athletics, or had ever been successful at any level in distance running. Despite an excess of control subjects from Addis Ababa, the control group is well matched to the general Ethiopian population (from which the athletes were selected) in geographical distribution as reported in Census data (Tesfaghiorghis, 1986). The athletes comprised two groups: athletes specializing in 5,000- to 10,000 m distances (5-10km; $N = 42$), and marathon runners (M; $N = 34$), and were collectively grouped as Endurance athletes (E; $N = 76$). Endurance athletes were truly elite athletes, regularly successful in international distance running, and included past and present World and Olympic champions, and World Record holders.

5.2.2 Data analysis

Molecular methods are as detailed in Chapter two. All subjects were screened for known polymorphisms C3594T, A10398G, and C10400T, which cause restriction site changes -3592 *Hpa* I, +10394 *Dde* I, and -10397 *Alu* I respectively. Subjects with −3592 *Hpa* I were determined to belong to African haplogroup L. They were then assigned to haplogroup L1 or L2 depending on the presence of HVS-I polymorphisms specific to either haplogroup (Maca-Meyer *et al.*, 2001). Samples that digested as +10394 and +10397 with *Dde* I and *Alu* I respectively were assigned to haplogroup M (Chen *et al.*, 1995;Passarino *et al.*, 1998) and displayed a variety of the HVS-I motifs (Maca-Meyer *et al.*, 2001). Samples which digested as -3592 *Hpa* I, +10394 *Dde* I, and -10397 *Alu* I, and featured the HVS-I motif 16223, allowing for some recurrent transitions, were classified as L3A (Rando *et al.*, 1998). Samples with this digestion pattern but displaying HVS-I motifs specific to other, primarily Eurasian, haplogroups such as I, J, and K (Richards *et al.*, 2000), were classified into the group E1. Samples that digested as -3592 *Hpa* I, -10394 *Dde* I, and -10397 *Alu* I may belong to one of many haplogroups, such as HV1 or PreHV (Richards *et al.*, 2000), which, for the purposes of this study, were classified into the group E2.

Chi-square tests were performed to establish mtDNA haplogroup frequency differences between groups defined by characteristics such as athletic status, place of birth and language family (Table 5.1). Statistical significance was declared at $P < 0.05$. For comparisons between endurance athletes and controls for mtDNA haplogroup distribution, each haplogroup was compared against the sum of all other haplogroups to establish differences in frequency. Odds ratios (OR) with 95 % confidence intervals were applied to each of these analyses to establish the direction of haplogroup frequency differences between groups. Place of birth and language family categories were defined as described in Chapter three.

Table 5.1. Distribution of subject groups throughout mitochondrial tree.

Athlete status	L1	L2	L3A	M	E1	E2	Total	Unique sequences
Control	20	8	29	21	10	20	**108**	90
Endurance	12	11	19	11	6	17	**76**	70
Total	**32**	**19**	**48**	**32**	**16**	**37**	**184**	160
Number of unique sequences	27	16	43	27	12	28	153	
P- Value (Vs. total of all others)	0.63	0.12	0.78	0.38	0.75	0.52		
OR	0.83	2.12	0.91	0.70	0.84	1.27		
95 % CI	(0.38-1.81)	(0.81-5.54)	(0.46-1.78)	(0.32-1.56)	(0.29-2.42)	(0.61-2.62)		

5.3 Results

5.3.1 Athlete status

The distribution of human mtDNA is shown by a simplified phylogenetic tree (Figure 5.1), which shows the orientation of the mtDNA classifications used in this study. It can be seen that there are a wide variety of mtDNA haplogroups present in Ethiopia, which have an ancient divergence (Table 5.1). Previous studies have shown a high proportion of East African L and M types in Ethiopia (Rando *et al.*, 1998;Passarino *et al.*, 1998;Salas *et al.*, 2002), which the findings of this study support. The distribution of mitochondrial types amongst endurance athlete and control groups is shown graphically in Figure 5.1.

Both athlete and control groups displayed high levels of haplotype diversity (C = 0.99, E = 0.99). No significant association was found between endurance athlete status and mtDNA haplogroup (Table 5.1, Figure 5.1), as endurance athletes did not differ significantly from controls (χ^2 = 3.47, d.f. = 5, P = 0.63). As can be seen from Figure 5.1, both groups showed similar distribution of typically African 'L' haplogroups (C = 53 %, E = 55 %), and athletes were equally diverse in their mtDNA haplogroup distribution. Many of the athletes do not share a common maternal or mtDNA ancestor since the time of 'mitochondrial Eve' (Fig 5.1). Given the distinct ethnicity displayed by the 5-10 km and marathon athletes, they were also compared to controls as distinct groups. They displayed no distinct differences relative to controls (C Vs. 5-10 km, χ^2 = 5.81, P = 0.33; C Vs. M, χ^2 = 3.43, P = 0.64). When marathon runners (N = 34) were compared to 5-10 km runners (N = 42), no significant differences were found in their mtDNA haplogroup distribution (χ^2 = 6.34, 5 d.f., P = 0.27). Furthermore, when controls were compared to 5-10 km and marathon athletes as separate groups, no significant differences were present (C Vs. 5-10 km, χ^2 = 5.8 = 0.33; C Vs. M, χ^2 = 3.4, P = 0.64).

Comparisons between the frequency of each haplogroup versus the sum of all others in athletes and controls did not reveal any frequency differences between groups. 95 % confidence limits of all odds ratios included 1.0 (Table 5.1). P was greater than 0.5 in all cases, except for L2 and M comparisons, where P = 0.12 and P = 0.38, respectively (Table 5.1).

5.3.2 Place of birth and language family

The haplogroup distribution amongst all subjects (athletes and controls) from different geographical regions of Ethiopia is displayed in Figure 5.2. The mtDNA haplogroup distribution of each region is similar, with all regions displaying similar proportions of African 'L' haplogroups (Addis Ababa: 59 %, Arsi: 50 %, Shewa: 44 %, Other: 57 %). No association was found between regional origin of subjects and their mitochondrial haplogroup ($\chi^2 = 8.5$, d.f. = 15, $P = 0.9$). Similarly, the mtDNA haplogroup distribution of subjects (athletes and controls) speaking languages from each family is shown in Figure 5.3. Again there was no association between language family and mitochondrial type ($\chi^2 = 5.4$, 5 d.f., $P = 0.37$). As can be seen in Figure 5.3, the haplogroup distributions of each language family are again very similar.

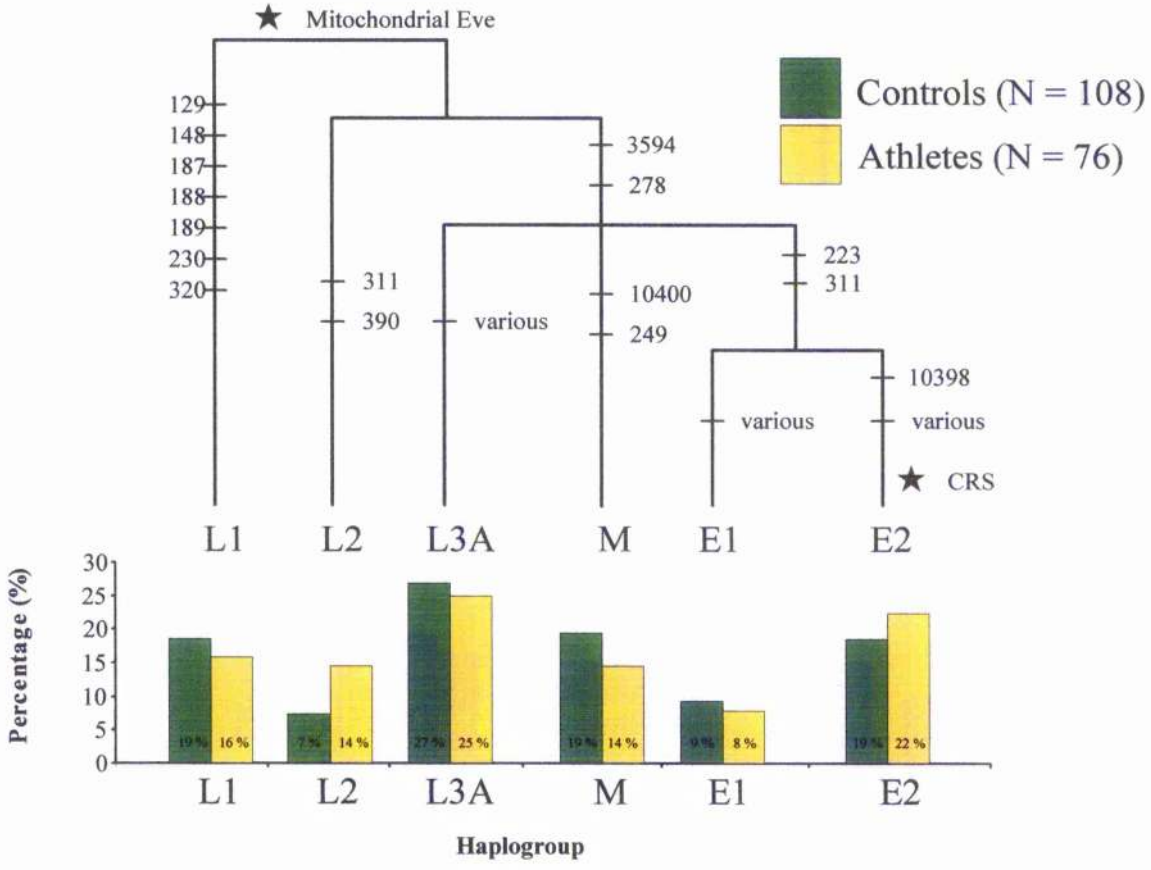


Figure 5.1. Human mitochondrial tree and percentage of each subject group in each haplogroup. Approximate positions of polymorphisms relative to the Cambridge Reference Sequence (CRS) (Anderson *et al.*, 1981) are shown (HVS-I polymorphisms are shown minus 16,000). Haplogroup topology is modelled upon more detailed human phylogenies (Macaulay *et al.*, 1999;Maca-Meyer *et al.*, 2001). Approximate position of the ancestral mtDNA sequence 'Mitochondrial Eve' is also shown. Percentages of each subject group are shown by the bars below the tree. No significant differences were present between groups. Haplogroup nomenclature has since amended so that haplogroup L1 is now split into haplogroups L0 and L1 (Kivisild et al., 2004), as in Chapter 6.

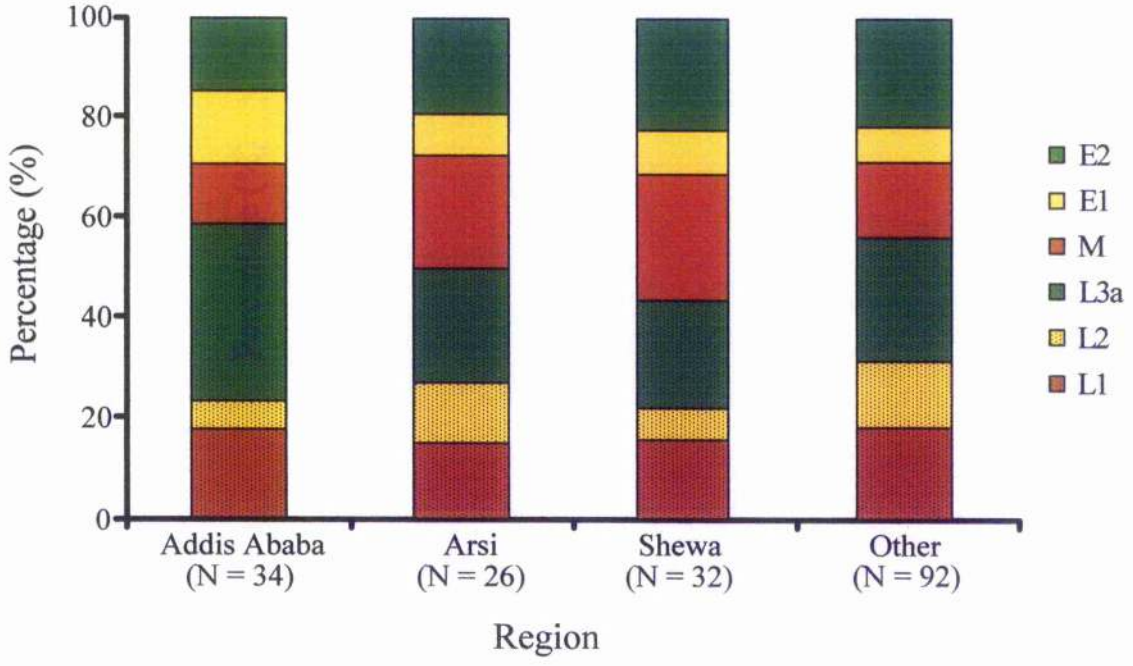


Figure 5.2. Haplogroup distribution of Ethiopian regions. The haplogroup distributions of all subjects (athletes and controls) are shown above.

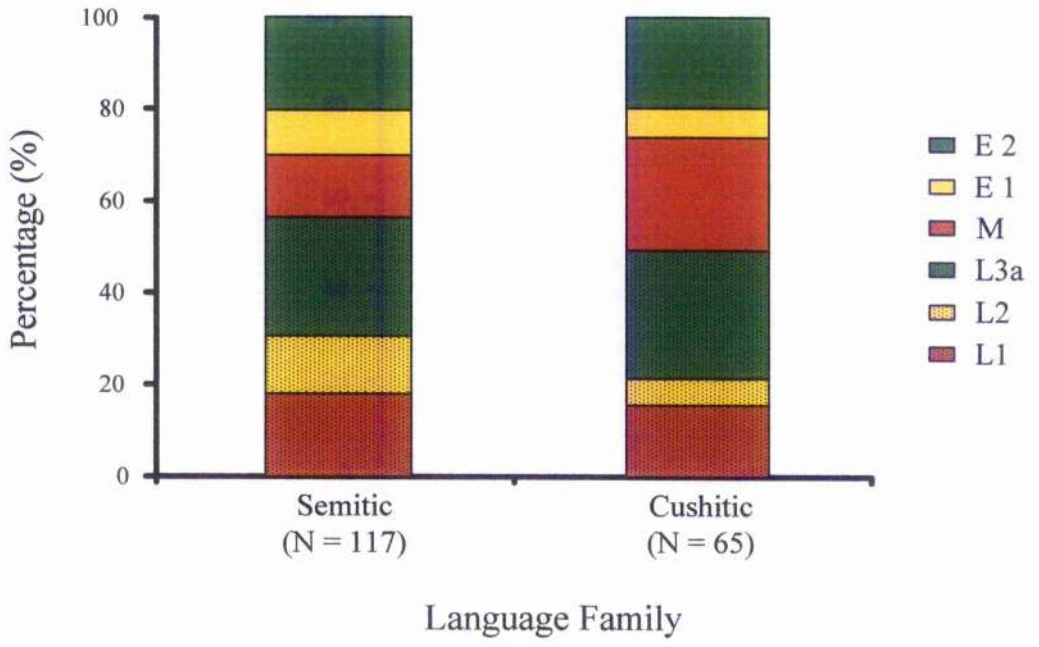


Figure 5.3. Haplogroup distribution of language families. The haplogroup distributions of all subjects (athletes and controls) are shown above, except wo subjects who spoke languages of Omotic origin, who were excluded from this analysis.

5.4 Discussion

Previous studies have suggested that mtDNA polymorphisms may affect endurance performance (see review by (Rupert, 2003)). Such influences on endurance performance have been suggested to be through polymorphisms in the coding region of mtDNA affecting mitochondrial function (Dionne *et al.*, 1993), or by polymorphisms in the non-coding region influencing the binding of transcription factors, potentially resulting in higher levels of mtDNA replication and subsequent increases in mitochondrial density (Murakami *et al.*, 2002). It was therefore of interest to establish if any mtDNA haplogroups were over-represented amongst elite Ethiopian athletes relative to the general population. The present study found that elite Ethiopian athletes display similar mitochondrial heterogeneity to the general Ethiopian population. The range of mtDNA haplogroups present in the elite athletes does not support the hypothesis that elite Ethiopian endurance athletes are from a genetically distinct group as classified by mtDNA.

Twenty six percent of control subjects, and a similar percentage of endurance athletes (30 %) belong to an L1 or L2 haplogroup (Figure 5.1) which are ancient African haplogroups that have diverged from other mitochondrial types c. 100,000 years ago (Chen *et al.*, 1995). These findings support previous reports showing a high prevalence of L1 and L2 types in East Africa (Passarino *et al.*, 1998;Salas *et al.*, 2002;Kivisild *et al.*, 2004). The finding that the elite endurance athletes are distributed in a similar fashion to the Ethiopian population does not support the hypothesis that mitochondrial polymorphisms are influential in the success of Ethiopian distance runners. Nor is it supportive of the belief that east African runners arise from a limited genetic isolate. The ancient divergence of L1 and L2 from other mitochondrial haplogroups, in conjunction with the absence of recombination in the matrilineal descent of mtDNA means that if mtDNA polymorphisms were influential in the success of Ethiopian distance runners, any polymorphism beneficial to endurance performance would need to have arisen over 100,000 years ago and therefore be present in almost all human populations, or have arisen repeatedly on separate occasions in each of the branches prevalent amongst the elite Ethiopian athletes. A previous study in Japanese subjects, found that certain HVS-I polymorphisms were over-represented in groups with higher baseline $\dot{V}o_{2\,max}$ and increased endurance trainability (Murakami *et al.*, 2002). However, the wide variation seen in the HVS-I region of the mtDNA D-loop of elite Ethiopian athletes is not consistent with the viewpoint that improved aerobic performance is due to polymorphisms in the HVS-I, as has been discussed in relation to Japanese subjects. Indeed, some of the polymorphisms suggested to influence endurance

trainability, such as 16298 in the Japanese cohort, were not present in any of the Ethiopian samples, athlete or control. The low numbers of identical HVS-I sequences (Table 5.2) further attests to the diversity of the athletes.

Previous findings, presented in Chapter three, have shown an over-representation of athletes from the Arsi region of Ethiopia. It may be that this geographical imbalance in the production of elite athletes reflects a differential distribution of mitochondrial types throughout Ethiopia, given that different geographical regions of Africa exhibit different distributions of mitochondrial haplogroups (Salas *et al.*, 2002). Similar over-representations are evident in Kenya, with many of the most successful Kenyan athletes originating from the Nandi tribe. Such observations have led to suggestions that there may be a genetic selection for variants accounting for improved endurance potential. However, the present study found that athletes from Arsi did not show a distinct distribution of mitochondrial types from control subjects originating from other geographical areas (Figure 5.2). As mtDNA is inherited in a maternal haploid manner and males do not pass on their mtDNA, for selection of a particular variant coding for improved aerobic capacity to occur, it would need to manifest itself through the conferral of maternal DNA. This is contrary to the stereotypical view that selection would be for the fittest male, who would not pass on any mtDNA to his offspring. However, gender-specific analysis finds that there is no association between mitochondrial type in either males or females. It seems that the distinct ethnicity of Ethiopian athletes relative to the Ethiopian population is independent of the influence of mtDNA haplogroup, either by genetic drift or selection on the basis of athletic success of a particular ethnic group.

Unlike nuclear genes, thousands of copies of the mitochondrial genome exist in each cell. The high mutation rate of mtDNA means that mutations may affect some but not all of these genomes. Therefore, different populations of mtDNA can exist in any given individual; a situation known as heteroplasmy. Heteroplasmic mutations account for most mtDNA-encoded pathologies (DiMauro & Schon, 2001), so may also account for mtDNA-encoded benefits to endurance performance. Often, mtDNA encoded pathologies are exhibited in tissues with a high energy demand, such as skeletal muscle. An example of this is the G7947A mutation, which has been associated with exercise intolerance and a high proportion (25 %) of ragged red skeletal muscle fibres (RRF) and a number of fibres deficient in Cytochrome Oxidase (80 %) (Grafakou *et al.*, 2003). Although mtDNA was not extracted from muscle cells, given the breadth of mitochondrial variation amongst the athletes, it remains unlikely that heteroplasmic polymorphisms are influential in the success of Ethiopian athletes, as such a scenario would still require that functional variants

have arisen independently in each of the lineages prevalent in the athlete groups or be present in many geographical populations. Recently, it has been suggested that even for homoplasmic mitochondrial mutations to induce a clinical phenotype, interaction with a nuclear modifier gene is necessary (Carelli *et al.*, 2003). The complex interaction between nuclear and mitochondrial genomes may also be a confounding factor in establishing a mitochondrial haplogroup that is over-represented in elite athletes due to functionality of the polymorphisms which it exhibits.

5.4.1 Conclusions

Although a number of studies have identified mtDNA polymorphisms as predetermining factors in exercise intolerance pathologies, the results regarding the influence of mtDNA polymorphisms and improved aerobic capacity are equivocal. The results of the present study suggest that mtDNA haplogroups prevalent in Ethiopia, and the variants they contain, are not a significant determinant of the exceptional success of Ethiopian athletes in international distance running.

Chapter 6

Mitochondrial DNA Lineages of Elite Kenyan Athletes

6.1 Introduction

As discussed in Chapter five, a number of studies have implicated mtDNA variation as being influential in the inter-individual differences in aerobic capacity, either indirectly (Lesage *et al.*, 1985;Bouchard *et al.*, 1999), or directly (Dionne *et al.*, 1993;Murakami *et al.*, 2002;Niemi & Majamaa, 2005). Encoding a number of proteins essential to OXPHOS, the mitochondrial genome is a potential candidate to contain variants influencing human performance. The matrilineal inheritance of mtDNA and linear accumulation of polymorphisms has allowed the construction of detailed mtDNA phylogenies (Maca-Meyer *et al.*, 2001). These phylogenies display the variation and diversity in human mtDNA and allow haplogroup identification through the analysis of a small number of haplogroup specific polymorphisms. Population movements and expansions have ensured that these haplogroups are often continent-specific, or present at widely differing frequencies in different populations. Haplogroups are groups of similar haplotypes, which only diverge when new mutations arise. This pattern of descent means that individual haplotypes share linked complexes of polymorphisms common to all sequences in a haplogroup. Associations between mtDNA haplogroups have been made with variations in risk of a number of pathologies such as Alzheimer's disease (Chagnon *et al.*, 1999) and Parkinson's disease (van der Walt *et al.*, 2003). In addition, a recent study by Niemi and Majamaa (2005) has shown that certain mtDNA haplogroups (J and K) are less common in endurance athletes, who rely heavily on effective OXPHOS during exercise, than in sprint athletes. However, as shown in Chapter five, there was no association between mtDNA haplogroup and elite Ethiopian athlete status.

The mtDNA variation in Kenya (Watson *et al.*, 1996;Brandstatter *et al.*, 2004) is less well characterised than that of Ethiopia (Kivisild *et al.*, 2004). There are known, however, to be a high number of African 'L' types (Brandstatter *et al.*, 2004), but a lower frequency of Eurasian haplogroups than that found in Ethiopia (Passarino *et al.*, 1998;Kivisild *et al.*, 2004). Each of these haplogroups has gained a number of unique polymorphisms, inherited in a matrilineal fashion which have the potential to influence mitochondrial function. The possibility exists, therefore, that haplogroups could predispose to variations in aerobic capacity. It has been suggested that mtDNA has undergone climatic selection for variants influencing efficiency of OXPHOS resulting in the presence of more efficient mitochondria in the tropics than in more polar climes (Mishmar *et al.*, 2003). Similar to Ethiopia, many of the haplogroups found commonly in Kenya, are found rarely outwith Africa. The present study, therefore, aimed to compare the frequency of mtDNA

haplogroups in elite Kenyan athletes to the general Kenyan population, and by doing so, establish the influence of mtDNA variation on elite Kenyan athlete status.

6.2 Methods

6.2.1 Subjects

Subjects are as detailed in Chapter two. 85 controls (40 males, 45 females) and 291 elite athletes were included in the study. Again, elite athletes were classified into national (N) (N = 221, 173 male, 48 female) and international (I) (N = 70, 59 male, 11 female).

6.2.2 Data analysis

Molecular methods are as detailed in Chapter two. All subjects were screened for known polymorphisms C3594T, A10398G, and C10400T, which cause restriction site changes -3592 *Hpa* I, +10394 *Dde* I, and -10397 *Alu* I, respectively. Subjects with +3592 *Hpa* I were determined to belong to African haplogroup L. They were then assigned to haplogroup L0, L1 or L2 depending on the presence of HVS-I polymorphisms specific to either haplogroup (Kivisild *et al.*, 2004;Salas *et al.*, 2004). Samples that digested as +10394 and +10397 with *Dde* I and *Alu* I, respectively, were assigned to haplogroup M (Chen *et al.*, 1995;Passarino *et al.*, 1998) and displayed a variety of the HVS-I motifs (Maca-Meyer *et al.*, 2001). Samples which digested as -3592 *Hpa* I, +10394 *Dde* I, and -10397 *Alu* I, and featured the HVS-I motif 16223, allowing for some recurrent transitions, were classified as L3A (Rando *et al.*, 1998). Subjects (N = 2) who displayed this digestion pattern but HVS-I motifs specific to other, primarily Eurasian, haplogroups such as I, and J (Richards *et al.*, 2000) were classified into the group E (the groups E1 and E2, as used in Chapter five, were not applicable in this analysis due to the lower subject numbers of these haplogroups). Samples that digested as -3592 *Hpa* I, -10394 *Dde* I, and -10397 *Alu* I may belong to one of many haplogroups (Pre-HV and HV1 in particular) (Richards *et al.*, 2000), which, for the purposes of this study, were classified into the group E. The classification of these haplogroups under the group E reflects their phylogenetic proximity (Maca-Meyer *et al.*, 2001) as shown by Figure 6.1.

Chi-squared tests were performed to establish mtDNA haplogroup frequency differences between groups defined by characteristics, such as athletic status, place of birth and language family. Statistical significance was declared at $P < 0.05$. Odds ratios (OR) with 95 % confidence intervals were applied to each of these analyses to establish the direction

of haplogroup frequency differences between groups. For comparisons between athletes and controls for mtDNA haplogroup distribution, each haplogroup was compared against the sum of all other haplogroups to establish differences in frequency. (d.f. = 2). Place of birth and language family categories were defined as described in Chapter four.

6.3 Results

The distribution of mtDNA haplogroups in Kenyan athletes relative to controls is shown graphically in Figure 6.1. It can be seen that the Kenyan population differs from Ethiopia in their mtDNA haplogroup frequencies, displaying a lower frequency of "Eurasian" haplogroups, and a subsequently higher frequency of L haplogroups.

6.3.1 Athlete status

All groups displayed similar high levels of haplotype diversity, based on HVS-I haplotypes (C = 0.99, N = 0.99, I = 0.99). However, when the mtDNA distribution was compared between subject groups, it was found that international athletes differed significantly in their distribution relative to controls (χ^2 = 12.6, 5 d.f., P = 0.027). National athletes did not differ significantly from controls (χ^2 = 8.7, 5 d.f., P = 0.12), or international athletes (χ^2 = 5.8, 5 d.f., P = 0.33). When each haplogroup was compared to the sum of all others, it was found that international athletes differed significantly from controls, showing an excess of L0 haplogroups (χ^2 = 4.8, 1 d.f., P = 0.028, OR = 2.37, 95 % CI = 1.09-5.2) (Figure 6.1). National athletes also differed significantly from controls when each haplogroup was compared to the sum of all others, showing an excess of M haplogroups (χ^2 = 4.5, 1 d.f., P = 0.034, OR = 4.36, 95 % CI = 1-19) (Figure 6.1).

6.3.2 Regional variation

It was recognised that the national and international athletes differ significantly from controls in their regional and ethnic distribution (Figures 4.2 and 4.3). For this reason, haplogroup distribution of subjects from the rift valley province (which produces a disproportionate number of elite athletes) was tested against the sum of all other regions. No significant differences in haplogroup frequency were found (L0: χ^2 = 2.3, P = 0.13. L1 χ^2 = 0.49, P = 0.49. L2: χ^2 = 1.1, P = 0.3. L3A: χ^2 = 1.6, P = 0.2. M χ^2 = 0.07, P = 0.8. E: χ^2 = 0.44, P = 0.51) (Figure 6.2).

6.3.3 Ethnic variation

Previous studies have shown that haplogroup frequencies vary between ethnic groups in Kenya (Watson *et al.*, 1997). Haplogroup differences between ethnic groups were tested as for regional variation. It was found that there were some significant differences between Bantu and Nilotic subjects when each haplogroup was tested against the sum of all others (L0: $\chi^2 = 1.64$, $P = 0.2$. L1 $\chi^2 = 0.09$, $P = 0.77$. L2: $\chi^2 = 0.09$, $P = 0.77$. L3A: $\chi^2 = 5.8$, $P = 0.016$, OR = 0.6, 95 % CI = 0.38-0.95. M: $\chi^2 = 4.6$, $P = 0.032$, OR = 2.4, 95 % CI = 0.96-6.02. E: $\chi^2 = 0.717$, $P = 0.4$) (Figure 6.3). As can be seen from Figure 4.3 in Chapter four, the international athletes showed an excess of Nilotic subjects relative to controls ($P = 0.001$). L3A haplogroups were under-represented in Nilotic subjects, but not in international athletes. However, Nilotic subjects showed an excess of M haplogroups relative to Bantu subjects. Similarly, M haplogroups were in excess in national athletes and tended toward an excess in international athletes (C Vs. N: $P = 0.034$, C Vs. I: $P = 0.082$).

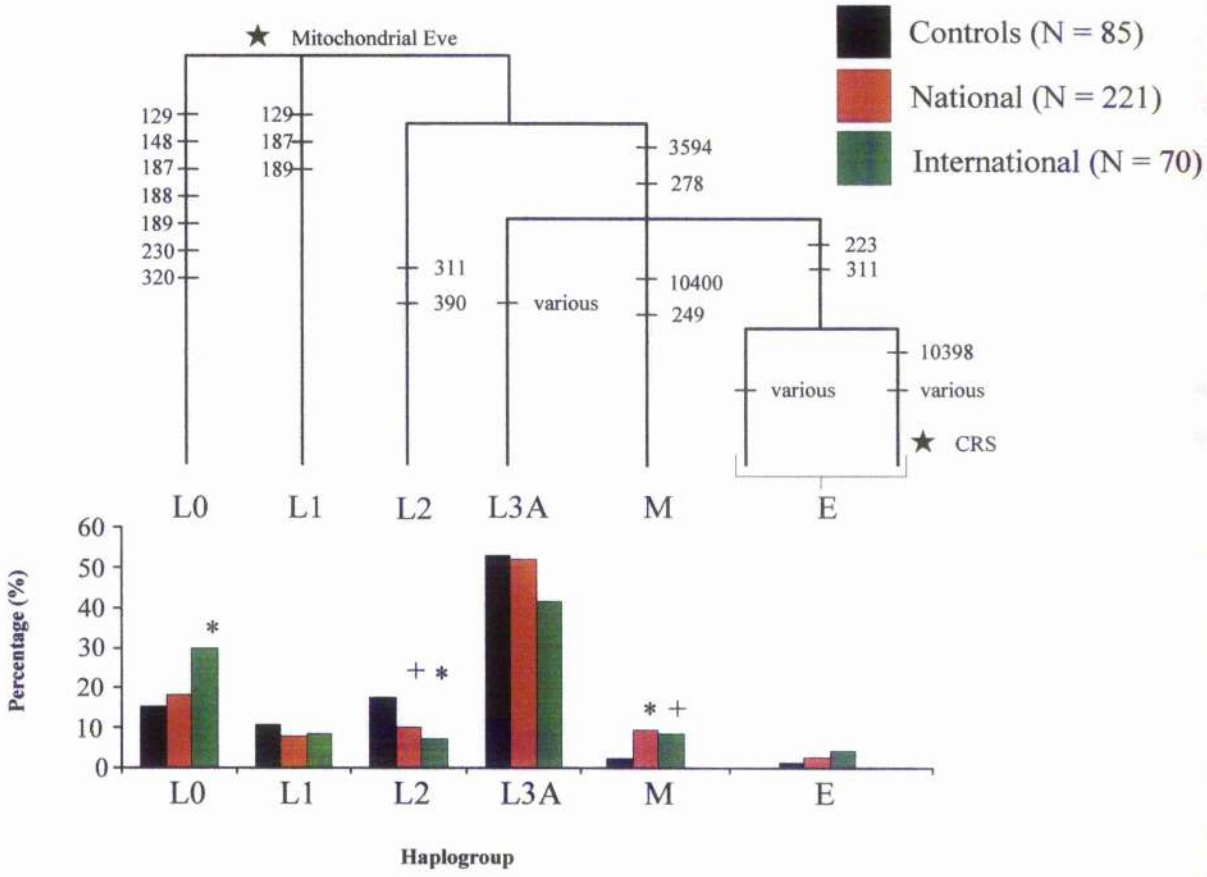


Figure 6.1 Human mitochondrial tree and haplogroup percentages. Approximate positions of polymorphisms relative to the Cambridge Reference Sequence (CRS) (Anderson *et al.*, 1981) are shown (HVS-I polymorphisms are shown minus 16,000). Haplogroup topology is modelled upon more detailed human phylogenies (Maca-Meyer *et al.*, 2001;Macaulay *et al.*, 1999; Kivisild *et al.*, 2004). Approximate positions of the ancestral mtDNA sequence 'Mitochondrial Eve', and the CRS are also shown. Percentages of each subject group are shown by the bars below the tree. Significant differences ($P < 0.05$) from controls are marked by * and tendencies toward differences are shown by +. Haplogroups L0 and L1 were previously classified as L1 in the mtDNA nomenclature, as in Chapter 5.

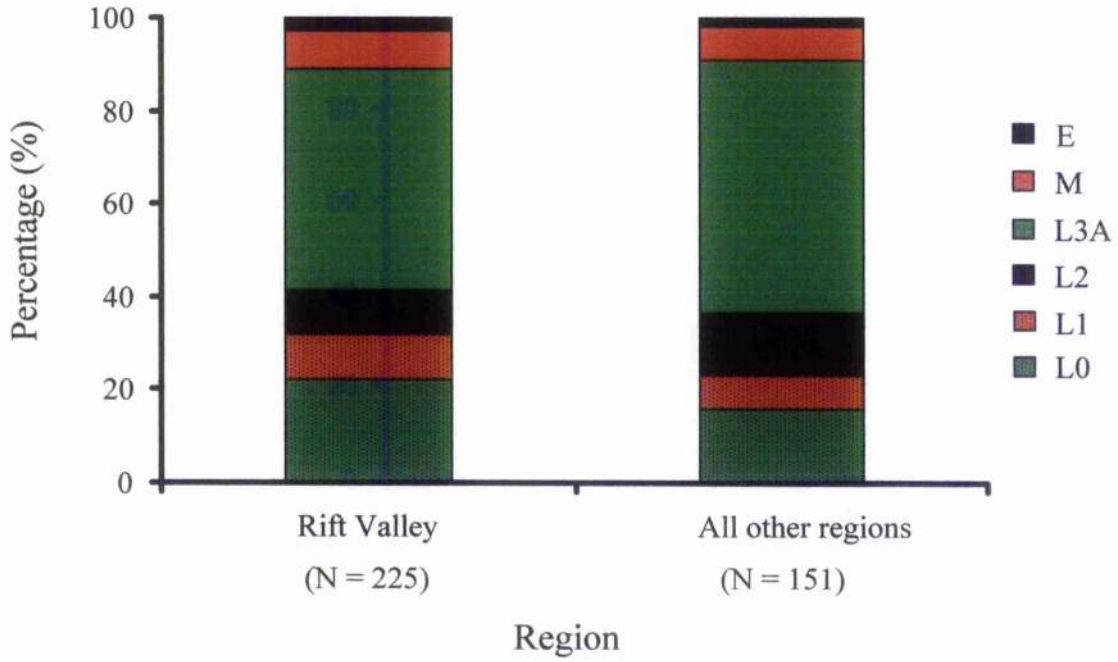


Figure 6.2 Haplogroup distribution of Rift Valley Vs Other regions. The percentages of Rift Valley controls versus all other regions belonging to each haplogroup is shown. No significant differences were present between groups.

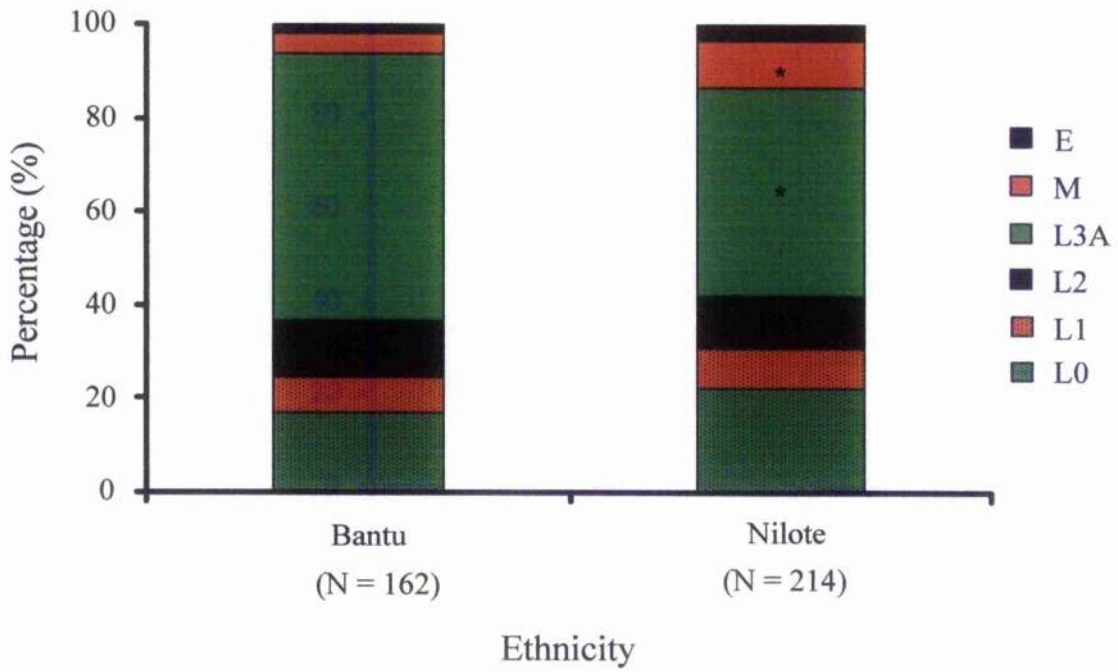


Figure 6.3 Ethnic distribution of haplogroups. The haplogroup distribution of controls belonging to each control group is shown. Differences from Bantu subjects are shown by * (M: $\chi^2 = 5.8$, $P = 0.016$, L3A: $\chi^2 = 4.6$, $P = 0.032$).

6.4 Discussion

The results of the present study suggest an association of mtDNA haplogroup with elite Kenyan athlete status. International athletes showed a significantly different distribution of mtDNA types relative to controls. They displayed an excess of L0 haplogroups (C: 15 %, N: 18 %, I: 30 %), and a tendency toward an excess of M haplogroups (C: 2 %, N: 10 %, I: 9 %). They also showed a lower frequency of L2 haplogroups than controls (C: 18 %, N: 10 %, I: 7 %). National athletes also showed differences from controls when each haplogroup was compared to the sum of all others. They exhibited an excess of M haplogroups and a tendency towards a lower frequency of L2 haplogroups, which supports the possibility that mtDNA variation in Kenya may be associated with elite athlete status. mtDNA haplogroups in Kenya have previously been shown to differ in frequency between different ethnic groups (Watson *et al.*, 1997). Nilotic subjects exhibited an excess of M haplogroups relative to Bantu subjects. As can be seen in Figure 4.3, national and international athletes exhibited an excess of Nilotic subjects relative to controls ($P < 0.001$).

6.4.1 Haplogroup association

The association of mtDNA haplogroups (L0 and M) with an increased frequency of being an elite Kenyan athlete may suggest that these haplogroups contain polymorphisms which influence some aspect of endurance performance or its trainability. However, the assigning of mtDNA haplogroups does not identify which variants may be causal to this association. Based on full genome sequencing, human mtDNA sequences have accrued upward of 40 mutations since their divergence from mitochondrial Eve (Maca-Meyer *et al.*, 2001). Each of these mutations has the potential to impact on aerobic capacity through subtle influences on mitochondrial function. However, even within a haplogroup, there is the potential for variation at the sub-haplogroup level. It would therefore be useful to look at the haplogroups associated with elite athlete status at higher resolution to establish any sub-haplogroup differences between athletes and controls.

It may seem intuitive that any association found in elite Kenyan athletes would be found in Ethiopian athletes. However, as seen in Chapter five, there was no association between mtDNA haplogroup distribution and elite athlete status in Ethiopia. Rather than high penetrance mutations, most of the associations with mtDNA haplogroups and disease phenotypes have been subtle and have often not been replicated in multiple populations.

For review, see Herrnstadt and Howell (2004). Often, the same pathogenic mutation can be associated with a number of different abnormalities; raising the possibility that mtDNA background and the presence of other polymorphisms can alter the penetrance of a given mutation. In addition, it has been shown that for the presence of a clinical phenotype, interaction with a nuclear modifier gene is necessary (Carelli *et al.*, 2003). It is not surprising, therefore, that any effect is not replicated. However, higher resolution analysis is now necessary to establish which polymorphisms may be influential in the association.

mtDNA encodes a number of proteins essential to mitochondria. Mitochondrial biogenesis is a complex process involving co-ordinated activation of the nuclear and mitochondrial genomes, and is one of the key adaptations to endurance training, resulting in an increased volume and density of mitochondria (Holloszy & Coyle, 1984). The molecular processes underlying this process are only beginning to be elucidated (Adhihetty *et al.*, 2003). It is possible that polymorphisms in mtDNA are influential in the adaptive response of mitochondria to endurance training, by modifying mitochondrial biogenesis. As well as encoding a number of essential proteins, recent studies have suggested that mtDNA may act as a sensing mechanism eliciting mitochondrial biogenesis through a decrease in mtDNA content after intense exercise (Marcuello *et al.*, 2005), elucidating another mechanism whereby mtDNA polymorphisms may influence aerobic capacity or its adaptive response to training. Although mtDNA was previously thought to be largely selectively neutral, it has been suggested that there has been climatic selection of mtDNA lineages dependent on their efficiency (Mishmar *et al.*, 2003;Ruiz-Pesini *et al.*, 2004). As well as a key role in ATP production to fuel exercise, OXPHOS has another key role: thermogenesis. The balance between these two functions is determined by the coupling of OXPHOS. Tightly coupled OXPHOS elicits efficient ATP production with lower thermogenesis. It has been suggested that this would be beneficial in the tropics, but loosely coupled mitochondria would be more useful in more Northern Climes. This would mean inefficient ATP production but an increased level of thermogenesis. In line with these suggestions, the basal metabolic rate of indigenous near-polar populations is higher than that of tropical populations (Leonard *et al.*, 2002). It is plausible, therefore, that the differences in haplogroup frequency reflect subtle influences of these haplogroups on OXPHOS efficiency. Based on this hypothesis, elite endurance athletes would require tightly coupled OXPHOS resulting in efficient ATP production. A side effect of this tight coupling is a higher production of deleterious reactive oxygen species, which may play a role in ageing (Huang & Manton, 2004). This hypothesis is supported by the finding that mice with loosely coupled OXPHOS live longer (Speakman *et al.*, 2004).

6.4.2 Population stratification

Despite the potential physiological link between mitochondrial DNA and aerobic capacity, there remains the possibility that the findings of the present study reflect the influence of sub-population variation in haplogroup frequency between ethnic groups of Kenya. Population stratification occurs when genetic frequencies differ between groups due to unknown differences in ancestral gene frequency (Tsai *et al.*, 2005). This can cause false positive or negative results due to association of any phenotype with a genotype subject to population stratification. That is to say that if the frequency of a particular haplogroup is high in the rift valley population, it will be associated with elite athlete status due to the fact that many elite athletes originate form this area. Although not all of the haplogroup associations found in Kenyan athletes were mirrored in haplogroup distribution of geographically or ethnically divided subjects, there was a higher frequency of M haplogroups in Nilotic subjects ($P = 0.032$) (Figure 6.3). As was shown in Chapter four, an excess of national and international athletes are of Nilotic origin ($P < 0.001$). The fact that an excess of M haplogroups was found in national athletes ($P = 0.034$), and a tendency toward an excess was found in international athletes ($P = 0.082$) suggests that this finding may have been influenced by population stratification. However, associations with haplogroups L0 and L2 and athlete status were not replicated in tests for population stratification.

6.4.3 Conclusions

The findings of the present study are that international standard Kenyan athletes display an excess of L0 haplogroups, and a lower frequency of L2 haplogroups. National standard athletes sustained these findings, displaying an excess of M haplogroups and a tendency toward a dearth of L2 haplogroups. Athletes display a distinct geographical and ethnic heritage relative to the general Kenyan population, which may mask an influence of population stratification on gene associations. Indeed, subjects of Nilotic origin displayed an excess of M haplogroups, which may suggest that the association with elite athlete status is as a result of population stratification. However, not all associations with elite athlete status were reflected when subjects were tested by regional or geographical classification. mtDNA remains a likely candidate for human performance and the current associations warrant further investigation at higher resolution.

Chapter 7

Y Chromosome Distribution of Elite Ethiopian Athletes

7.1 Introduction

As discussed, east African athletes have enjoyed unparalleled success in distance running. Female East African runners are now becoming successful, but historically, it has been the males who have had most success. Given the regional clustering of elite east African athletes, as evidenced in Chapters three and four, anecdotal proposals have also suggested that this is due to athletes arising from some form of genetic isolate (Manners, 1997;Entine, 2001b). It is suggested that periods of genetic isolation have led to genetic selection amongst these particular populations for variants conferring benefits in exercise performance. However, as discussed, there is no evidence to this effect and the levels of diversity in mtDNA data (Chapters five and six) do not support such a hypothesis.

In a similar way that mtDNA offers a snapshot of human history through the female lineage, the Y chromosome offers a similar view of the male lineage. Although some of the Y chromosome undergoes recombination with the X chromosome during meiosis, the majority (>90 %) does not, and is known as the non-recombining Y (NRY). This has allowed the construction of detailed phylogenies, which allow haplogroups to be defined through analysis of haplogroup specific polymorphisms. An example of a Y chromosome phylogeny is shown in Figure 7.1. The Y chromosome is relatively gene poor, containing an estimated 156 genes (Skaletsky *et al.*, 2003) over 23Mb. The human gene map does not currently contain any entries on the Y chromosome that are believed to be influential in human fitness-related phenotypes (Wolfarth *et al.*, 2005). Despite being non-essential for survival, shown by the presence of females who do not have a Y chromosome, Y chromosome haplogroups have been associated with a number of diverse phenotypes such as infertility, prostate cancer, and height (Reviewed by Jobling and Tyler-Smith (2003)). In principle, therefore, the Y chromosome may contain genes that influence endurance performance. Given that the success of east African runners is primarily, although not exclusively, a male phenomenon, it is an interesting candidate to study as it may in principle contain gene variants that influence performance in males. The haploid nature of inheritance of the Y chromosome also means that it is a useful indicator of individual or population ancestry.

The present study, therefore, aimed to compare the Y chromosome haplogroup frequencies between elite Ethiopian athletes relative to the Ethiopian general population and establish if any haplogroup was associated with elite endurance athlete status. It further aimed to

substantiate claims that elite east African athletes arise from a limited genetic subset by assessing the Y chromosome haplogroups present in elite athletes relative to controls.

7.2 Methods

7.2.1 Subjects

A total of 216 male subjects provided written informed consent to participate in the study. Subjects included 44 elite endurance athletes (E) consisting of 23 5-10km (5-10 km) athletes and 21 marathon athletes (M). Controls comprised one group of students at college in Addis Ababa (C; $N = 93$), (Chapter 3), and an additional control group from the Arsi province (A; $N = 79$), which is known to produce a disproportionate number of elite athletes (Figure 3.1). Subject status was as described in Chapter 2.

7.2.2 Data analysis

Molecular methods were as detailed in Chapter 2. Y-chromosomal binary markers typed to allow haplogroup identification are shown on the phylogenetic tree in Figure 7.1. Haplogroups were defined according to the Y chromosome consortium nomenclature (Y Chromosome Consortium, 2002).

Y chromosome haplogroup frequency differences between groups were tested by exact tests of population differentiation (Raymond & Rousset, 1995) using Arlequin version 2.0.1.1 (Schneider *et al.*, 2000). Subsequent exact tests were performed; comparing each haplogroup versus the sum of all others, to establish between which groups the differences lay. Odds ratios (OR) with 95 % confidence intervals were applied to each of these analyses to establish the direction of haplogroup frequency differences between groups. Haplogroup frequency differences between groups defined by regional and ethnic background were also examined by exact tests. Statistical significance was declared at $P < 0.05$.

7.3 Results

The distribution of Ethiopian subjects: athletes and controls, is shown in Figure 7.1. It can be seen that they are widely distributed throughout the human Y chromosome phylogeny,

and are similar to those previously characterised in Ethiopia (Semino *et al.*, 2002;Underhill *et al.*, 2000).

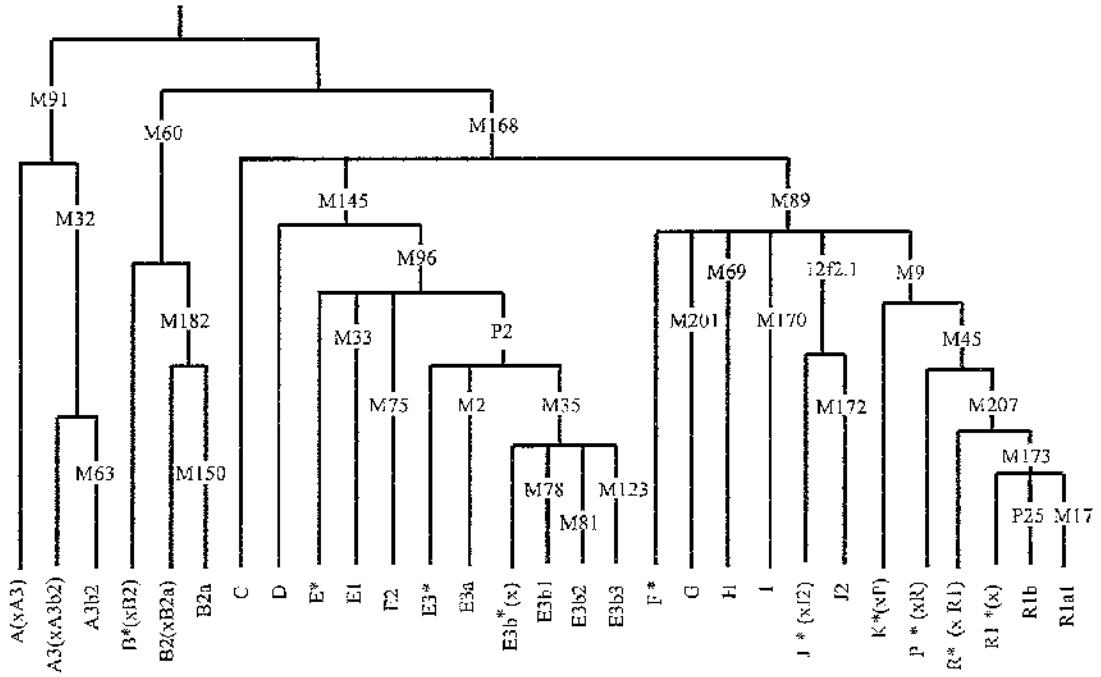
7.3.1 Athlete Status

Y chromosome haplogroup distribution was compared between each of the three subject groups: C, A, and E. It was found that E differed significantly from C ($P = 0.04$) and A ($P = 0.014$). A did not differ significantly from C ($P = 0.83$). Subsequent analyses of each haplogroup compared to the sum of all others between subject groups showing differences revealed that endurance athletes differed significantly from C ($P = 0.03$) and A ($P = 0.045$), exhibiting an excess of E* haplogroups. E* haplogroups were only present in endurance athletes so no OR could be calculated. E also differed significantly from C ($P = 0.028$, OR = 0.25, 95 % CI = 0.07-0.89) and A ($P = 0.014$, OR = 0.22, 95 % CI = 0.96-0.77), exhibiting a lower frequency of E3b1 (C = 23 %, A = 25 %, E = 7 %). Although it did not reach significance, E showed a tendency toward an excess of K(xP) haplogroups relative to Arsi controls ($P = 0.062$, OR = 7.8, 95 % CI = 0.84-72.12). No other significant differences were present ($P > 0.15$).

Given that the marathon athletes displayed a distinct ethnic profile to the 5-10 km athletes. The groups of athletes were analysed as discrete groups in comparison to control populations. It was found that C differed significantly from 5-10 km athletes ($P = 0.04$), but not from M ($P = 0.29$). Arsi controls differed significantly from both 5-10km ($P = 0.04$) and M ($P = 0.018$). Nor did marathon athletes differ significantly in their Y chromosome haplogroup distribution from 5-10 km athletes ($P = 0.45$). Subsequent analyses of each haplogroup compared to the sum of all others between subject groups showing differences revealed that 5-10 km athletes differed significantly from C ($P = 0.035$), exhibiting an excess of E* haplogroups. Although it did not reach statistical significance, 5-10km athletes also showed a tendency toward a higher frequency of E* haplogroups than the Arsi control group ($P = 0.052$). 5-10km also differed significantly from C (0.018, OR = 6.18, 95 % CI = 1.51-25.29), exhibiting an excess of E3* haplogroups, and a lower frequency of E3b1 haplogroups relative to the Arsi control group ($P = 0.037$, OR = 0.13, 95 % CI = 0.02-1.06) (C = 23 %, A = 25 %, 5-10km = 4 %, M = 10 %). C did not differ significantly from 5-10km athletes in E3b1 haplogroups but showed a tendency toward a lower frequency ($P = 0.08$, OR = 0.16, 95 % CI = 0.02-1.23). Marathon athletes showed an excess of K(xP) haplogroups relative to Arsi controls ($P = 0.029$, OR = 2.56, 95 % CI = 1.28-132.35). No other significant differences were present between the groups ($P > 0.14$).

7.3.2 Regional and Ethnic distribution of Y Chromosome Haplogroups

It has been shown that the Ethiopian athlete cohorts show a distinct distribution from controls in their regional and ethnic affiliations (Figures 3.2 and 3.3). It is plausible that Y chromosome haplogroup frequencies differ between regional or ethnic groups in Ethiopia. For this reason, Y chromosome haplogroup distribution was tested between regional and ethnic subject groupings to assess the possibility that the frequency differences between athletes and controls groups were as a result of their regional or cultural status rather than a biological effect. It was found that regional groups did not differ significantly in their haplogroup frequencies (Addis Ababa Vs. Arsi, $P = 0.072$; Addis Ababa Vs. Shewa P= 0.93; Addis Ababa Vs. Other $P = 0.18$; Arsi Vs. Shewa $P = 0.94$; Arsi Vs. Other $P = 0.1$; Shewa Vs. Other $P - 0.38$) (Figure 7.2). Haplogroup frequencies did not differ between Semitic and Cushitic subjects ($P = 0.14$) (Figure 7.3).



Control (C) (93)	-	2	16	-	0	2	-	-	**0**	-	1	**4**	-	15	**23**	-	6	-	-	-	-	22	4	**4**	-	-	-	-	-
Arsi Control (A) (79)	-	0	18	-	0	0	-	-	**0**	-	0	**9**	-	15	**25**	-	5	-	-	-	-	23	4	1	-	-	-	-	-
5-10 km (23)	-	0	17	-	0	0	-	-	**9**	-	0	**22**	-	9	**4**	-	4	-	-	-	-	26	4	**4**	-	-	-	-	-
Marathon (M) (21)	-	0	14	-	0	0	-	-	**5**	-	0	**0**	-	5	**10**	-	10	-	-	-	-	38	5	14	-	-	-	-	-
Endurance (E) (44)	-	0	15	-	0	0	-	-	7	-	0	**11**	-	7	7	-	7	-	-	-	-	32	5	9	-	-	-	-	-

Figure 7.1 Y chromosome distribution of Ethiopian athletes and control subjects. The percentage of each subject group belonging to each haplogroup is shown. Haplogroups which differed in frequency between athletes and controls are shown in bold.

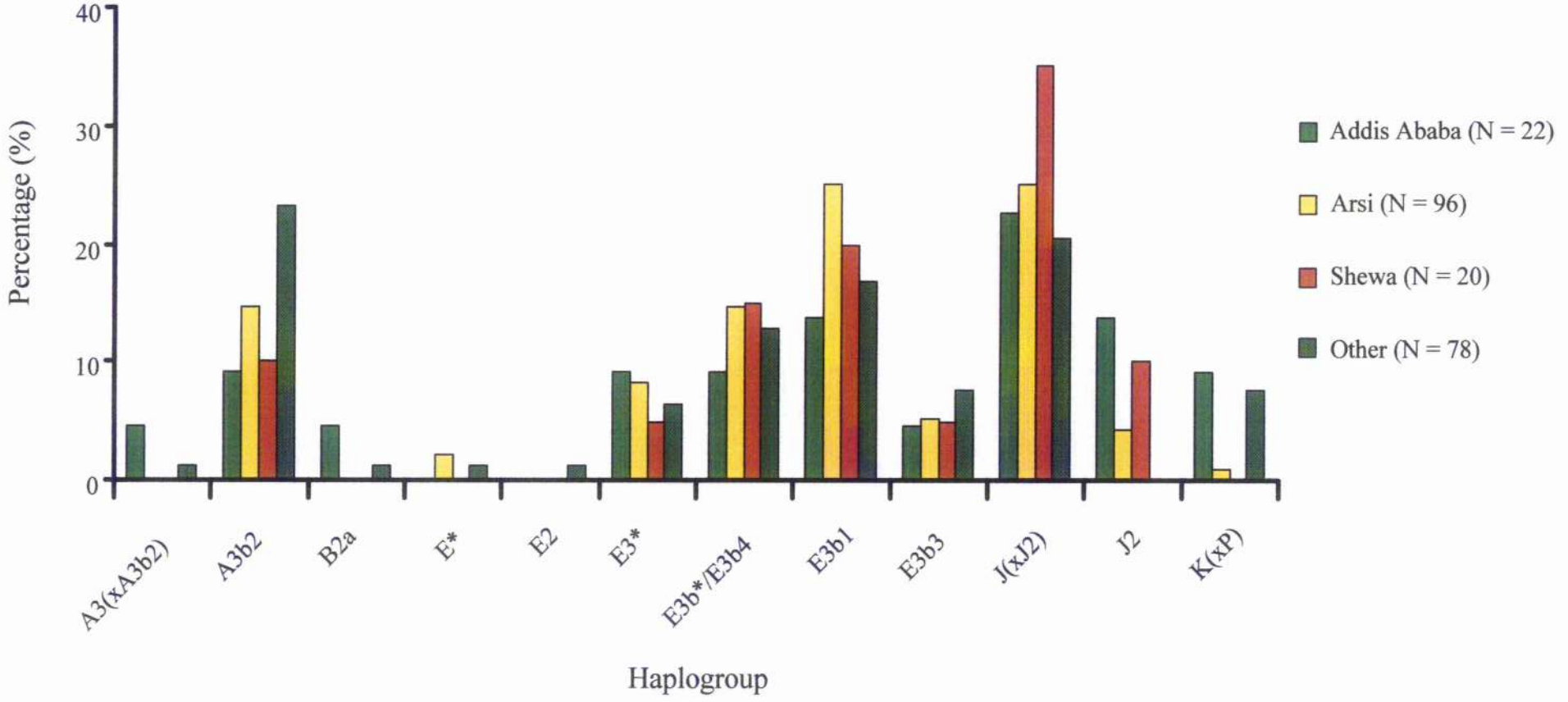


Figure 7.2 Y chromosome haplogroup distribution of Ethiopian regions. Percentages of subjects from each region are shown. No significant differences were present between regions.

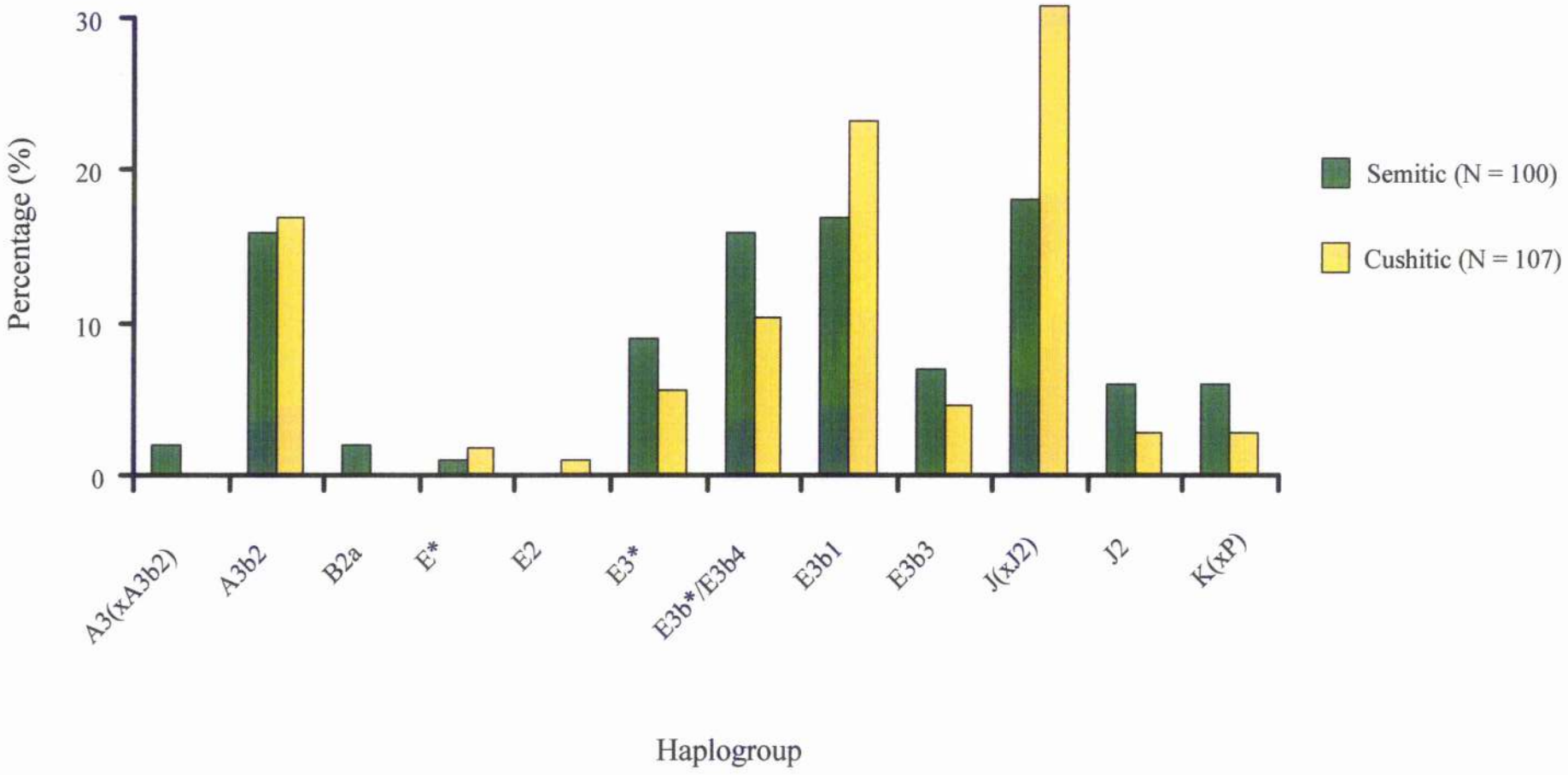


Figure 7.3 Y chromosome haplogroup distribution of ethnic groups. Percentages of subjects belonging to each haplogroup are shown. No significant differences were present ($P = 0.14$).

7.4 Discussion

The results of the current chapter show differences in Y chromosome haplogroup distribution between elite athlete and control groups (Figure 7.1). Endurance athletes combined (E) showed an excess of E* haplogroups relative to C ($P = 0.03$) and A ($P = 0.045$), and a tendency toward an excess of K(xP) haplogroups relative to the Arsi controls ($P = 0.06$). E also display a lower frequency of E3b1 haplogroups relative to C ($P = 0.028$) and A ($P = 0.014$). When the athletes were compared to controls as separate groups, the source of the above differences was elucidated. 5-10 km athletes displayed an excess of E* haplogroups relative to controls ($P = 0.035$), and a tendency toward an excess relative to A ($P = 0.52$), neither of which were replicated in the marathon athletes ($P > 0.19$). 5-10 km also displayed an excess of E3* haplogroups relative to controls ($P = 0.018$), and a dearth of E3b1 haplogroups relative to A ($P = 0.029$). Marathon athletes displayed an excess of K(xP) haplogroups relative to Arsi controls ($P = 0.029$). These results show that Y chromosome haplogroups are associated with elite athlete haplogroup status in Ethiopia. However, given the discrepancy of results between 5-10 km and marathon athletes, it is unclear if this is a biological effect or if some other phenomenon is influencing the association.

7.4.1 Y chromosome haplogroup association with athlete status

Although the human gene map for fitness related phenotypes now contains in excess of 150 entries; there are no entries on the Y chromosome (Wolfarth *et al.*, 2005). However, as mentioned in Chapter one, the study of the genetics of physical performance is at a very early stage and this is not conclusive evidence that the Y-chromosome does not have a biological influence on physical performance in males. Variation in the Y chromosome has been associated with a number of diverse phenotypes such as infertility (Kuroki *et al.*, 1999), blood pressure (Charchar *et al.*, 2002), and height (Ellis *et al.*, 2001). One of the propositions for the success of east African runners is that they display a distinct somatotype that is beneficial to distance running performance (Larsen, 2003). Elite Kenyan runners have shown improved running economy relative to Scandinavian runners (Saltin *et al.*, 1995b). It has been proposed that their relative long leg length and slender build contribute to their superior running economy (Larsen *et al.*, 2004). Existing associations of the Y chromosome and stature (Ellis *et al.*, 2001) highlight one example of how the Y chromosome may influence elite performance, through influencing anatomical phenotypes. As can be seen from Figure 7.1, the excess of E3* haplogroups in 5-10 km athletes relative

to controls is not replicated in marathon athletes, none of whom have an E3* haplogroup (Figure 7.1). In fact, none of the differences from controls are replicated in both 5-10 km and marathon athlete groups, which does not appear to support a biological influence on performance. Although both comparisons do not reach statistical significance, the frequencies in both 5-10 km and marathon athletes of haplogroups E* and E3b1 go in the same direction relative to the control groups (Figure 7.1). This adds some credence to the possibility of a biological influence of these haplogroups on elite athlete status. However, the frequencies of haplogroups E3* and K(xP) are in opposite directions relative to controls. For example, the haplogroup E3* is present at a frequency of 23 % in 5-10 km athletes relative to 8 % and 4 % in the Arsi and general control group, respectively. However, this haplogroup is not found in any of the marathon athletes. It is difficult to conceive, given the similarity of these disciplines and the underlying physiological requirements for success, how a Y chromosome haplogroup could confer an advantage, or disadvantage, in one discipline that would not be transferred to the other. If the associations could be replicated in another population such as the Kenyan athletes, it would add credence to a biological influence on the variation in physical performance. However, it is recognized that a failure to replicate association studies does not necessarily falsify those showing positive associations (Ellis & Harrap, 2003).

7.4.2 Population stratification

Gene frequency differences in ancestral populations can lead to genotype frequency differences between groups of differing ancestral background (Tsai *et al.*, 2005). Given the distinct ethnic and environmental distribution of the elite athletes relative to controls (as shown in Chapter 3), the possibility that these associations were a reflection of this, rather than a biological effect was tested. The 5-10 km and marathon athletes also differed from each other in their ethnic origin (Chapter 3), which may partially explain the discrepant findings between the groups. However, it was found that there was no association between Y chromosome haplogroup and either regional origin ($P = 0.7$) or ethnicity ($P = 0.14$). However, influences of population stratification cannot be ruled out. It may be that the athletes arise from distinct populations within these regions. If there were something special about of the genetics of the Arsi population that influenced their success, it may have been expected that the Arsi control group (A) would have differed in their haplogroup distribution from the general control group (C). It may also have been considered that more differences would have been found between the athletes and general controls than from the Arsi control group, given that more of the athletes are from Arsi. No differences were

found between the control groups in their haplogroup distribution ($P = 0.8$). However, more differences were present between the athletes and the Arsi control group. This raises the question of whether that control group is reflective of the Arsi region. For example, the Arsi samples were taken at a human rights rally in the region, but perhaps those at the rally were from a different social group than the general population of the region and that from which the athletes originate.

7.4.3 Conclusions

Although elite Ethiopian athletes display a wide variety of Y chromosome haplogroups spanning most of the human phylogeny (Y Chromosome Consortium, 2002). This does not support the anecdotal belief that elite Ethiopian athletes arise from a limited genetic isolate (Manners, 1997;Entine, 2001b). However, both athlete groups displayed a distinct Y chromosome distribution from control subjects. Haplogroups E*, E3*, and K(xP) appear to confer an increased potential to become an endurance athlete, whereas haplogroup E3b1 is negatively associated with elite athlete status. Although regional or ethnic variation within Ethiopia did not associate with Y chromosome haplogroup distribution, an influence of population stratification cannot be ruled out. Although the Y chromosome has not previously been associated with athletic performance and has no obvious candidate genes identified to date; the possibility that these haplogroup associations with athletic performance are as a result of a biological effect cannot be ruled out. The replication of these associations in another athletic population such as Kenya would add credibility to a biological influence of the Y chromosome on elite athlete status.

Chapter 8

ACE Gene Variation in Elite East African Runners

8.1 Introduction

Early familial studies suggested that there was a genetic component to endurance performance, through findings of high heritability in physiological indices of performance such as $\dot{V}o_{2\,max}$ (Lesage *et al.*, 1985;Fagard *et al.*, 1991). More recently, advances in molecular technology have allowed a number of potential candidate genes for endurance performance to be identified. A yearly review of candidate genes for performance-related phenotypes has now identified in excess of 150 gene variants which may contribute to the wide variation in human physical performance (Wolfarth *et al.*, 2005).

Perhaps the most studied of the potential 'performance genes' is the Angiotensin Converting Enzyme (ACE) gene. Angiotensin converting enzyme influences blood pressure through the generation of the vasoconstrictor Angiotensin II, and inactivation of the vasodilator bradykinin. The most studied of the ACE polymorphisms is the I/D polymorphism, characterized by the presence (I, Insertion) or absence (D, Deletion) of a 287 bp fragment in an Alu sequence in intron 16 of the ACE gene. The ACE I/D polymorphism has been associated with a wide variety of physical fitness phenotypes. Although associations with physical performance have been far from unequivocal (Jones *et al.*, 2002); in general, the ACE D allele has been associated with strength-and power-related performance (Folland *et al.*, 2000;Nazarov *et al.*, 2001;Woods *et al.*, 2001), and the ACE I allele with endurance performance (Hagberg *et al.*, 2001;Jones *et al.*, 2002). For example, a study of Australian rowers found a higher frequency of the I allele relative to controls (Gayagay *et al.*, 1998), while there was an over-representation of the I allele in British Olympic standard runners competing in distances of over 5,000 m (Myerson *et al.*, 1999). More recently, a study of South African triathlon competitors has shown an excess of the I allele in the fastest 100 finishers relative to controls (Collins *et al.*, 2004). The I allele has also been associated with an increase in type I muscle fibres (Zhang *et al.*, 2003), and the II genotype with higher $\dot{V}o_{2\,max}$ in studies of post-menopausal women (Hagberg *et al.*, 1998;Hagberg *et al.*, 2002). Not all studies, however, support an association between the I allele and endurance performance; some have found no association with endurance athlete status amongst groups of mixed sporting discipline and race (Rankinen *et al.*, 2000b). Others however, have even shown the D allele to be associated with training-related gains in $\dot{V}o_{2\,max}$ or with $\dot{V}o_{2\,max}$ itself (Rankinen *et al.*, 2000a;Zhao *et al.*, 2003). Such contrasting results have been attributed to small subject numbers (Rankinen *et al.*, 2000a) and to the inadequate stratification of subjects (Woods *et al.*, 2000). Given the disparity in results to date, it is clear that any association may be found only in

homogenous cohorts of athletes stratified by their endurance status and perhaps by their level of performance.

The I/D polymorphism has been estimated to explain up to 47 % of the variance in circulating ACE levels in a Caucasian population (Rigat *et al.*, 1990), with the I allele being associated with lower plasma and tissue ACE levels than the D allele (Alvarez *et al.*, 2000;Danser *et al.*, 1995;Rigat *et al.*, 1990;Woods *et al.*, 2004). Given that the I/D polymorphism is an intronic Alu sequence, a direct influence on circulating ACE levels is not readily apparent. The I/D polymorphism is therefore considered to be a genetic marker in linkage disequilibrium with a functional variant influencing ACE levels. The association between I/D and ACE levels has been described in Caucasians, but more recent studies in African populations have shown that other variants of the ACE gene are more closely associated with circulating ACE levels (Cox *et al.*, 2002;Zhu *et al.*, 2000;Zhu *et al.*, 2001). A transition at np 22982, in the sequence AF118569 as in Rieder et al.(1999), or 31958 as in Cox et al (2002), has been found to show the largest phenotypic differences between genotypes (Zhu *et al.*, 2000). Although all markers tested in Europeans showed significant differences in ACE levels between each genotype (probably due to linkage disequilibrium between genotypes), the most marked difference in ACE levels in both Afro-Caribbean and European subjects was found between genotypes at A22982G. Absolute linkage disequilibrium between I/D and 22982 has been shown for Caucasian populations (Soubrier *et al.*, 2002) and strong haplotype association has been shown in African populations (Cox *et al.*, 2002;Zhu *et al.*, 2000). The I allele at I/D has been shown to be in linkage disequilibrium with the A allele at A22982G and the D with the G, respectively (Soubrier *et al.*, 2002). Consequently, the A allele has been associated with lower circulating ACE levels than the G allele (Cox *et al.*, 2002;Soubrier *et al.*, 2002), except in one study where the A allele was associated with higher ACE levels (Zhu *et al.*, 2000). The variant at 22982 has been suggested to be a potential functional variant due to the proximity to a splice site (Zhu *et al.*, 2000), which may be influential in the production of alternative splice forms (Sugimura *et al.*, 1998). Although ACE genotype has been shown to have an influence on circulating and tissue ACE levels, it is unclear if this is the mechanism responsible for any association of the ACE genotype and performance.

The aim of the present study, therefore, was to describe the association between the ACE genotype at I/D and 22982 and elite east African endurance athlete status. In addition, the relationship between ACE genotype at these loci and plasma ACE activity was assessed in the Kenyan control group. If the influence of the ACE I/D polymorphism on endurance performance is due to the effects on plasma ACE levels, it is likely that the 22982

polymorphism would also be associated, and perhaps more strongly, with elite endurance athlete status in Africans.

8.2 Methods

8.2.1 Subjects

Subjects were as described in Chapter two. 291 elite Kenyan endurance athletes (232 male, 59 female) and 85 control subjects (40 male, 45 female) were included in the present study. 70 of the athletes (59 male, 11 female) had competed internationally representing Kenya (I, $N = 70$). Other athletes, classified as National (N), had competed at national level within Kenya ($N = 221$, 173 male, 48 female). All athletes had competed in distances from 3,000 m to marathon, where the energy source is predominantly aerobic (Gastin, 2001). As some previous studies have identified associations with genotype and performance at the highest performance levels (Nazarov *et al.*, 2001;Yang *et al.*, 2003), international athletes were considered separate from national athletes to ensure that any association was not masked by inadequate classification of athlete status.

76 Ethiopian elite athletes and 109 general Ethiopian controls (94 male, 15 female) were analysed for ACE genotype. In addition, 92 controls (79 male, 13 female) from the region of Arsi (a region producing a disproportionate number of elite athletes) were also genotyped. Elite athletes were separated into 5-10 km ($N = 42$; 23 male, 19 female) and marathon athletes ($N = 34$; 21 male, 13 female).

96 Caucasian subjects of European origin were also genotyped at both I/D and 22982 to test for levels of linkage disequilibrium between the markers in Europeans.

8.2.2 Genotype determination and molecular techniques

Techniques for genotype determination at the ACE I/D polymorphism and A22982G were as detailed in Chapter two. All controls also provided a supine 10 ml venous blood sample. Blood samples were drawn into iced plain tubes, from which plasma was drawn. Samples were centrifuged immediately at 10,000 rpm for 10 min and the supernatant stored at –80°C. Samples were then sent in dry ice to Germany, where the following ACE assay was performed. ACE activity was assayed by a modified fluorometric method using carbobenzoxy-phenyl-alanyl-histidyl-leucine (Z-Phe-His-Leu) as substrate. Plasma (50 µl) was diluted with cold phosphate buffer (70 mM, pH 8.0 containing 300 mM sodium

chloride) to a final volume of 450 µl. The enzyme reaction was started by adding 50 µl of a 10 mM Z-Phe-His-Leu solution to the samples and incubated at 37°C. At the end of the incubation time (30 min) the reaction was terminated by transferring 100 µl aliquots from the incubation mixture into 1 ml of 0.1 M NaOH. All subsequent steps in the assay were continued in the dark, beginning with the addition of 25 µl of 2%ortho-phthaldialdehyde solution (in DMSO) to the samples. After 30 min, the reaction was terminated by addition of 1 ml of 0.8 M HCl, precipitates were spun down by a 3,000 g centrifugation step for 3 min, and fluorescence was measured within 60 min. Zero time blank values were subtracted from the corresponding test values. All assays were performed in duplicate, from which the derived mean was used in all data analysis. Results were expressed as nmol His-Leu $ml^{-1}min^{-1}$ (Woods *et al.*, 2004).

8.2.3 Data analysis

Subjects were tested at both loci for HWE by Chi-square tests with one degree of freedom. Subjects were tested as a whole and then separated by subject group and gender. For analysis of Ethiopian data, low numbers of female subjects did not allow statistical comparison of athletes versus controls. To allow some form of comparison among females, general Ethiopian controls were combined with Arsi controls and 5-10 km athletes were combined with marathon athletes. In Kenyan subjects, female athletes (I and N) were considered as one group for comparison to controls due to low subject numbers. 18 tests were performed in total in Ethiopian subjects and 22 in Kenyan subjects. Inter-group genotype frequency differences were tested using contingency Chi-square tests.

Distribution of ACE activity levels was tested for normality using the Ryan-Joiner test. Between genotype differences for ACE activity were assessed by ANOVA, followed by Fisher's pairwise comparisons. Statistical significance was declared at $P < 0.05$. Linkage disequilibrium measures D and D' (Hedrick & Kumar, 2001) were calculated after calculation of expected haplotype frequencies using an EM algorithm to calculate maximum likelihood frequencies.

8.3 Results

8.3.1 Ethiopian ACE genotype results

When tested for HWE, of the 18 tests performed, all test showed that subjects were in HWE. Of the 96 European samples genotyped at both ACE loci, all subjects homozygous at I/D were also homozygous at 22982. Subjects homozygous for the I allele at the I/D locus were homozygous for the A allele at 22982 and likewise for the D allele at I/D with the G allele at 22982. Ethiopian subjects did not show complete linkage disequilibrium: D = 0.16, and D' = 0.76.

I/D genotype frequencies did not differ between controls and either athlete group (5-10km:d.f. = 2, χ^2 = 5, P = 0.08. M: d.f. = 2, χ^2 = 2.4, P = 0.31) (Figure 8.1). Nor did the Arsi control population differ significantly from the controls (d.f. = 2, χ^2 = 3.8, P = 0.15) or from the athletes in I/D genotype frequencies (5-10 km:d.f. = 2, χ^2 = 0.4, P = 0.82. M: d.f. = 2, χ^2 = 0.22, P = 0.89). However, when comparisons were made by allele frequency, it was found that 5-10 km athletes differed significantly from controls (d.f. = 2, χ^2 = 2.0, P = 0.036) (Figure 8.2), displaying an excess of the I allele (C: 0.28, A: 0.36, 5-10: 0.40, M: 0.35). However, they did not differ significantly from the Arsi control group (d.f. = 1, χ^2 = 0.4, P = 0.526), who also did not differ from the marathon athletes (d.f. = 2, χ^2 = 0.029, P = 0.864). The Arsi controls, however, showed a tendency to differ from the general control group (d.f. = 2, χ^2 = 3.4, P = 0.066). The marathon athletes did not differ significantly from controls (d.f. = 2, χ^2 = 1.3, P = 0.25). When subjects were separated by gender, I/D genotype frequency differences became apparent between male marathon athletes and general controls (d.f. = 2, χ^2 = 9.2, P = 0.01) (Figure 8.3). However, no other differences were apparent between subject groups. Comparisons of allele frequencies between male subject groups mirrored these findings, with male marathon athletes differing significantly from general controls (d.f. = 2, χ^2 = 4.7, P = 0.03), and Arsi controls displaying a tendency to differ from general controls (d.f. = 2, χ^2 = 3.7, P = 0.053). Female endurance athletes did not differ significantly from controls in their genotype distribution (d.f. = 2, χ^2 = 0.35, P = 0.84). I/D allele frequency was not significantly different between female athletes and controls (d.f. = 2, χ^2 = 0.3, P = 0.56).

When subject groups were compared by genotype at 22982, no significant differences were present (Figure 8.4). Controls did not differ from 5-10 km (χ^2 = 0.48, P = 0.78) or marathon athletes (χ^2 = 1.2, P = 0.54) in 22982 genotype. Nor did the Arsi controls differ

form either athlete group (A Vs. 5-10 km, $\chi^2 = 0.51$, $P = 0.78$; A Vs. M, $\chi^2 = 0.35$, $P = 0.84$). Furthermore, when 22982 allele frequencies were compared between subject groups, no frequency differences were apparent. No genotype frequency differences became apparent when subject groups were separated for gender. There were also no differences in 22982 allele frequency between subject groups when compared as a whole or when separated for gender.

Although Ethiopian athletes differ from controls in their regional origin and ethnicity (Chapter 3), Place of birth of Ethiopian subjects was not associated with ACE genotype at I/D (d.f. = 8, $\chi^2 = 3.1$, $P = 0.93$). When subjects from Arsi, the region known to produce a disproportionate number of athletes were compared to subjects from all other regions, no differences were present (d.f. = 2, $\chi^2 = 0.1$, $P = 0.95$). Nor did I/D genotype differ significantly between subjects of Semitic or Cushitic ethnicity (d.f. = 2, $\chi^2 = 4.0$, $P = 0.14$). Nor did genotype at 22982 differentiate between these groups (Region: d.f. = 8, $\chi^2 = 2.7$, $P = 0.95$; Ethnicity: d.f. = 2, $\chi^2 = 2.6$, $P = 0.27$). Similarly, when subjects from Arsi were compared to those from all other regions, no differences were present in 22982 genotype frequencies (d.f. = 2, $\chi^2 = 0.66$, $P = 0.72$).

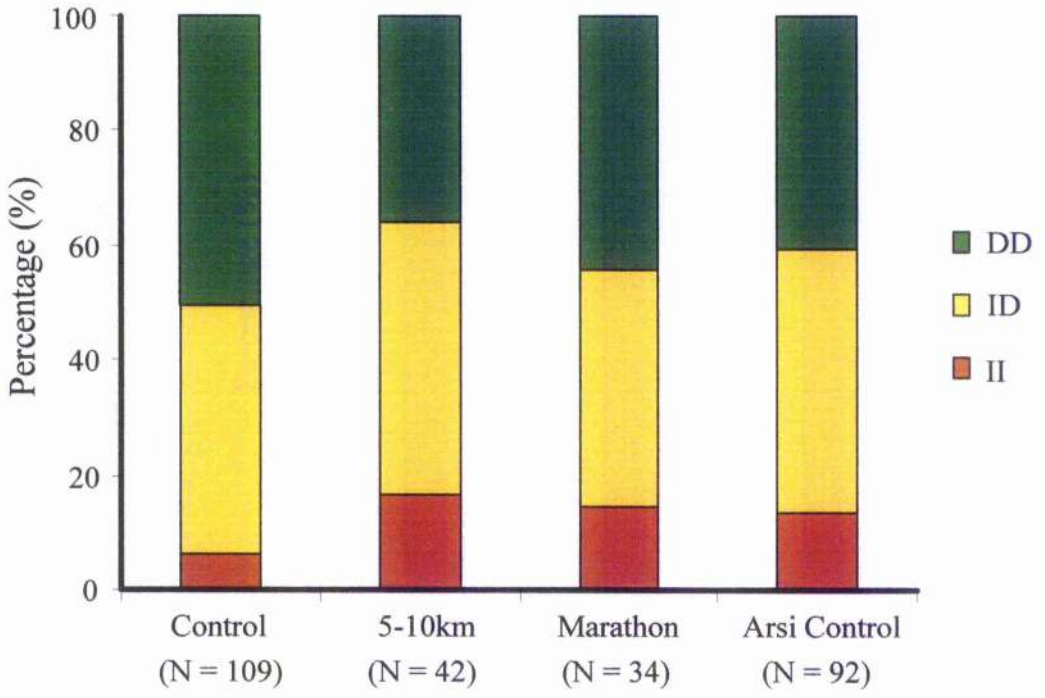


Figure 8.1 I/D genotypes of Ethiopian subject groups (male and female) are shown. There were no significant differences between groups.

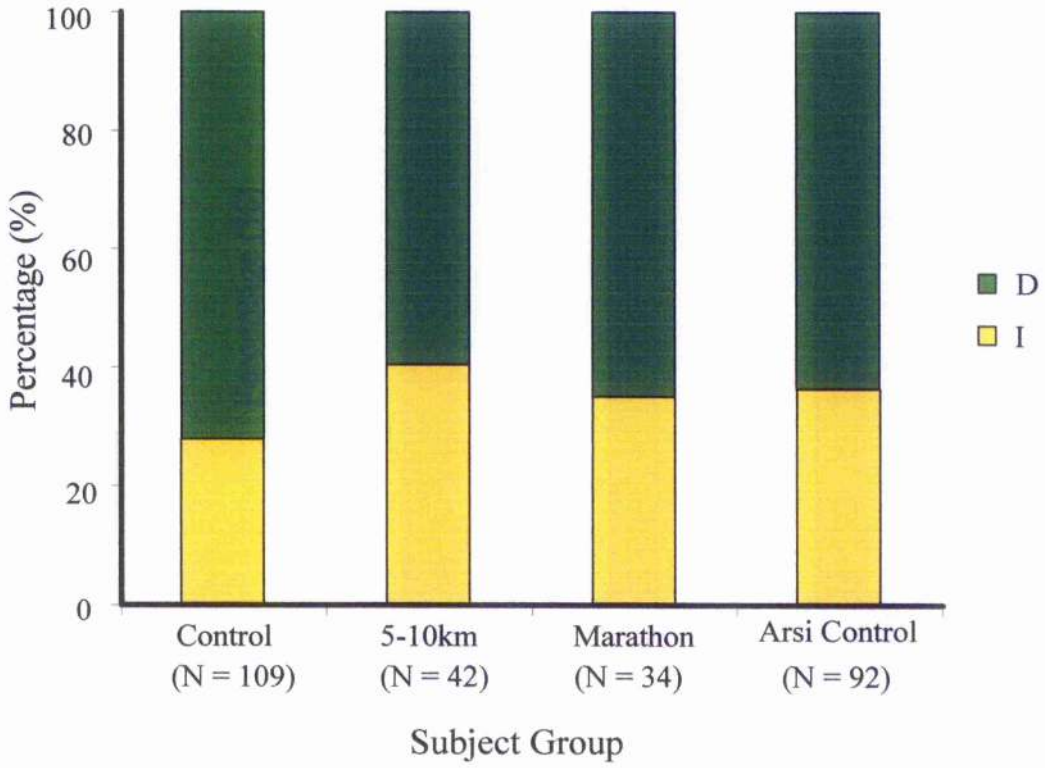


Figure 8.2 I/D allele frequencies are shown. 5-10km athletes differ significantly from controls (P = 0.036), but no other differences are present.

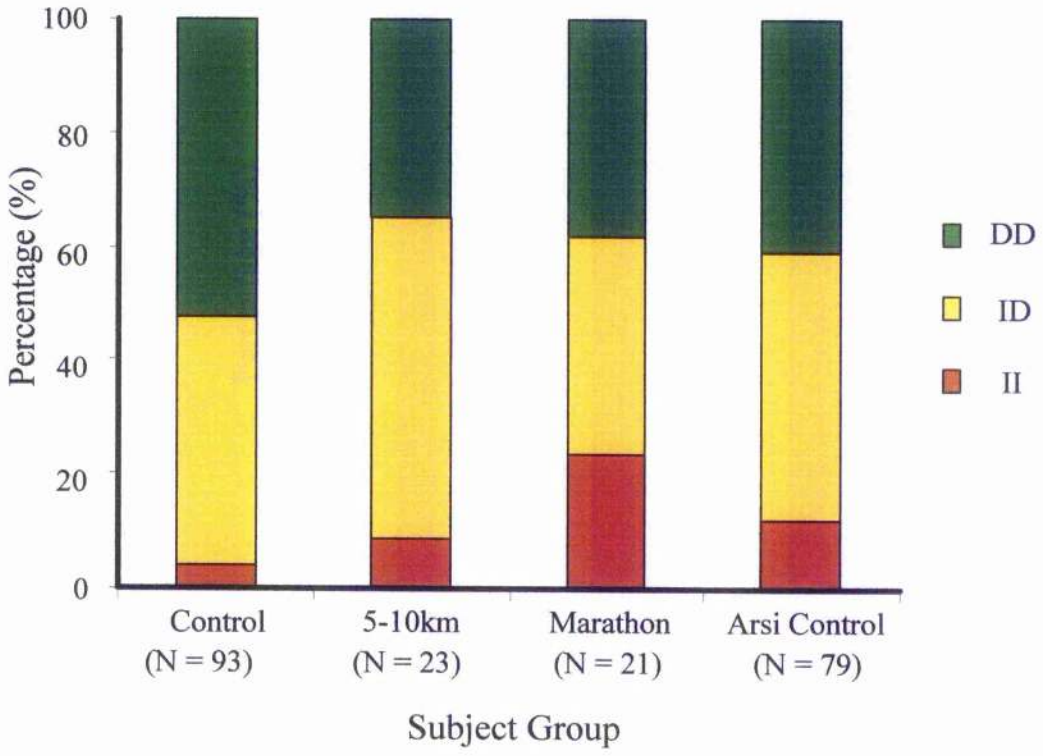


Figure 8.3 I/D genotypes are shown for male Ethiopian subjects. Marathon runners differ significantly from controls ($P = 0.01$) but no other differences are present.

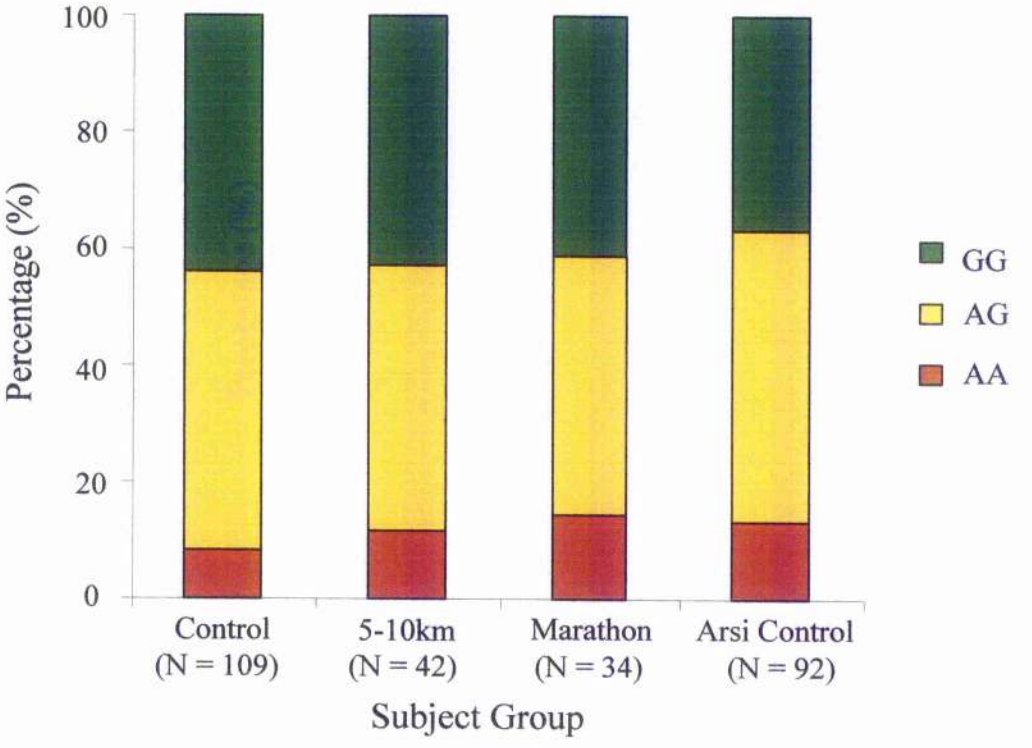


Figure 8.4 22982 Genotypes of Ethiopian subjects (male and female) are shown. No differences in genotype frequency are present between groups.

8.3.2 Kenyan ACE genotype results

Of the 22 tests, only international athletes showed a deviation from HWE for I/D genotype (P = 0.04), showing an excess of heterozygotes. In the Kenyan samples, linkage disequilibrium between I/D and A22982G was not complete: D = 0.12, and D' = 0.59.

Controls did not differ from the general Kenyan population (Kenyan Central Bureau of Statistics, 2003) in their geographical distribution throughout Kenya (Figure 4.2). Although athletes differed significantly from controls in ethnic distribution (Figure 4.3) and geographical distribution throughout Kenya (Figure 4.2), neither of these factors associated with I/D genotype (Ethnicity: d.f. = 2, χ^2 = 0.66, P = 0.72; Place of Birth: d.f. = 8, χ^2 = 11.4, P = 0.18). Nor did 22982 genotype associate with ethnicity or regional distribution (Ethnicity: d.f. = 2, χ^2 = 1.2, P = 0.56; Place of Birth: d.f. = 8, χ^2 = 4.6, P = 0.8).

I/D genotype frequencies were similar in both athlete groups and controls (Figure 8.5). No significant differences were found in genotype frequencies between subject groups (d.f. = 4, χ^2 = 4.1, P = 0.39). Controls did not differ significantly from national athletes (d.f. = 2, χ^2 = 0.6, P = 0.74) or international athletes (d.f. = 2, χ^2 = 2.0, P = 0.36). Nor did any significant difference arise when athlete groups were combined and compared to controls (d.f. = 2, χ^2 = 0.46, P = 0.80). I Allele frequencies were also similar between control and athlete groups (C: 0.38, N: 0.42, I: 0.38), and no significant differences were found between athletes and controls (d.f. = 2, χ^2 = 0.97, P = 0.62). When groups were separated for gender, male athletes did not differ significantly from male controls in I/D genotype frequencies (d.f. = 2, N: χ^2 = 2.3, P = 0.31; I: χ^2 = 0.20, P = 0.9). Low numbers of female athletes (N = 48, I = 11) did not allow female international athletes to be considered separately, but when all female athletes were compared to controls, no significant difference was found (d.f. = 2, χ^2 = 0.05, P = 0.98).

Whole group ACE activity was (Mean ± SD) 24.6 ± 6.9 nmol His-Leu $ml^{-1}min^{-1}$. The mean age of subjects of each genotype were II: 25 ± 6 yrs, I/D: 31 ± 9 yrs, II: 28 ± 9 yrs. AA: 28 ± 8 yrs, AG: 30 ± 9 yrs, GG: 29 ± 9 yrs. The mean ACE activity for each ACE I/D genotype is shown in Figure 8.6 and for each 22982 genotype in Figure 8.7. I/D genotype was associated with ACE activity (II: 22.20 ± 6.24, ID: 23.4 ± 6.83, DD: 27.01 ± 6.59 nmol His-Leu $ml^{-1}min^{-1}$; P = 0.034), and explained almost 13 % of the variance in ACE activity levels; II and ID genotypes did not differ significantly in ACE activity, while DD genotype had a significantly higher ACE activity than both II and ID genotypes (Figure

8.6). ACE activity was more strongly associated with genotype at 22982 (AA: 20.28 ± 5.26, AG: 24.81 ± 6.21, GG: 29.54 ± 6.6 nmol His-Leu $ml^{-1}min^{-1}$; $P < 0.001$). Genotype at 22982 was calculated to explain over 24 % of the variance in ACE activity levels. All genotypes at 22982 differed significantly from each other in ACE activity (Figure 8.7). In this east African population, although I/D is associated with ACE activity levels, the association is to a lesser extent than with genotype at 22982.

There were no significant differences in 22982 genotype frequency between groups (d.f. = 4, $\chi^2 = 5.67$, $P = 0.23$; Figure 8.8). Controls did not differ significantly from either athlete group (d.f. = 2, N: $\chi^2 = 0.11$, $P = 0.95$; I: $\chi^2 = 4.45$, $P = 0.11$). In addition, there was no difference in 22982 genotype frequencies when controls were compared to all athletes combined (d.f. = 2, $\chi^2 = 0.66$, $P = 0.72$). Frequencies of the A allele at 22982, which associates strongly with lower ACE activities in Africans as the I allele does in Europeans, were similar in athletes and controls (C: 0.53, N: 0.52, I: 0.41), and revealed no significant differences between subject groups (d.f. = 2, $\chi^2 = 5.3$, $P = 0.07$). As for I/D, separating for gender did not reveal any male specific effect of ACE genotype on elite athlete status when compared to controls in either national or international athletes (d.f. = 2, N: $\chi^2 = 0.23$, $P = 0.89$; I: $\chi^2 = 23.6$, $P = 0.16$). Analysis of combined female athletes against controls did not reveal any association between genotype and athlete status (d.f. = 2, $\chi^2 = 0.10$, $P = 0.95$).

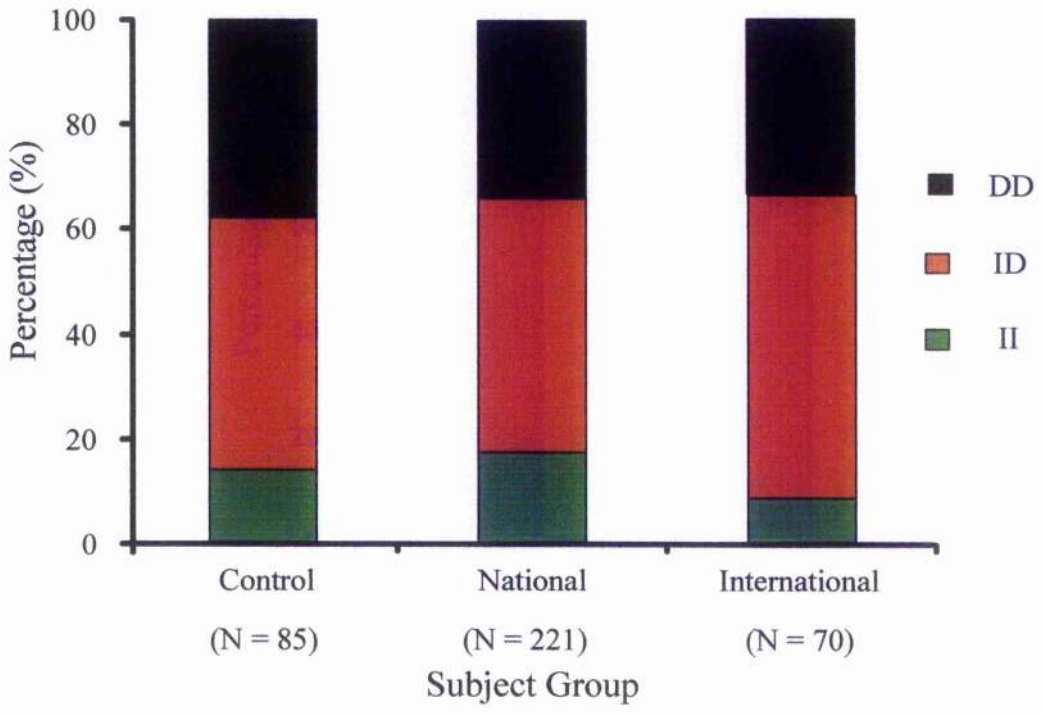


Figure 8.5. ACE I/D genotype frequencies in athletes and controls. The number of subjects for each genotype is indicated. No significant differences in genotype frequency were present between groups.

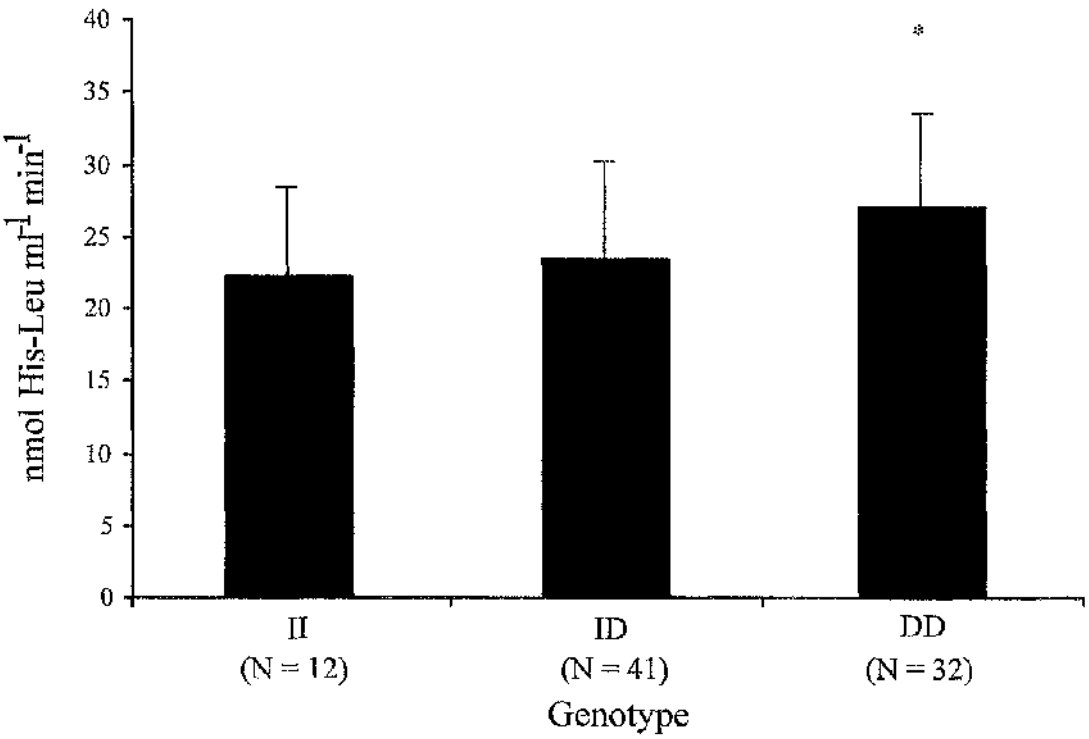


Figure 8.6. Circulating ACE activity levels (mean ± SD) according to I/D genotype. * indicates significant differences from II and ID groups.

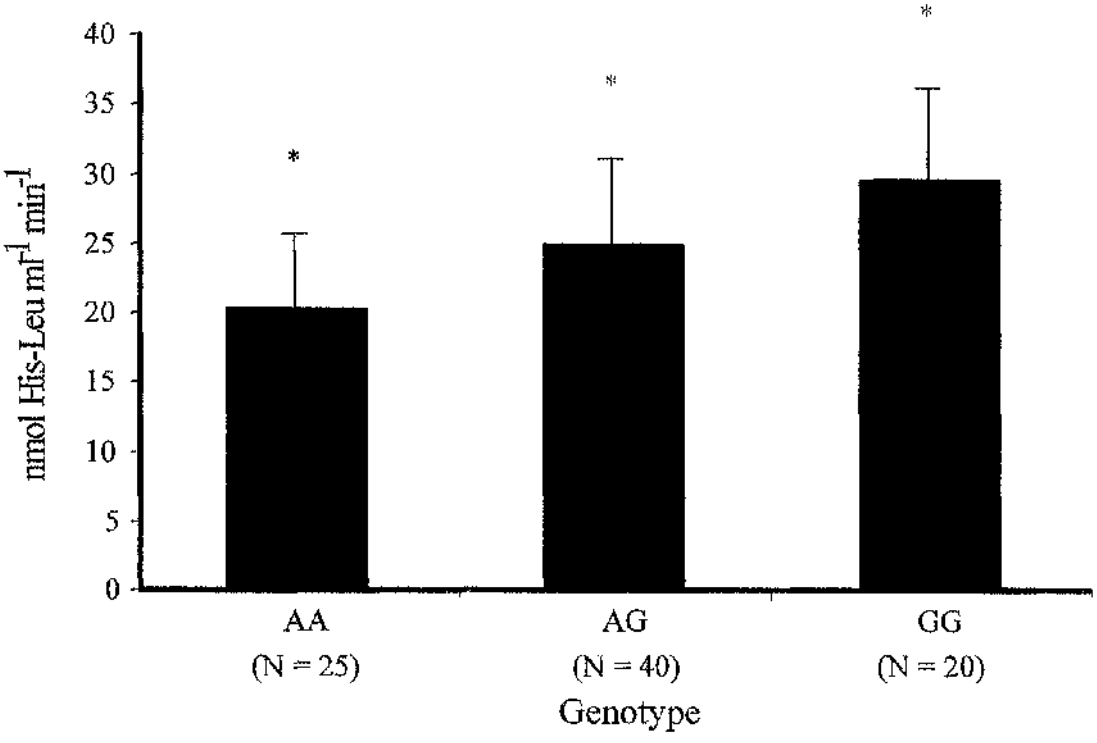


Figure 8.7. Circulating ACE activity levels (mean ± SD) according to 22982 genotype. * indicates significant differences between groups.

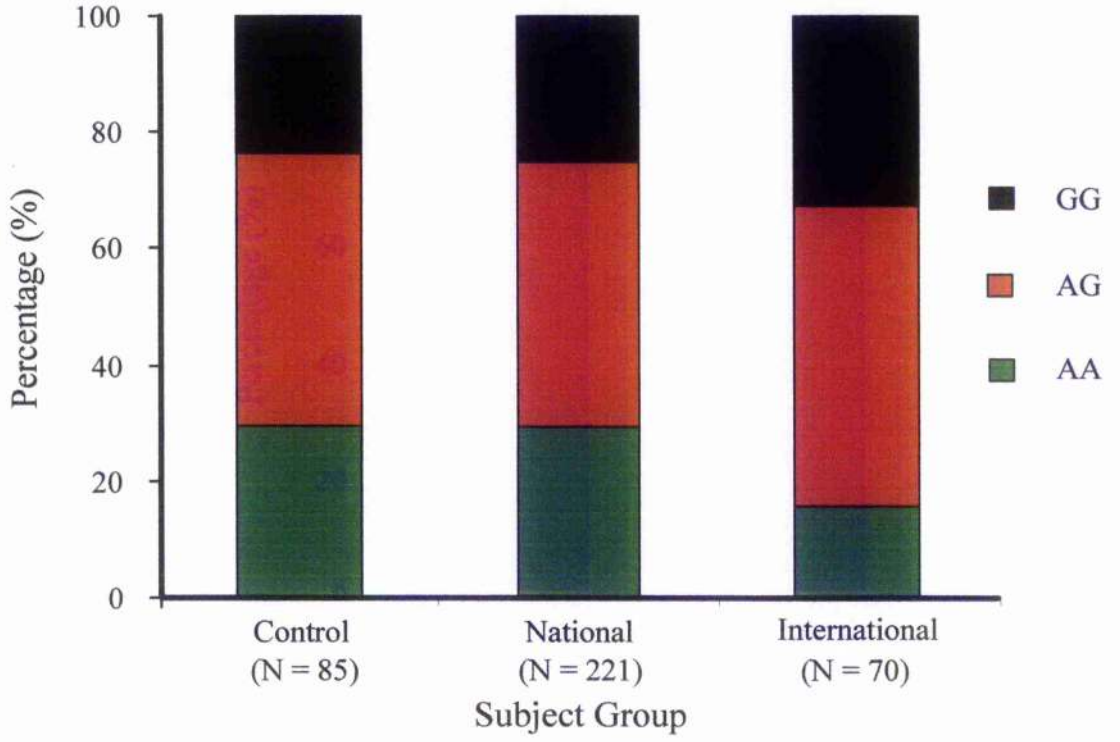


Figure 8.8. ACE 22982 genotype frequencies in athletes and controls. The number of subjects for each genotype is indicated. No significant differences in genotype frequency were present between groups.

8.4 Discussion

When Ethiopian subjects were separated by gender, male marathon runners displayed an excess of the I allele relative to the general control population (P = 0.01) (Figure 8.2). However, they did not differ from the Arsi control population. There were no differences between any Ethiopian subject groups for genotype at 22982, whether considered as a whole or separated by gender, which brings into question whether the differences in I/D genotype frequencies were as a result of a genetic effect on performance or some other reason. In Kenyan subjects, the results of the present study do not support an association between elite east African endurance athlete status and variation in the ACE gene. No significant differences in genotype or allele frequency at either of the loci tested were found between elite endurance athletes and controls. ACE genotype was independent of ethnicity and regional distribution of subjects. Although some studies report differences in the allele frequency of performance genes such as ACE and α-actinin-3 when the very best athletes were compared to controls (Nazarov *et al.*, 2001;Yang *et al.*, 2003), no significant differences in genotype or allele frequency were found in the present study even when only

the international athletes (i.e. including many Olympic and World champion distance runners) were compared to controls.

Despite some disparity amongst past reports, as reviewed by Jones et al (Jones *et al.*, 2002), ACE I/D genotype has been associated with exercise performance: the I allele with endurance and the D allele with power phenotypes. Therefore, it may have been expected that the I allele would be more frequent amongst elite endurance athletes than controls. Although this was the case in male Ethiopian subjects, it is questionable if this is a "real" effect. Many of the studies supporting the association between the I allele and endurance performance have been conducted in Caucasian subjects, where the I/D genotype is strongly associated with ACE activity (Rigat *et al.*, 1990), with no associations in black African subjects, where the link between ACE I/D genotype and ACE activity is not as strong (Zhu *et al.*, 2000). African populations show higher levels of haplotype diversity (Reich *et al.*, 2001) and are therefore more useful in assessing associations with SNPs by allowing contrasts to be seen where a pair of SNPs may be in tight linkage disequilibrium in European populations. It has been shown that the I/D polymorphism is not in linkage disequilibrium with many other polymorphisms in the ACE gene, especially in African subjects (Rieder *et al.*, 1999). Even in European populations, there is a major genetic subdivision which separates the deletion clade into two distinct haplotypes: H1 and H7 (Rieder *et al.*, 1999). Such findings may help further explain some of the controversy in the literature regarding the association between ACE genotype and physical performance. A polymorphism at np 22982 has been shown to be the variant with the largest deviation in ACE levels between opposing homozygotes in Afro-Carribean subjects in a study of seven polymorphisms spanning 13 Kb (Zhu *et al.*, 2000). This polymorphism is within 6 bp of an exon splice junction and may cause alternative splice forms (Zhu *et al.*, 2000), which may have a functional role in the inter-genotype variation in circulating ACE levels. This is in agreement with the results of the present study where genotype at 22982 is more strongly associated with variation in ACE levels than at I/D. It can be seen in Kenyan controls (Figure 8.7) that the G allele is associated with higher circulating ACE activity than the A allele. Data from the European subjects has shown that the I allele is in complete linkage disequilibrium with the A allele at 22982. However, this is not the case in the African subjects of the present study where linkage disequilibrium (D') between the two loci is 0.76 in Ethiopians and 0.59 in Kenyans. It is conceivable, therefore, that there would be a stronger association of 22982 genotype with performance than I/D. This makes it surprising that an influence of I/D genotype on elite male Ethiopian athletes status would not be complemented by a concurrent excess of AA genotypes. In Kenyan subjects, there is also no association between ACE genotype and elite athlete status, despite genotype at this

locus explaining over 24 % of the variance in ACE activity levels in this population. Also, the difference in marathon runners may also be expected to exhibit in 5-10km athletes based on associations in the literature (Myerson *et al.*, 1999) and their similar physiological requirements. This is not the case, as 5-10km athletes do not differ significantly from controls in their I/D or 22982 genotype distribution. It can be questioned, therefore, what is causing the effect of association between ACE I/D genotype in males and Ethiopian athlete status. One possible contributory factor is that association studies are prone to type I errors through multiple testing. In the Ethiopian comparisons, 52 chi-square tests were performed in total. At a probability level of 0.05, it would be expected that 2.6 tests were positive by chance (0.05 x 52 = 2.6). There were 2 positive tests, which compares well to this projection. However, it must be stated that 28 tests were carried out in Kenyan subjects and none were positive. Another possibility is that the general Ethiopian control group was not reflective of the Ethiopian population. Figure 8.1 shows that the control population shows a lower frequency of II genotypes than all other groups. This may, of course, reflect that there is a slight advantage in having an II genotype for an endurance athlete. However, there is no available data from the Ethiopian population on ACE genotype frequencies to check this. Additional samples are currently being collected from the general Ethiopian population to check this hypothesis.

Previous findings that many of the most successful East African athletes reside and train at altitude (Chapter three) may have supported a role for the ACE gene in the determination of their success in distance running. The ACE I allele has been associated, although not conclusively (Dehnert *et al.*, 2002) with the ability to tolerate high altitude conditions (Woods & Montgomery, 2001) as well as with endurance phenotypes (Jones *et al.*, 2002). It was considered, therefore, that there may have been a role for the ACE gene in the success of east African athletes, many of whom live and train at altitudes of over 2,500 m (Chapters three and four). It is conceivable that there has been selection for the I allele amongst such a population for the altitude tolerance phenotype it may confer. The association of the I allele with endurance performance phenotypes may have adapted this population toward endurance performance. However, the present results do not support this hypothesis.

Although some studies have shown that ACE genotype has an influence on physical performance and cardiovascular mediators of physical performance such as ventricular hypertrophy (Myerson *et al.*, 2001), no convincing physiological explanations for an influence of circulating ACE levels or the I allele on endurance performance have been proposed (Rankinen *et al.*, 2000b). As ACE gene variation has been associated with

adaptation to training rather than baseline performance phenotypes, it has been suggested that the exercise-related changes in ACE activity are more important than basal activity (Woods *et al.*, 2004). Although it has been found that the exercise-related change in ACE levels was independent of genotype after acute bouts of exercise (Woods *et al.*, 2004), the changes in ACE levels as a result of chronic training (as is the case in the athletes of the present study) are unknown. In addition, studies using ACE inhibitors have shown little effect on parameters of endurance performance such as $\dot{V}o_{2\,max}$ in humans (Predel *et al.*, 1994) and endurance time in rats (Bahi *et al.*, 2004). However, rather than an effect of lower ACE levels, there may be an indirect effect on endurance performance through other, as yet unknown, physiological mediators, perhaps even through complex patterns of linkage disequilibrium with other genes.

Endurance capacity is not solely a product of physiological factors such as $\dot{V}o_{2\,max}$ (Myburgh, 2003). Other physiological parameters, such as running economy, are major contributors. Parameters such as running economy, again, are multifactorial in their determination, being the product of many factors: physiological, biochemical and biomechanical. It is possible that ACE genotype influences endurance performance not through effects on aerobic capacity, but through influences on one of the many other factors contributing to endurance performance. Such a scenario may partially account for the negative results when investigating the association of ACE genotype with $\dot{V}o_{2\,max}$ (Rankinen *et al.*, 2000b) and the response to training (Rankinen *et al.*, 2000a), while there have been some positive associations with endurance performance (Gayagay *et al.*, 1998;Montgomery *et al.*, 1998;Myerson *et al.*, 1999). The ACE I allele has been associated with an increased percentage of type I muscle fibres in Japanese subjects (Zhang *et al.*, 2003). However, the physiological basis for this association is as yet unknown, although it was considered that this might be solely a linkage disequilibrium effect rather than a functional influence of the ACE gene. Nevertheless, given that levels of linkage disequilibrium can vary in different populations, a linkage disequilibrium effect found in Japanese subjects may not be present in other populations, and the use of the I/D marker as a determinant of such a phenotype would, therefore, be limited. The evidence to date, including the results of the present study, suggests that any influence of the ACE gene on endurance performance is not mediated through direct effects of the renin-angiotensin system but may be as a result of another gene(s) being in linkage disequilibrium. The present study finds that although circulating ACE levels are associated strongly with 22982 genotype in Kenyans, there is no association with elite endurance athlete status.

8.4.1 Conclusions

In conclusion, the current results do not support the hypothesis that ACE gene variation is associated with elite endurance athlete status in east African distance runners. Currently, there is controversy over the influence of ACE genotype on endurance performance, and this study does not support a role for ACE gene variation on the inter-individual or inter-population differences in endurance performance. The absence of an association between the I/D polymorphism and especially genotype at A22982G with elite athlete status also suggests that the ACE gene does not contribute significantly to the phenomenal success of east African endurance runners in international distance running competition. This is the largest and most homogenous genetic association study to date involving elite athletes and finds no association between ACE gene variation and elite endurance athlete status in east African distance runners.

Chapter 9

General Discussion

The primary objectives of the experiments described in the previous chapters were:

i. To investigate the environmental characteristics associated with elite east African status, and in doing so, better understand the reasons for the phenomenal success of east Africa in distance running. This was achieved by investigating the regional and ethnic origins of east African athletes, in response to evidence that they are not equally distributed throughout east Africa, but are confined mainly to distinct regions and ethnic groups. The suggestion that the distance that east Africans travel to school as children is influential in their success was also investigated.

ii. To establish if the distinct ethnic characteristics of elite athletes are indicative that the athletes arise from a limited genetic isolate, selected for endurance performance, through analysis of non-recombining DNA sequences, which allow the ancestry of individuals and populations to be traced. This was achieved by analysis of mtDNA and Y chromosome variation in elite east African distance runners relative to controls and also by comparing regions producing a large number of athletes to other regions to establish the level of population stratification throughout Ethiopia and Kenya.

iii. To investigate any genetic influences on the performance of east Africa in distance running through the analysis of existing candidate genes for human performance, and also by using the novel technique of mtDNA and Y chromosome haplogroup analysis of east African athletes relative to the general east African population. In doing so, this would provide the first information on the association between elite athlete status and genetic variation in African subjects.

iv. To validate existing associations of the ACE gene with human performance by establishing the ACE genotype frequency in a cohort of the world's most successful athletes. To date, all genetic associations with human physical performance have been in Caucasian subjects. The experiment in Chapter eight aimed to replicate existing associations in east African subjects. If the associations could be replicated in a cohort of the world's most successful athletes, of different ethnic background from those in the existing literature, it would add credibility to these genes having a biological influence on endurance performance.

The studies described in Chapters three and four described the environmental and ethnic factors associated with elite east African status. Those in Chapters five to seven were the

first to assess the association of mtDNA and Y chromosome haplogroups with athletic performance. This method allowed the identification of any genetic substructure in the east African athletes and the populations from which they are drawn. It also, particularly for mtDNA, allowed identification of a key candidate gene by haplogroup analysis rather than by the genotyping of many individual SNPs. To haplogroup the individuals, given the haploid nature of the genome, was a more effective method of assessing the association between mtDNA variation and elite athlete status and this study was the first to do so. The study of ACE genotype in Chapter eight identified no association between ACE and elite athlete status. ACE is the most studied of human performance genes and has produced many conflicting results. It was recognised that the I/D polymorphism is unlikely to be causal in the variation in ACE levels, and given lower levels of linkage disequilibrium in Africans it was necessary to assess the ACE gene variation at another loci, which is more strongly associated with the ACE phenotype and is potentially causal to the variation in ACE levels. This study was the first to do so, and assessed the influence of ACE gene variation on elite athlete status in African populations. In this chapter, the findings and general conclusions of this series of studies are discussed.

9.1 Environmental influences on east African success

The results from the studies in Chapters three and four highlight the importance of environmental factors in the success of elite East African athletes. Both Kenyan and Ethiopian athletes displayed distinct environmental backgrounds relative to their source populations. The neighbouring proximity of these countries means that they are both bisected by the Rift Valley, which ensures that both countries have regions of high altitude. From the concurrent findings in Ethiopia and Kenya, it appears that the regional origin of elite east African athletes has an influence on their success. In Ethiopia, the region of Arsi accounted for a particularly high number of elite athletes relative to controls. The region of Arsi accounts for less than 5 % of the Ethiopian population and only 3 % of the control group but 24 % of the 5-10 km athletes and 38 % of the marathon athletes (Figure 3.1). Similarly, Shewa accounts for only 12 % of controls but 19 % and 35 % of the 5-10km and marathon athletes, respectively (Figure 3.1). Kenya shows similar over-representations of certain regional origins amongst the elite athletes. The Rift Valley province, which includes areas such as Eldoret and the Nandi hills accounted for only 20 % of control subjects but 65 % and 81 % of the national and international standard Kenyan athletes respectively (Figure 4.2). It is clear, therefore, that regional origin is strongly associated with elite athlete status in east Africa. However, although a few possibilities were evident,

the causal factor mediating this association is unclear. All of the regions that showed over-representations of athletes lie at altitude, which raises the possibility that altitude inhabitation was causal to the over-representation. Altitude training is believed to be of benefit in improving endurance performance through the haematological adaptations it provokes (Wilber, 2001). There is debate over which method of hypoxic training is most effective and although there is some evidence that the "live high-train low" method is effective in improving sea level performance (Stray-Gundersen *et al.*, 2001), it is not clear if training at altitude, or simulated altitude is beneficial to endurance performance. Many east African athletes were born at altitude and many, particularly in Ethiopia where the entire national squad trains together, train at altitude (~2,500m). It is unclear at present how this may influence their performance and it has been commented that other areas of the world (Hamilton, 2000), and indeed Kenya (Manners, 1997) have people of similar lifestyle living at similar altitude, who do not show the same levels of athletic success. A study by Schmidt et al (2002) found that total haemoglobin and blood volume were influenced synergistically by training and altitude exposure in endurance athletes native to moderate altitude (~2,500m). This would certainly add credence to the belief that the altitude at which east African athletes are born and raised is influential in their success. The $\dot{V}o_{2\,max}$ of athletes native to altitude and athletes native to sea level was similar at their respective altitudes, but it was presumed that the $\dot{V}o_{2\,max}$ of the altitude native athletes would increase at sea level (Schmidt *et al.*, 2002). The extent to which this is true however, is unclear. The study by Saltin et al (1995b) found that those Kenyan athletes ($N = 2$) who were tested at altitude (~2,000m) showed an increase in $\dot{V}o_{2\,max}$ when tested at sea level of ~12 %. However, it has also been shown that high altitude natives do not gain much in $\dot{V}o_{2\,max}$ when tested in acute normoxia (Favier *et al.*, 1995). It has been shown that second generation Tibetan lowlanders acclimatise to altitude more quickly than Caucasians indigenous to altitude (Marconi *et al.*, 2004), and there is intriguing evidence of adaptations to chronic hypoxia that result in efficient aerobic performance on descent from high altitude (Marconi *et al.*, 2005), suggesting that east Africans have some long term adaptations which are beneficial to distance running through their inhabitation of high altitude regions. Evidence of different strategies to cope with hypobaric hypoxia (Beall, 2003) used by geographically isolated indigenous populations adds complication to the applicability of data from one population to another. Although there is some evidence of a genetic component to altitude adaptation in Himalayans, the genes that mediate the adaptations, are only slowly being elucidated (Rajput *et al.*, 2006). Despite the uncertainty over the extent to which altitude influences the success of east African athletes, the fact

that so many elite east African athletes originate from altitudinous regions justifies further research into the area.

As mentioned in Chapters three and four, there are other possible explanations for the regional disparities in athlete production in Ethiopia and Kenya. Ethiopian athletic coaches informed the study team that most of Ethiopia's best distance runners are from Arsi and that they go on talent scouting missions around Arsi to find new talent. It may be the case that the over-representation of athletes from Arsi is a reflection of a bias in talent scouting. The fact that track and field athletes also showed an excess of subjects from Arsi supports this possibility. It was found that 18 % of track and field athletes were from Arsi compared to 3 % of the controls (Figure 3.1), and although there is some evidence to suggest that anaerobic performance may be improved by altitude training (Nummela & Rusko, 2000), the benefit of altitude to a throw or jump athlete is not as potentially influential as it may be to an endurance athlete. In Kenya, the success of the Nandi is well publicised (Manners, 1997;Bale & Sang, 1996;Entine, 2001b) and the Nandi region has long been recognised as being a hotbed of athletic talent. Perhaps the most famous east African athlete of all time: Kipchoge Keino is from the Nandi region, and it is likely that his success is a motivation to other young Nandi. Similarly, in Ethiopia, Haile Gebresellasie is from the region of Arsi, which is likely to stimulate the interest of other aspiring runners from Arsi. This trend shows no sign of stopping with the emergence of Kenenisa Bekele, the world record holder at 5,000 and 10,000m, who is also from Arsi. Although these athletes are national heroes, it is likely that the population of Arsi identify strongly with their success. Although the prevalence of talent scouting in these regions is discussed as a possible determinant of the geographical imbalance in athlete production, the possibility remains that there is real biological reason behind the success of these runners that made them attractive to coaches in the first instance.

The regional clustering of athletes found in Chapters three and four may have been considered to reflect a genetic similarity among the athletes, as has been suggested, despite a paucity of evidence (Entine, 2001b;Manners, 1997). It was also found that both Ethiopian and Kenyan athletes displayed a distinct ethnicity relative to controls, with an excess of Cushitic and Nilotic athletes, respectively, relative to controls. Again, this may appear to suggest a genetic similarity among the athletes, given that gene frequencies often vary among different ethnic groups (Watson *et al.*, 1997). However, Ethiopian and Kenyan ethnic groups are geographically clustered, so these ethnic differences may be secondary to the geographical disparities between athletes and controls. As shown by Chapters five to seven, the athletes are not a genetically distinct group, as defined by mtDNA or Y

chromosome variation. It is perhaps unlikely, therefore, that this reflects a variation in the frequency of potential performance genes in athletes relative to the general population. It seems more likely that the distinct ethnic affiliations of the elite athletes are as a result of their geographical distribution.

Although there are ambiguities in the causality of regional and ethnic variations in athlete production, findings that do provide compelling evidence, which may be argued to be causal to the success of east African athletes, relate to the distances they run to school each day as children. Saltin et al. (1995b) showed that running to school raised aerobic capacity in children relative to those who were not physically active on a regular basis. Chapters three and four now show that the distances that they run to school distinguish between athletes and controls. It seems, therefore, that to be successful in distance running, having a high level of physical activity each day as children offers an advantage. This finding may also offer some insight into the success of east Africa in distance running relative to other nations. Even in the control groups, 25 % of Kenyans and 36 % of Ethiopians travelled farther than 5km to school each day and 22 % of Kenyans and 24 % of Ethiopians travelled to school by running (Figures 3.3, 3.4, 4.5, and 4.6). This is unheard of in Western populations, where very few children would travel such distances to school and even fewer would do so by foot. The activity levels of African children contrast with the increasing prevalence of childhood obesity in western populations, such as the UK (Ebbeling *et al.*, 2002). Intriguingly, there was also an indication in Kenyans that the athletes were more likely than controls to have run to school when covering equivalent distances, without the availability of motorised transport. It was found that of the athletes and controls travelling over 5 km per day to school, with no motorised transport available, 50 % of controls chose to run to school, compared to 84 % and 91 % of national and international athletes respectively. This may reflect that they gained more enjoyment from running; equally, and perhaps more intriguingly, it may reflect that they were inherently more able to cover these distances by running. Although no compelling evidence has yet been found for a genetic component in the success of east African athletes, if there is a genetic component to distance running, it is likely that in east Africa, those with gene variants beneficial to distance running realise their potential through using it regularly. While gene variants conferring benefits to physical performance may be prevalent elsewhere, they may not be recognised without being used.

9.2 Genetic influences on east African success

Results from Chapters five, six and seven highlight the level of genetic diversity in elite east African athletes. As mentioned, it has been suggested that elite east African athletes may display a form of genetic homogeneity predisposing them to athletic success. The practise of cattle raiding, prevalent in the Kalenjin people in Kenya, to whom the Nandi belong, has been proposed to serve as a mechanism for natural selection for athletic success to occur. This practise involved journeys of 100 miles or more at speed on foot to steal cattle and drive them home before their owners could catch up. It has been suggested that the better a young man was at cattle raiding, the more cattle he could accumulate, the more wives he could buy and therefore children he could father (Manners, 1997). It has also been suggested that because the Kalenjin tend to intermarry, which may potentiate a genetic isolate effect where levels of these putative performance genes would flourish (Manners, 1997). It is also suggested that linguistic evidence supports a genetic link between the Kalenjin of Kenya and the Oromo of Ethiopia, who are the main constituents of the Arsi region of Ethiopia (Manners, 1997). It is acknowledged in the above piece, however, that there is no genetic evidence to support the speculation. However, the results presented in the previous chapters do offer some insight into the claims, particularly in Ethiopia, where results from both mtDNA and Y chromosome are available and show congruence. If cattle raiding had caused a genetic selection for endurance performance, this would be manifested primarily in the Y chromosome results, as any population selection through males would be evident in a differentiation of Y chromosome haplogroup types prevalent in the elite athletes, and the population from which they are drawn. As can be seen in Chapter seven, although there were some significant differences present between athletes and controls, there remained a high degree of Y chromosome haplogroup diversity in elite Ethiopian athletes relative to the general population (Figure 7.1) (Semino *et al.*, 2002;Underhill *et al.*, 2000). Of course, the Y chromosome variation in elite Kenyan athletes has not yet been evaluated, so the speculations of Manners (1997) cannot be directly evaluated in Kenyan athletes as yet, but it is likely that they also exhibit a high degree of Y chromosome haplogroup variation, as they do in their mtDNA haplogroup distribution. However, the finding that the excess of Cushitic languages in Ethiopian athletes was coincident to an absence of such languages in the Kenyan athletes refutes the claim that the concurrent success of the Kalenjin in Kenya and the Oromo in Ethiopia is reflective of a similar linguistic ancestry.

9.2.1 mtDNA and Y chromosome characteristics of Ethiopian and Kenyan athletes

The idea that generations of intermarrying have caused a degree of genetic adaptation in certain Kenyan populations thereby predisposing them to athletic performance can be evaluated by the findings of Chapters five to seven. If there was selection in one region of east Africa, this would be manifested by that region displaying a lower diversity of mtDNA haplogroups than other regions. Results from mtDNA in both Kenyan and Ethiopian populations show that this is not the case. Although there were some statistical differences between Kenyan athletes and controls in haplogroup distribution (Figure 6.1), athletes showed similar levels of diversity in their haplogroup distributions to the general population. Athletes also displayed a broad range of haplogroups covering most of the human phylogeny, for both mtDNA (Maca-Meyer *et al.*, 2001) and Y chromosome (Y Chromosome Consortium, 2002). In fact, some of the most successful athletes in history share more recent common maternal and paternal ancestors with the author of this thesis than they do with some of their fellow athletes. The level of genetic similarity between the Kenyan and Ethiopian populations can also be assessed from the results of Chapters five and six. Although they display a similar range of mtDNA haplogroups, alike those previously reported in east Africa (Salas *et al.*, 2002), there are some key differences. Ethiopian mtDNA is characterised by a high frequency of Eurasian haplogroups (M, E1 and E2 in the previous chapters) (~45 %) (Kivisild *et al.*, 2004) which has been suggested to be a result of gene flow from the Near East back into Ethiopia. Kenya, however, displays a much lower level of these haplogroups (~10 %). Although there is little other published evidence on Kenyan mtDNA distribution (Brandstatter *et al.*, 2004), the lower level of Eurasian haplogroups, and subsequently higher levels of African specific "L" groups, presumably reflects a lower level of gene backflow from the Near East. The different genetic background of these populations not only highlights the inadequacy of a black versus white classification of subjects in inferring genetic background, but also the differences between geographically proximal populations.

Of course, the significant differences between athletes and controls for mtDNA haplogroups in Kenya and Y chromosome in Ethiopia may reflect a biological influence of genetic variation on athletic performance. This is especially plausible in mtDNA, as mtDNA is a strong candidate genome with the potential to influence human performance. Many studies have highlighted a maternal effect in the heritability of aerobic capacity (Lesage *et al.*, 1985;Bouchard *et al.*, 1999;Bouchard *et al.*, 1998). Furthermore, a number

of loci in this genome are known to influence physical performance, primarily in exercise intolerance pathologies (Wolfarth *et al.*, 2005). Studies have also suggested an influence of mtDNA on some of the inter-individual differences in aerobic capacity and its response to training (Dionne *et al.*, 1993;Murakami *et al.*, 2002). In addition, the study by Niemi at el. (2005) found an association between mtDNA haplogroups and elite endurance athlete status in Finnish athletes. Although the current results do not identify the particular polymorphisms that may be influential in the haplogroup frequency distribution, it highlights that polymorphisms on the L0 branch of the human mtDNA phylogeny may be influential in attaining elite endurance athlete status in Kenyans. Each branch of the human mtDNA phylogeny is defined by a series of mutations (normally between 40 and 70). Each of these mutations, in principle, could influence mitochondrial function, or mitochondrial biogenesis, and have an influence on aerobic performance. Also, there is evidence that rather than being selectively neutral, there has been selection for certain mtDNA types which has shaped regional mtDNA variation. It has been suggested that there has been climatic selection of mtDNA lineages dependent on their efficiency (Ruiz-Pesini *et al.*, 2004;Mishmar *et al.*, 2003). As well as a key role in ATP production to fuel exercise; OXPHOS is also important in thermogenesis. The balance between these two functions is determined by the coupling of OXPHOS. Tightly coupled OXPHOS elicits efficient ATP production, which would be beneficial in the tropics, but loosely coupled mitochondria would be more useful in more Northern Climes where less efficient ATP production would elicit a higher degree of thermogenesis. However, many of the mtDNA types found in east Africa are found in other areas of Africa, including west Africa, and Eurasia (Salas *et al.*, 2002), so the possibility that the general east African mtDNA distribution alone is responsible for their endurance performance is unlikely. Also, those in other tropical climes would carry a similar advantage, through efficient OXPHOS.

Similarly, the results finding differences in Y chromosome haplogroup distribution between elite Ethiopian athletes and controls may reflect an influence of the Y chromosome on attaining elite athlete status. However, the fact that the associations between Y chromosome haplogroups and elite athlete status were not replicated in both 5-10km and marathon athletes is not supportive of a biological influence of Y chromosome on performance. Although both comparisons did not reach statistical significance, the frequencies in both 5-10 km and marathon athletes of haplogroups E* and E3b1 go in the same direction relative to the controls. This adds some credence to the possibility of a biological influence of these haplogroups on elite athlete status. However, the frequencies of haplogroups E3* and K(xP) were in opposite directions relative to controls. For example, the haplogroup E3* is present at a frequency of 23 % in 5-10 km athletes relative

to 8 % and 4 % in the Arsi and general control group, respectively, but was not found in any of the marathon athletes (Figure 7.1). It is difficult to conceive how a Y chromosome haplogroup could confer an advantage in 5-10km events but not in marathon competition. This may indicate that, at least some of, the associations between elite athlete status and Y chromosome haplogroups are as a result of factors other than a direct biological influence on performance. As mentioned, there is no clear link has yet been established between the Y chromosome and athletic performance (Wolfarth *et al.*, 2005), however, as discussed in chapter seven, there are ways in which the Y chromosome may influence performance. Y chromosome variants have been associated with a number of phenotypes (Jobling & Tyler-Smith, 2003), and interestingly, with stature (Ellis *et al.*, 2001). This sustains the possibility that the Y chromosome may have a biological influence on performance, given the belief that anthropometric characteristics may be associated with the success of east African runners and their superior running economy (Larsen *et al.*, 2004;Saltin *et al.*, 1995b). Height has previously been shown to be negatively correlated with running economy (Maldonado *et al.*, 2002), albeit in 800m - 1500m runners, not in runners competing at distances of 5000m or more. In addition, recent work has identified that the Y chromosome is associated with cardiovascular phenotypes such as blood pressure (Charchar *et al.*, 2002), and although the Y chromosome has been physically mapped, relatively little is known about the function of the identified genes. The fact that the Y chromosome is associated with blood pressure also raises the possibility that it may be influential in other cardiovascular phenotypes. However, until more is known about the genes on the Y chromosome and their function, it is not possible to speculate on the functionality of their role in attaining elite athlete status. If the associations in Ethiopian subjects were replicated in the Kenyan athlete, this would add credence to the idea of a biological function on athletic performance. However, as in mtDNA (Carelli *et al.*, 2003), the association between the Y chromosome and any phenotype is complicated by factors such as genetic background (Ellis & Harrap, 2003), and the interaction between Y chromosome and autosomal variants (Charchar *et al.*, 2003) adds difficulty to repeating associations in different populations. Therefore, if the Ethiopian associations were not replicated in Kenyans, it would not rule out a role for a biological effect (Ellis & Harrap, 2003), but of course, would not be supportive of one. This was the case in the discrepancy between the mtDNA results between the Ethiopian and Kenyan populations, which raised the question of why they were not replicated. Previous associations between various measures of aerobic performance and mtDNA (Murakami *et al.*, 2002;Dionne *et al.*, 1993;Niemi & Majamaa, 2005) highlight the potential of mtDNA as a candidate genome for human performance. It has been suggested that polymorphisms in the D-loop of

mtDNA may influence mtDNA replication (Murakami *et al.*, 2002). For this to impact positively on mitochondrial volume, it would require a co-ordinated interaction with the nuclear genome, as mitochondrial biogenesis is a process involving the complex interaction of the nuclear and mitochondrial genomes (Adhihetty *et al.*, 2003). It is therefore possible that the interaction between nuclear and mitochondrial genes is a confounding factor in replicating associations (Carelli *et al.*, 2003). As previously discussed, the contrasting mtDNA haplogroup frequencies between Kenyan and Ethiopian subjects highlight the different genetic background of these populations, which may also partly explain the discrepant findings (Herrnstadt & Howell, 2004). Even given the associations between certain haplogroups and elite athlete status, the diversity of the athlete populations further attests to the likelihood that a multitude of genes is influential in an elite athlete phenotype. Of course, even when all of these genes are considered in combination, they alone are not sufficient to explain elite athlete status (Myburgh, 2003).

9.2.2 ACE gene variation and east African performance

Although it has been established that the majority of human genetic variation is within populations rather than between (Cavalli-Sforza & Feldman, 2003), and that the levels of genetic variation within Africa are greater than between Africa and Eurasia (Yu *et al.*, 2002), there may still be a role for genetics in inter-population differences in physical performance. There is controversy over the genetics of race and its utility in biomedical research (Burchard *et al.*, 2003;Cooper *et al.*, 2003), and much of the debate is focused around how genetics varies between skin colour groups. As mentioned, it is flawed to extrapolate skin colour as a proxy for genetics, given the diversity within skin colour groups. However, most of the literature on genetics of race is based around the genetic characterisation of populations on variants neutral to selection. If a population meets some form of selection pressure, such as high altitude or famine, genetic variants conferring benefit in such circumstances are likely to proliferate. If these genes confer additional benefits in oxygen transport or in economical use of available energy, a plausible mechanism for genetic variation to influence inter-population differences in performance arises. A potential example of this is the ACE gene. The I allele of the ACE gene has been implicated as important in high altitude tolerance (Montgomery *et al.*, 1998), shown by an increased frequency of Caucasian climbers with at least one I allele being able to tolerate high altitude. It is plausible, therefore, in high altitude populations, such as those found in east Africa, that there has been a selective advantage for those carrying the I allele. As discussed earlier, the I allele has also been associated with endurance performance (Jones

et al., 2002), which raised the possibility that the highland east Africans may have been adapted towards endurance performance indirectly. The results of Chapter eight, however, do not support the importance of the ACE gene in such a scenario. Findings of no significant association between the ACE I/D genotype and elite endurance athlete status did not support previous findings (Myerson *et al.*, 1999;Scanavini *et al.*, 2002;Gayagay *et al.*, 1998). The study of ACE gene variation in east African subjects was complicated by studies in African populations that have shown other variants elsewhere in the ACE gene to be more closely associated with circulating ACE levels than the I/D polymorphism (Cox *et al.*, 2002;Zhu *et al.*, 2000;Zhu *et al.*, 2001) in African populations, unlike Caucasian populations. A potentially functional variant elsewhere in the ACE gene (np A22982G in the sequence AF118569 as in Rieder et al.(1999), or 31958 in Cox et al. (2002)), has been found to show the largest phenotypic differences between genotypes (Zhu *et al.*, 2000). Although all markers tested in Europeans showed significant differences in ACE levels between each genotype (probably due to linkage disequilibrium between genotypes), the most marked difference in ACE levels in both Afro-Caribbean and European subjects was found between genotypes at A22982G. This meant that to be sure that an absence of association in east African athletes was not simply due to a lack of LD between I/D and the functional ACE variant, it was necessary to study A22982G. However, although the A22982G variant was strongly associated with circulating ACE levels in Kenyans, it was not associated with elite athlete status in either Ethiopians or Kenyans. Previous studies of the ACE gene have failed to find an association of the I allele with aerobic performance (Rankinen *et al.*, 2000a;Rankinen *et al.*, 2000b;Taylor *et al.*, 1999), which has been suggested to perhaps reflect a lack of homogeneity in the mixed athlete cohorts used and the fact that the ACE association with endurance performance in some cohorts may not be mediated by cardio-respiratory factors such as $\dot{V}o_{2\,max}$ (Jones *et al.*, 2002). However, it has been highlighted that the studies which tend to show positive associations are those with smaller, better defined cohorts, while those with larger cohorts often find negative results (MacArthur & North, 2005). Part of the problem in such association studies, and one that also troubles the field of exercise physiology in general, is the fact that is very difficult to work with truly elite athletes, which may further confound genetic associations. The study presented in Chapter eight, however, is the largest and most homogenous in the literature, contains many of the world's most successful athletes, yet still found no association between ACE gene variation and elite athlete status. The fact that A22982G genotype associates well with the ACE phenotype in Kenyans but not athletic performance does not support a role for the ACE gene in elite endurance athlete status. It may be the case that a complex pattern of LD is responsible for the published associations to date, with the ACE

gene in LD with another proximal variant influencing ACE levels. Of course, again, genetic background cannot be ruled out as influential in the disparities in results.

9.2.3 Applicability of current findings and considerations for future research

As discussed in Chapter 1, a number of genes have been associated with human physical performance. In the main, the associations have been in Caucasian populations, but it is likely that genetic variation also influences the variation in physical performance in other populations. What is unclear however, is the extent to which the genetic influence on performance may differ between populations. It has been found that the response of $\dot{V}o_{2\,max}$ to a standardised training program is a highly heritable phenotype (Bouchard *et al.*, 1999), which suggests that there is likely to be a significant genetic component influencing the adaptation to training of a key component of endurance performance. As many of the top Kenyan runners are from rural villages, rather than towns (Larsen, 2003), a recent study by Larsen et al (2005) investigated the extent to which the $\dot{V}o_{2\,max}$ training response differed between Nandi town and village boys. Significantly, the study found that there was no difference between the increase in $\dot{V}o_{2\,max}$ of town and village boys. Even more interesting, perhaps, was the finding that the magnitude of the training response was similar to that previously found in Caucasian boys (Fournier *et al.*, 1982). This perhaps demystifies some of the success of the Nandi runners, by showing that it may not be as evident that they have a genetic advantage making them such successful runners. Nevertheless, the studies comparing elite African runners to Caucasian runners, although flawed from a genetic perspective, show interesting findings which merit further investigation. The work of Saltin et al (1995b), shows some major differences between Kenyan and Scandinavian runners, such as a markedly lower level of ammonia accumulation, which warrants further investigation. Although the reason or effect of ammonia accumulation during exercise is not conclusive, hyperammonaemia has been associated with the onset of fatigue, as reviewed by Banister et al. (1990). This is therefore an area that requires further investigation into the reasons for the reduced ammonia accumulation in Kenyan runners. Once the pathways for ammonia generation have been elucidated, genetic variants that may influence the response could be identified. The findings of higher ammonia accumulation relative to lactate accumulation in African runners compared to Caucasian (Weston *et al.*, 2000) highlight the problems associated with extrapolating results from such studies to east African athletes. Although physiological studies need to be done to establish the reasons behind the success of east

African runners, the results from these studies should be extrapolated with caution (Hamilton & Weston, 2000). However, physiological studies can generate useful information on the determinants of elite performance, and which factors distinguish between elite and sub-elite performers.

Knowledge of physiological pathways important to exercise performance is vital if the study of genetics of physical performance is to advance. The example of the ACE gene highlights the shortcomings of many of the current methods. ACE gene variants have been associated with dozens of different phenotypes, pathological and non-pathological. Although there have been associations with many aspects of physical performance, clear physiological explanations underlying such associations have been lacking. The α-actinin-3 (ACTN3) R577X polymorphism is an example of a genetic variant that has a clear physiological implication resulting from the polymorphism. The finding that the X allele is associated with a deficiency of α-actinin-3 (North *et al.*, 1999) elucidates a clear reason for association of the variant with elite performance (Yang *et al.*, 2003). This is an example that should be followed for future genetic studies: they should be designed based on knowledge of the physiological processes that regulate performance. Association studies, in particular, need to consider the physiological grounding for any association to be considered valid. Association studies should always be interpreted with caution as they have a chance of eliciting false positive results, especially when they are subject to multiple testing (Ioannidis, 2003). An investigation into the reproducibility of association studies found that of the 166 associations studied, only 6 were replicated in over 75 % of repeat studies (Hirschhorn *et al.*, 2002). It is likely that some of these one-off associations are type I errors. However, an identifiable physiological impact on performance obviously strengthens the likelihood of any association being due to a biological influence on performance. It must be acknowledged that, as association studies, the results of the previous chapters have their limitations. The marginal associations found in the mtDNA and Y chromosome chapters, although potentially significant, do not identify causal factors and are, by their design, findings of association. Although none of the chapters above had corrections for multiple testing, the significant findings presented are more frequent than would be expected as a result of a multiple testing alone (Chapter 6: 4 of 21 tests were positive, and initial test between groups showed that international athletes differed from controls. Chapter 7: 6 of 27 tests were positive, and endurance athletes differed significantly from both control groups in their haplogroup distribution). Also, the direction of both athlete groups often go in the same direction relative to controls, which does support a "real" association rather than one as a result of chance alone. The above

discussion on the statistical chance raises another relevant issue. Genetic association studies are primarily used in the study of disease. They measure the association between specific genetic variants and the likelihood of exhibiting an easily measurable pathological phenotype. Often a single SNP can render a complete physiological pathway dysfunctional, resulting in a pathological phenotype. However, this is in stark contrast to the example of physical performance, and elite performance in particular, where we are primarily interested in variants that affect the fine-tuning of a system rather than its complete loss of function. To investigate and find variants which can improve performance by 1 % is exceedingly more difficult than finding those which result in a clear phenotype. When we consider that in elite athletic competition, the difference between winning and losing can be 0.1 %, the picture looks bleak, particularly when available technologies allow the analysis of only a few variants at a time. However, improving technologies such as microarray, and genetic studies with physiological grounding are likely to yield more sturdy results than at present.

Another significant confounding factor in association studies is population stratification. As the athletes in the current studies are of a distinct ethnic and environmental background relative to the controls, it is impossible to be sure that genetic associations are not as a result of an underlying genetic structure in the population from which they arise. For example, in the association of elite Kenyan athletes with the mtDNA haplogroup M (Figure 6.1), it was also found that subjects of Nilotic origin exhibited an excess of M haplogroups (Figure 6.3). Given that the athletes have an excess of subjects of Nilotic origin (Figure 4.3), it may be that the athletes' ethnic status alone is the reason for the association. For each association with elite athlete status in the current series of experiments, the association was further tested against geographical and ethnic classifications to test for population stratification, which, apart from the above example, yielded negative results. It is proposed, therefore, that the results of the present series of studies, although offering some insight the success of east African athletes, raise many more to be answered. Particularly, the findings on the geographical clustering of elite east African athletes raises interesting questions on why this might be the case. Based on the current understanding on the implications of chronic altitude exposure for athletic performance, one is not able to confidently suggest that the regional disparity in athlete production is as a result of altitude, but further investigation is warranted. Similarly, the finding that more of the elite athletes had chosen to run to school, when faced with similar distances to travel, raises the question of whether they were inherently more able to do so.

9.3 Why are east Africans successful in distance running?

Given the focus of the previous chapters, it may seem appropriate to offer a revised explanation of the reasons behind the success of east African distance runners. It is clear that the distance that east African athletes run to school as children accounts for a significant component of their success, not only relative to other nations, but also serves as a mechanism within east Africa to determine those who may be successful in distance running. This is perhaps likely to be the single most influential factor in their success based on current results. The regional distribution of east African athletes also appears to be influential in their status as elite athletes, although the reason for this is unclear. It may be due to the high altitude of these regions, or perhaps as a result of a prevalence of athletics in these regions due to the beliefs of those running east African athletics, which in itself may have a biological grounding. Differences between east Africans and others are more likely to be the result of their demanding training (Billat *et al.*, 2003) than any genetic advantage confined to east Africans, given the high levels of genetic diversity in the east African athletes: Ethiopian and Kenyan. The idea that east African athletes have some form of genetic advantage is pervasive, as it allows other less successful competitors to ascribe their success to a factor outwith their control. Furthermore, the fact that even amongst such a cohort of athletes, containing a number of the finest athletes the world has ever seen, there remains considerable allelic variation in previously identified, and novel, candidate genes attests to the complexity of the elite athlete phenotype. Although other nations are likely to contain sections of the population who have high levels of childhood physical activity and, certainly, other populations are indigenous to altitude, east Africa displays a culture of distance running success and environmental factors that are beneficial to endurance performance. This combination of factors is unique, and central in the success of east African runners relative to other populations.

9.4 General Conclusions

A number of conclusions can be drawn from the studies presented in the previous chapters. Although these conclusions are based on the largest, most successful, and most homogenous cohort in the published literature, current understandings on the importance of genetic background on the penetrance of variants influencing phenotypic variation may not allow extrapolation of these results to other populations.

i. Elite Ethiopian distance runners are of a distinct ethnic and environmental background, relative to the general Ethiopian population, and were found to have a distinct profile for all variables investigated in comparison to the general Ethiopian population. An excess of the athletes were from the regions of Arsi and Shewa relative to controls, both of which are altitudinous. Also athletes displayed a different ethnic profile to controls in that they tended to speak languages of Cushitic origin. Perhaps the most significant finding in Ethiopian athletes was that they travelled longer distances to school each day then controls and that more of them did so by running. Members of the Ethiopian marathon team displayed a more pronounced version of this profile relative to controls. This profile is associated with endurance success, and it is likely that these environmental conditions along with the cultural and motivational factors inherent in Ethiopia are contributors to East African distance running success.

ii. Environmental characteristics of elite Kenyan athletes relative to the general Kenyan population mirrored those found in Ethiopia. An excess of Kenyan runners were from the Rift Valley province, were Kalenjin and spoke languages of Nilotic origin relative to controls. Athletes also travelled farther to school than controls and mainly did so by running. International athletes tended to show a pronounced version of this profile compared with national athletes. The finding of concurrent results in Ethiopia and Kenya highlights the importance of such environmental factors in east African distance running success.

iii. International standard Kenyan athletes displayed an excess of L0 haplogroups, and a lower frequency of L2 haplogroups. Haplogroup frequencies in national athletes supported these findings. National athletes displayed an excess of M haplogroups, and a tendency toward a dearth of L2 haplogroups relative to controls. Athletes displayed a distinct geographical and ethnic heritage relative to the general Kenyan population, displaying an excess of subjects of Nilotic origin. Indeed, subjects of Nilotic origin displayed an excess of M haplogroups, suggesting that associations with elite athlete status may have been influenced by population stratification. Findings of differences in mtDNA haplogroup frequency between Kenyan athletes and controls were not replicated in Ethiopian athletes, where elite athletes did not differ from controls in mtDNA haplogroup frequency.

iv. However, both athlete groups displayed a distinct Y chromosome distribution from control subjects. Haplogroups E*, E3*, and K(xP) appear to confer an increased

potential to become an endurance athlete, whereas haplogroup E3b1 is negatively associated with elite athlete status. Although regional or ethnic variation within Ethiopia did not associate with Y chromosome haplogroup distribution, an influence of population stratification could not be ruled out. Although the Y chromosome has not previously been associated with athletic performance and has no obvious candidate genes identified to date; the possibility that these haplogroup associations with athletic performance are as a result of a biological effect cannot be ruled out. Although certain mtDNA and Y chromosome haplogroups were associated with elite east African athlete status, athletes showed similar levels of diversity to controls and, indeed, the general population. This finding does not support hypotheses that east African athletes arise from genetically isolated populations selected for endurance performance.

v. ACE gene variation is not a determinant of east African distance running success. Although levels of LD vary in east Africans relative to Caucasian populations, A22982G genotype, which was associated strongly with circulating ACE levels in Kenyan controls, was not associated with elite endurance athlete status. Although the ACE gene is the most studied of the candidate genes for human performance, this study did not support a role for the ACE I allele at I/D or A allele at A22982G in elite endurance athlete status.

Appendix 1

UNIVERSITY OF GLASGOW

ETHICS COMMITTEE FOR NON CLINICAL RESEARCH INVOLVING HUMAN SUBJECTS

RESEARCH SUBMISSION

Name of person(s) submitting research proposal **Dr Yannis Pitsiladis, Dr Richard Wilson; Dr William Goodwin**

Position held **Lecturers in IBLS and Department of Forensic Medicine and Science**

Department/Group/Institute/Centre - **Centre for Exercise Science and Medicine (CESAME), IBLS**

Name of Principal Researcher (if different from above) _____

Position held _____

Date of submission **January, 2002**

Project Title: **MtDNA haplotypes and the African endurance athlete: Does matrilineal genetics contribute to the dominance of African athletes in international endurance athletics?**

1. Describe the basic purposes of the research proposed.

Background of investigation

While African runners dominate international distance running events, the reason(s) behind their disproportionate success is unclear. Possible explanations range from favourable genetic conferred physical characteristics (Ama et al. 1986) to socio-economic factors (Phillips 1976). Previous attempts to unravel the genetic component focused on comparisons of highly heritable characteristics, such as skeletal structure and differences in distribution of muscle fibre types in "black" and "white" athletes (Ama et al. 1986, Weston et al. 1999, Weston et al. 2000). Many of the advocates of this approach seemed to ignore the fact that race is socially determined (i.e. often based on skin colour) and that these heritable characteristics are not evenly distributed among "race" groups.

A more recent approach has compared various heritable characteristics of populations from different geographical areas of the world. For example, Saltin et al. (1995a,b) compared aerobic capacity, muscle morphology and enzyme activities of Kenyan and Scandinavian runners. The superior running ability of Kenyan athletes was attributed primarily to their high aerobic capacity as well as their good running economy. However, most geographical areas comprise of heterogeneous populations and genetic differences within racially or geographically defined groups are greater than between groups (Barbujani et al. 1997). Another approach is therefore needed to assess the genetic basis of the current dominance of athletes of African descent in distance running.

MtDNA analysis has been used extensively since the pioneering work of Vigilant et al. (1991) and has lead to the identification of mtDNA haplogroups that are specific to either Africans, Europeans, or Asians/Amerindians (Alves-Silva et al. 2000). Haplogroup allocation of a given mtDNA lineage allows the assessment of its (sub)continental origin, so that the matrilineal ancestry of admixed populations can be evaluated. Ethiopia boasts a success record in international distance running second only to Kenya. With the third highest population in Africa of about 65 million, the Ethiopian population displays a high degree of heterogeneity, with the indigenous people experiencing influxes of Caucasoids since neolithic times as a result of immigration from Arabia and the Mediterranean. African influxes to the area have also come from the Sudan area. Genetic studies of autosomal loci have estimated that 40 % of the DNA is of Caucasoid origin while 60 % is African (Cavalli-Sforza et al. 1994). A recent study of mtDNA and Y-chromosome polymorphisms predicts approximately the same degree of admixture (Passarino et al. 1998). The mitochondrial DNA types of the Ethiopian population can be assigned to one of four broad groups: African (25 %), Caucasoid (25 %), Indian Caucasoid (20 %) and 30 % that cannot be attributed to specific geographical areas. The initial aim of this research is, therefore, to examine whether the distribution of mtDNA haplotypes in Ethiopian endurance athletes differs significantly from that of the general Ethiopian population. If so, matrilineal genetics would provide the first real evidence of a genetic basis of the current dominance of athletes of African descent in distance running.

2. Outline the design and methodology of the project. Please include in this section details of the proposed sample size.

We have been invited by Dr Bezabe Wolde (Secretary General of the Ethiopian Olympic Committee, General Secretary of the Ethiopian Athletics Federation) to visit Ethiopia and carry out this research (see Appendix 1).

Plan of Investigation
We propose to:

1. Collect saliva samples from past and present members of the Ethiopian national athletics team (5,000m to marathon), and from a sample of the general Ethiopian population;
2. Extract DNA from these samples;
3. Profile and classify the mitochondrial DNA; and
4. Analyse the distribution of mitochondrial DNA types found in the endurance athletes and compare this to the distribution within the sample of the general population.

Methods/Design of investigation
Approximately 70-80 members, past and present, of the Ethiopian athletic national team and approximately 300 subjects from the general Ethiopian population (i.e. students studying in Addis) will be invited to take part in this study. All subjects will be asked to complete a questionnaire (see Appendix 2) to establish the following background information:

1. Place of birth (and that of their immediate ancestors). This should aid identification of any particular regions with a disproportionately high number of successful athletes.
2. Language (and that of their immediate ancestors). This may provide further information on ethnic background.
3. Means of transport to and from school. A common view is that the superior running ability of African athletes may be primarily due to the large distances covered by the athletes running to school every day.
4. Athletic achievements (Ethiopian athletic national team only). This information is necessary to ensure inclusion of elite athletes only in the final data analysis.

MtDNA samples collected will be analysed and assigned to the previously described haplotypes (Passarino et al. 1998). The relative proportions of each haplotype within all groups will be compared to elucidate any patterns in their distribution. Statistical analysis will be carried out on the data to determine whether there are differences between the relative proportions in each group.

All subjects will be fully informed of the procedures and will required to give written or verbal consent, as appropriate, for their participation in the study (see Appendix 3).

3. Describe the research procedures as they affect the research subject and any other parties involved.

Saliva samples will be obtained from the athletes at training grounds in the presence of the Ethiopian national athletic coaches. Sampling of control subjects (i.e. students) will be carried out at Kotebe Teacher Training College in Addis Ababa. Samples will be non-invasive and procedures pose no health risk to the subjects.

4. What in your opinion are the ethical considerations involved in this proposal? (You may wish for example to comment on issues to do with consent, confidentiality, risk to subjects, etc.)

The ethical concerns are minor. The risks during testing are negligible and all saliva samples will be collected, stored and disposed of according to standard health and safety procedures.

Volunteers will be fully informed of the nature of the study and all testing procedures and will provide written informed consent (see Appendix 3). As a great majority of the general population is illiterate, an impartial witness (who will also act as a translator) will be present during the entire informed consent discussion. In such cases, verbal informed consent will be accepted.

DNA profiling will be carried out using the saliva samples collected. Saliva samples will be stored until successful publication of the work (probably no more than five years after completion of the project).

The individual data will be treated in the strictest confidence and will be revealed to no one outside the research group. The data will not be made available to coaching staff.

5. Outline the reasons which lead you to be satisfied that the possible benefits to be gained from the project justify any risks or discomforts involved.

The research undertaken will contribute to our understanding of why so many African athletes dominate endurance track events.

6. Who are the investigators (including assistants) who will conduct the research and what are their qualifications and experience?

Dr Yannis Pitsiladis PhD MMedSci BA, Dr Will Goodwin PhD, Dr Richard Wilson PhD, Dr Evelina Georgiades (postdoctoral research assistant). The principal investigators have wide ranging research experience over periods of up to 25 years without incident.

7. Are arrangements for the provision of clinical facilities to handle emergencies necessary? If so, briefly describe the arrangements made.

N/A

8. In cases where subjects are identified from information held by another party (for example, a doctor or hospital) describe the arrangements whereby you gain access to this information.

N/A

9. Specify whether subjects will include students or others in a dependent relationship.

N/A

10. Specify whether the research will include children or those with mental illness, disability or handicap. If so, please explain the necessity of using these subjects.

N/A

11. Will payment be made to any research subject? If so, please state the level of payment to be made, and the source of the funds to be used to make the payment.

NO

12. Describe the procedures to be used in obtaining a valid consent from the subject. Please supply a copy of the information sheet provided to the individual subject.

Volunteers will be fully informed of the nature of the study and all testing procedures. As a great majority of the general population is illiterate, an impartial witness (who will also act as a translator) will be present during the entire informed consent discussion. In such cases verbal informed consent will be accepted. Sampling will NOT proceed unless there is a positive response from subjects.

13. Comment on any cultural, social or gender-based characteristics of the subject which have affected the design of the project or which may affect its conduct.

None

14. Give details of the measures which will be adopted to maintain the confidentiality of the research subject.

The information obtained will be anonymized and individual information will not be passed on to anyone outside the study group. The results of the tests will not be used for selection purposes.

15. Will the information gained be anonymized? If not, please justify.

Yes

16. Will the intended group of research subjects, to your knowledge, be involved in other research? If so, please justify.

No

17. Date on which the project will begin (1 February, 2002) and end (1 February, 2003)

18. Please state location(s) where the project will be carried out.

Ethiopian training ground and Kotebe Teacher Training College in Addis Ababa.

Information sheet

Institute of Biomedical and Life Sciences
University of Glasgow

INFORMATION SHEET

Study title: Mitochondrial DNA haplotypes and the African endurance athlete: Can matrilineal genetics explain the dominance of black African athletes in international endurance athletics?

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What is the purpose of the study? A team of University researchers are visiting your area to collect a small saliva sample for analysis. This sample may enable us to investigate why African athletes dominate endurance running events.

Why have I been chosen? You have been selected as a possible participant in this investigation because you are a well-trained endurance athlete and/or in good health. 300-400 volunteers are being sought.

Do I have to take part? It is up to you to decide whether or not to take part. If you decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

What will happen to me if I take part? You will be asked to complete a short questionnaire giving some personal information about yourself and immediate family. This will include your place of birth and your native language. It is also asked that you provide the same information for your parents and grandparents (only blood relatives). This is to help define from which area your immediate ancestors originate. We also ask what distance you travelled to school each day and what means (if any) of transport you used. A common view is that the superior running ability of African athletes may be primarily due to the large distances covered by the athletes running to school every day.

You will then be required to provide a small saliva sample for subsequent analysis. This procedure is safe and as such will not pose any risk to your health.

What are the side effects of taking part? There are no side effects.

What are the possible disadvantages and risks of taking part? There are no disadvantages or risks to your health.

What are the possible benefits of taking part? We hope that you will find out more about why African athletes dominate endurance running events.

Will my taking part in this study be kept confidential? All information which is collected about you during the course of the research will be kept strictly confidential.

What will happen to the results of the research study? Results will be published in a peer-reviewed scientific journal once the study is completed. You will not be identified in any publication.

If you wish to find out more about this investigation, you can contact:

Dr Yannis Pitsiladis
Lecturer, Institute of Biomedical and Life Sciences
West Medical Building
University of Glasgow
Glasgow, G12 8QQ
Phone: 0141 330 3858
Fax: 0141 330 6542
e-mail: Y.Pitsiladis@bio.gla.ac.uk

Consent Form

I .....

give my consent to the research procedures which are outlined above, the aim, procedures and possible consequences of which have been outlined to me

Signature .....

Date .....
Signed _____ Date _____
(Proposer of research)

Where the proposal is from a student, the Supervisor is asked to certify the accuracy of the above account.

Signed _____ _____ Date _____
(Supervisor of student)

COMMENT FROM HEAD OF DEPARTMENT/GROUP/INSTITUTE/CENTRE

This is an interesting project that should contribute to our understanding of the genetical basis for successful athletic performance. It is my understanding that the researchers have made the necessary formal collaborative arrangements with the collaborating group in Ethiopia, and that approval to carry out this study has been granted by the Ethiopian Authorities.

General Information

First name:		Date of Birth: (Day / Month / Year)			
Surname:					
Gender	Female ☐ Male ☑				
Place of birth:	Arusi	Native Language:	Amhara	Nationality:	Amhara

Nationality of grandparents:	1) Mother's mother	2) Mother's father	3) Father's mother	4) Father's father
	Amhara	Amhara	Amhara	Amhara
Place of birth of grandparents:	5) Mother's mother	6) Mother's father	7) Father's mother	8) Father's father
	Arusi	Arusi	Arusi	Arusi
Language of grandparents:	9) Mother's mother	10) Mother's father	11) Father's mother	12) Father's father
	Amhara	Amhara	Amhara	Amhara

				Other (please specify)
What is/was your best athletic discipline?	5km ☐	10km ☑	Marathon ☐	

			Other
How far was your school from your home?	less than 5km ☐	more than 5km but less than 20km ☑	

				Other
How did you normally get to school?	walk ☐	run ☑	transport ☐	

List your most significant athletic achievements (with dates and times if possible):

It is so many. We can not count it.

DNA sample ☐ (number): A.S.t

Appendix 2

UNIVERSITY OF GLASGOW

ETHICS COMMITTEE FOR NON CLINICAL RESEARCH INVOLVING HUMAN SUBJECTS

RESEARCH SUBMISSION

Name of person(s) submitting research proposal **Dr Yannis Pitsiladis, Dr Richard Wilson**

Position held **Lecturers in IBLS**

Department/Group/Institute/Centre - **Centre for Exercise Science and Medicine (CESAME), IBLS**

Name of Principal Researcher (if different from above) _____

Position held _____

Date of submission:

Project Title: **MtDNA haplotypes and the African endurance athlete: Does matrilineal genetics contribute to the dominance of African athletes in international endurance athletics?**

_____

1. Describe the basic purposes of the research proposed.

Background of investigation

While African runners dominate international distance running events, the reason(s) behind their disproportionate success is unclear. Possible explanations range from favourable genetic conferred physical characteristics (Ama et al. 1986) to socio-economic factors (Phillips 1976). Previous attempts to unravel the genetic component focused on comparisons of highly heritable characteristics, such as skeletal structure and differences in distribution of muscle fibre types in "black" and "white" athletes (Ama et al. 1986, Weston et al. 1999, Weston et al. 2000). Many of the advocates of this approach seemed to ignore the fact that race is socially determined (i.e. often based on skin colour) and that these heritable characteristics are not evenly distributed among "race" groups.

A more recent approach has compared various heritable characteristics of populations from different geographical areas of the world. For example, Saltin et al. (1995a,b) compared aerobic capacity, muscle morphology and enzyme activities of Kenyan and Scandinavian runners. The superior running ability of Kenyan athletes was attributed primarily to their high aerobic capacity as well as their good running economy. However, most geographical areas comprise of heterogeneous populations and genetic differences within racially or geographically defined groups are greater than between groups (Barbujani et al. 1997). Another approach is therefore needed to assess the genetic basis of the current dominance of athletes of African descent in distance running.

MtDNA analysis has been used extensively since the pioneering work of Vigilant et al. (1991) and has lead to the identification of mtDNA haplogroups that are specific to either Africans, Europeans, or Asians/Amerindians (Alves-Silva et al. 2000). Haplogroup allocation of a given mtDNA lineage allows the assessment of its (sub)continental origin, so that the matrilineal ancestry of admixed populations can be evaluated. In 2002 we carried out a study examining whether the distribution of mtDNA haplotypes in Ethiopian endurance athletes differed significantly from that of the general Ethiopian population (approved by the university ethics committee). We now would like to extend the initial study and to examine whether the distribution of mtDNA haplotypes in Kenyan endurance athletes differs significantly from that of the general Kenyan population. If so, matrilineal genetics would provide the first real evidence of a genetic basis of the current dominance of athletes of African descent in distance running.

2. Outline the design and methodology of the project. Please include in this section details of the proposed sample size.

We have been invited by Dr Michael K Boit ("living Kenyan athletic legend" and current Chairman of the Department of Exercise and Sport Science, Kenyatta University) to visit Kenya and carry out this research.

Plan of Investigation
We propose to:

1. Collect saliva samples from past and present members of the Kenyan national athletics team (5,000m to marathon), and from a sample of the general Kenyan population (Control Group);
2. Extract DNA from these samples;
3. Profile and classify the mitochondrial DNA;
4. Collect blood samples (5 mls) from selected individuals from the Control Group; and
5. Analyse the distribution of mitochondrial DNA types found in the endurance athletes and compare this to the distribution within the sample of the general population.

Methods/Design of investigation
Approximately 70-80 members, past and present, of the Kenyan athletic national team and approximately 300 subjects from the general Kenyan population (i.e. students studying in Kenyatta University) will be invited to take part in this study. All subjects will be asked to complete a questionnaire (amended from the previously approved questionnaire so as to include Kenyan languages/regions, see Appendix 1) to establish the following background information:

5. Place of birth (and that of their immediate ancestors). This should aid identification of any particular regions with a disproportionately high number of successful athletes.
6. Language (and that of their immediate ancestors). This may provide further information on ethnic background.
7. Means of transport to and from school. A common view is that the superior running ability of African athletes may be primarily due to the large distances covered by the athletes running to school every day.
8. Athletic achievements (Kenyan athletic national team only). This information is necessary to ensure inclusion of elite athletes only in the final data analysis.

MtDNA samples collected will be analysed and assigned to the previously described haplotypes. The relative proportions of each haplotype within all groups will be compared to elucidate any patterns in their distribution. Statistical analysis will be carried out on the data to determine whether there are differences between the relative proportions in each group.

All subjects will be fully informed of the procedures and will required to give written or verbal consent, as appropriate, for their participation in the study (see Appendix 2 for previously approved Information sheet).

3. Describe the research procedures as they affect the research subject and any other parties involved.

Saliva samples will be obtained from the athletes at training grounds in the presence of the Kenyan national athletic coaches. Sampling of control subjects (i.e. students) will be carried out at Kenyatte University in Nairobi. Saliva and blood samples will be obtained by standard procedures previously approved by the university ethics committee and should therefore pose no health risk to the subjects.

4. What in your opinion are the ethical considerations involved in this proposal? (You may wish for example to comment on issues to do with consent, confidentiality, risk to subjects, etc.)

The ethical concerns are minor. The risks during testing are negligible and all saliva and blood samples will be collected, stored and disposed of according to standard health and safety procedures.

Volunteers will be fully informed of the nature of the study and all testing procedures and will provide written informed consent (see Appendix 2). For any illiterate subjects, an impartial witness (who will also act as a translator) will be present during the entire informed consent discussion. In such cases, verbal informed consent will be accepted.

DNA profiling will be carried out using the saliva samples collected. Saliva samples will be stored until successful publication of the work (probably no more than five years after completion of the project).

The individual data will be treated in the strictest confidence and will be revealed to no one outside the research group. The data will not be made available to coaching staff.

5. Outline the reasons which lead you to be satisfied that the possible benefits to be gained from the project justify any risks or discomforts involved.

The research undertaken will contribute to our understanding of why so many African athletes dominate endurance track events.

6. Who are the investigators (including assistants) who will conduct the research and what are their qualifications and experience?

Dr Yannis Pitsiladis PhD MMedSci BA, Dr Richard Wilson PhD. The principal investigators have wide ranging research experience over periods of up to 25 years without incident.

8. Are arrangements for the provision of clinical facilities to handle emergencies necessary? If so, briefly describe the arrangements made.

N/A

8. In cases where subjects are identified from information held by another party (for example, a doctor or hospital) describe the arrangements whereby you gain access to this information.

N/A

9. Specify whether subjects will include students or others in a dependent relationship.

N/A

10. Specify whether the research will include children or those with mental illness, disability or handicap. If so, please explain the necessity of using these subjects.

N/A

13. Will payment be made to any research subject? If so, please state the level of payment to be made, and the source of the funds to be used to make the payment.

NO

14. Describe the procedures to be used in obtaining a valid consent from the subject. Please supply a copy of the information sheet provided to the individual subject.

Volunteers will be fully informed of the nature of the study and all testing procedures. An impartial witness (who will also act as a translator) will be present during the entire informed consent discussion for illiterate subjects. In such cases verbal informed consent will be accepted. Sampling will NOT proceed unless there is a positive response from subjects.

13. Comment on any cultural, social or gender-based characteristics of the subject which have affected the design of the project or which may affect its conduct.

None

14. Give details of the measures which will be adopted to maintain the confidentiality of the research subject.

The information obtained will be anonymized and individual information will not be passed on to anyone outside the study group. The results of the tests will not be used for selection purposes.

15. Will the information gained be anonymized? If not, please justify.

Yes

16. Will the intended group of research subjects, to your knowledge, be involved in other research? If so, please justify.

No

17. Date on which the project will begin (immediately) and end (1 February, 2005)

19. Please state location(s) where the project will be carried out.

Kenyan training facilities and Kenyatta University in Nairobi.

Information sheet / Consent forms

Institute of Biomedical and Life Sciences
University of Glasgow

INFORMATION SHEET

Study title: Mitochondrial DNA haplotypes and the African endurance athlete: Can matrilineal genetics explain the dominance of black African athletes in international endurance athletics?

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What is the purpose of the study? A team of University researchers are visiting your area to collect a small saliva sample for analysis. This sample may enable us to investigate why African athletes dominate endurance running events.

Why have I been chosen? You have been selected as a possible participant in this investigation because you are a well-trained endurance athlete and/or in good health. 300-400 volunteers are being sought.

Do I have to take part? It is up to you to decide whether or not to take part. If you decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

What will happen to me if I take part? You will be asked to complete a short questionnaire giving some personal information about yourself and immediate family. This will include your place of birth and your native language. It is also asked that you provide the same information for your parents and grandparents (only blood relatives). This is to help define from which area your immediate ancestors originate. We also ask what distance you travelled to school each day and what means (if any) of transport you used. A common view is that the superior running ability of African athletes may be primarily due to the large distances covered by the athletes running to school every day.

You will then be required to provide a small saliva and blood sample for subsequent analysis. This procedure is safe and as such will not pose any risk to your health.

What are the side effects of taking part? There are no side effects.

What are the possible disadvantages and risks of taking part? There are no disadvantages or risks to your health.

What are the possible benefits of taking part? We hope that you will find out more about why African athletes dominate endurance running events.

Will my taking part in this study be kept confidential? All information which is collected about you during the course of the research will be kept strictly confidential.

What will happen to the results of the research study? Results will be published in a peer-reviewed scientific journal once the study is completed. You will not be identified in any publication.

If you wish to find out more about this investigation, you can contact:

Dr Yannis Pitsiladis
Lecturer, Institute of Biomedical and Life Sciences
West Medical Building
University of Glasgow
Glasgow, G12 8QQ
Phone: 0141 330 3858
Fax: 0141 330 6542
e-mail: Y.Pitsiladis@bio.gla.ac.uk

Consent Form

I .....

give my consent to the research procedures which are outlined above, the aim, procedures and possible consequences of which have been outlined to me

Signature .....

Date .....

Signed _____ Date _____
(Proposer of research)

Where the proposal is from a student, the Supervisor is asked to certify the accuracy of the above account.

Signed _____ _____ Date _____
(Supervisor of student)

General Information

First name:

Surname:

Date of Birth: Day Month Year

Gender: Female ☐ Male ☑

Place of birth: KISII | Native Language: NANDI | Nationality: KENYAN

Place of birth of grandparents: *(choose below)*	1) Mother's mother	2) Mother's father	3) Father's mother	4) Father's father
	RV	RV	RV	RV

1) Nairobi Province; 2) Central Province; 3) Rift Valley Province; 4) Western Province; 5) Coast Province; 6) Eastern Province; 7) Northeastern Province; 8) Nyanza Province

Language of grandparents: *(choose below)*	5) Mother's mother	6) Mother's father	7) Father's mother	8) Father's father
	NANDI	NANDI	NANDI	NANDI

1) Luhya, 2) Gusii, 3) Kuria, 4) Akamba, 5) Kikuyu, 6) Embu, 7) Meru, 8) Mbere, 9) Thanaka, 10) Swahili, 11) Mijikenda, 12) [illegible] 13) Pokomo, 14) Taita, 15) Taveta, 16) Luo, 17) Maasai, 18) Samburu, 19) Turkana, 20) Teso, 21) Njemps, 22) Elmolo, 23) Munch, 24) Muukweni, 25) Pokot, 26) Tugen, 27) Kipsigis, 28) Elkony, 29) Boni, 30) Somali, 31) Rendille, 32) Orma, 33) Pokot, 34) Galdera, 35) Other (please specify)

What is/was your best athletic discipline? 5km ☐ 10km ☑ Marathon ☐ Other (please specify)

How far was your school from your home? less than 2km ☐ | 2-5 km ☐ | 5-10 km ☑ | 10-15 km ☐ | Other (please specify)

How did you normally get to school? walk ☐ run ☑ transport ☐ Other (please specify)

List your most significant athletic achievements (with dates and times if possible):

WON GOLD MEDAL, OLYMPICS
3000 M. STEEPLECHASE

List the most important reasons for becoming an athlete (eg. Kenyan tradition, significant others etc)?

- TALENT
- BETTER THAN OTHERS IN SCHOOL

DNA sample ☐ (number): LL029

Appendix 3

Genotype - The allele combination in a pair of chromosomes at a certain polymorphic site

Haplogroup - A group of similar haplotypes. Often defined by slowly mutating markers, relative to those used in the determination of haplotypes

Haplotype - The combination of alleles in a chromosomally linked segment of DNA

Insertion/Deletion polymorphism - A polymorphism defined by the presence or absence of additional nucleotides in a sequence

Linkage disequilibrium - non-random association of alleles at different loci. Alleles are said to be in complete linkage disequilibrium if they are always found together

Phenotype - Any measurable characteristic in an individual. Often the physical manifestation of a particular genotype.

Polymorphism - Change in the DNA sequence at a frequency of >1 %

Restriction Fragment Length Polymorphism - A polymorphism that results in different fragment lengths when digested with a restriction enzyme

References

Adhihetty, P. J., Irrcher, I., Joseph, A. M., Ljubicic, V., & Hood, D. A. (2003). Plasticity of skeletal muscle mitochondria in response to contractile activity. *Exp.Physiol* **88**, 99-107.

Alvarez, R., Terrados, N., Ortolano, R., Iglesias-Cubero, G., Reguero, J. R., Batalla, A., Cortina, A., Fernandez-Garcia, B., Rodriguez, C., Braga, S., Alvarez, V., & Coto, E. (2000). Genetic variation in the renin-angiotensin system and athletic performance. *Eur.J.Appl.Physiol* **82**, 117-120.

Ama, P. F., Lagasse, P., Bouchard, C., & Simoneau, J. A. (1990). Anaerobic performances in black and white subjects. *Med.Sci.Sports Exerc.* **22**, 508-511.

Ama, P. F., Simoneau, J. A., Boulay, M. R., Serresse, O., Theriault, G., & Bouchard, C. (1986). Skeletal muscle characteristics in sedentary black and Caucasian males. *J.Appl.Physiol* **61**, 1758-1761.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-465.

Bahi, L., Koulmann, N., Sanchez, H., Momken, I., Veksler, V., Bigard, A. X., & Ventura-Clapier, R. (2004). Does ACE inhibition enhance endurance performance and muscle energy metabolism in rats? *J.Appl.Physiol* **96**, 59-64.

Baker, J. & Horton, S. (2003). East African running dominance revisited: a role for stereotype threat? *Br.J.Sports Med.* **37**, 553-555.

Bale, J. & Sang, J. (1996). *Kenyan Running. Movement culture, geography, and global change.* Frank Cass and Co Ltd.

Banister, E. W. & Cameron, B. J. (1990). Exercise-induced hyperammonemia: peripheral and central effects. *Int.J.Sports Med.* **11 Suppl 2**, S129-S142.

Beall, C. M. (2003). High-altitude adaptations. *Lancet* **362 Suppl**, s14-s15.

Billat, V., Lepretre, P. M., Heugas, A. M., Laurence, M. H., Salim, D., & Koralsztein, J. P. (2003). Training and bioenergetic characteristics in elite male and female Kenyan runners. *Med.Sci.Sports Exerc.* **35**, 297-304.

Blanchard, A., Ohanian, V., & Critchley, D. (1989). The structure and function of alpha-actinin. *J.Muscle Res.Cell Motil.* **10**, 280-289.

Bosch, A. N., Goslin, B. R., Noakes, T. D., & Dennis, S. C. (1990). Physiological differences between black and white runners during a treadmill marathon. *Eur.J.Appl.Physiol Occup.Physiol* **61**, 68-72.

Bouchard, C., An, P., Rice, T., Skinner, J. S., Wilmore, J. H., Gagnon, J., Perusse, L., Leon, A. S., & Rao, D. C. (1999). Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. *J.Appl.Physiol* **87**, 1003-1008.

Bouchard, C., Daw, E. W., Rice, T., Perusse, L., Gagnon, J., Province, M. A., Leon, A. S., Rao, D. C., Skinner, J. S., & Wilmore, J. H. (1998). Familial resemblance for VO2max in the sedentary state: the HERITAGE family study. *Med.Sci.Sports Exerc.* **30**, 252-258.

Bouchard, C., Malina, R. M., & Perusse, L. (1997). *Genetics of Fitness and Physical Performance* Human Kinetics.

Brandstatter, A., Peterson, C. T., Irwin, J. A., Mpoke, S., Koech, D. K., Parson, W., & Parsons, T. J. (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int.J.Legal Med.* **118**, 294-306.

Burchard, E. G., Ziv, E., Coyle, N., Gomez, S. L., Tang, H., Karter, A. J., Mountain, J. L., Perez-Stable, E. J., Sheppard, D., & Risch, N. (2003). The importance of race and ethnic background in biomedical research and clinical practice. *N.Engl.J.Med.* **348**, 1170-1175.

Carelli, V., Giordano, C., & d'Amati, G. (2003). Pathogenic expression of homoplasmic mtDNA mutations needs a complex nuclear-mitochondrial interaction. *Trends Genet.* **19**, 257-262.

Cavalli-Sforza, L. L. & Feldman, M. W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nat.Genet.* **33 Suppl**, 266-275.

Chagnon, P., Gee, M., Filion, M., Robitaille, Y., Belouchi, M., & Gauvreau, D. (1999). Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population. *Am.J.Med.Genet.* **85**, 20-30.

Charchar, F. J., Tomaszewski, M., Padmanabhan, S., Lacka, B., Upton, M. N., Inglis, G. C., Anderson, N. H., McConnachie, A., Zukowska-Szczechowska, E., Grzeszczak, W., Connell, J. M., Watt, G. C., & Dominiczak, A. F. (2002). The Y chromosome effect on blood pressure in two European populations. *Hypertension* **39**, 353-356.

Charchar, F. J., Tomaszewski, M., Strahorn, P., Champagne, B., & Dominiczak, A. F. (2003). Y is there a risk to being male? *Trends Endocrinol.Metab* **14**, 163-168.

Chen, Y. S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K., & Wallace, D. C. (2000). mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *Am.J.Hum.Genet.* **66**, 1362-1383.

Chen, Y. S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S., & Wallace, D. C. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet.* **57**, 133-149.

Coetzer, P., Noakes, T. D., Sanders, B., Lambert, M. I., Bosch, A. N., Wiggins, T., & Dennis, S. C. (1993). Superior fatigue resistance of elite black South African distance runners. *J.Appl.Physiol* **75**, 1822-1827.

Collins, M., Xenophontos, S. L., Cariolou, M. A., Mokone, G. G., Hudson, D. E., Anastasiades, L., & Noakes, T. D. (2004). The ACE gene and endurance performance during the South African Ironman Triathlons. *Med.Sci.Sports Exerc.* **36**, 1314-1320.

Cooper, R. S., Kaufman, J. S., & Ward, R. (2003). Race and genomics. *N.Engl.J.Med.* **348**, 1166-1170.

Cox, R., Bouzekri, N., Martin, S., Southam, L., Hugill, A., Golamaully, M., Cooper, R., Adeyemo, A., Soubrier, F., Ward, R., Lathrop, G. M., Matsuda, F., & Farrall, M. (2002). Angiotensin-1-converting enzyme (ACE) plasma concentration is influenced by multiple ACE-linked quantitative trait nucleotides. *Hum.Mol.Genet.* **11**, 2969-2977.

Danser, A. H., Schalekamp, M. A., Bax, W. A., van den Brink, A. M., Saxena, P. R., Riegger, G. A., & Schunkert, H. (1995). Angiotensin-converting enzyme in the human heart. Effect of the deletion/insertion polymorphism. *Circulation* **92**, 1387-1388.

Dehnert, C., Weymann, J., Montgomery, H. E., Woods, D., Maggiorini, M., Scherrer, U., Gibbs, J. S., & Bartsch, P. (2002). No association between high-altitude tolerance and the ACE I/D gene polymorphism. *Med.Sci.Sports Exerc.* **34**, 1928-1933.

DiMauro, S. & Schon, E. A. (2001). Mitochondrial DNA mutations in human disease. *Am.J.Med.Genet.* **106**, 18-26.

Dionne, F. T., Turcotte, L., Thibault, M. C., Boulay, M. R., Skinner, J. S., & Bouchard, C. (1993). Mitochondrial DNA sequence polymorphism, VO2max, and response to endurance training. *Med.Sci.Sports Exerc.* **25**, 766-774.

Duey, W. J., Bassett, D. R., Jr., Torok, D. J., Howley, E. T., Bond, V., Mancuso, P., & Trudell, R. (1997). Skeletal muscle fibre type and capillary density in college-aged blacks and whites. *Ann.Hum.Biol.* **24**, 323-331.

Ebbeling, C. B., Pawlak, D. B., & Ludwig, D. S. (2002). Childhood obesity: public-health crisis, common sense cure. *Lancet* **360**, 473-482.

Ellis, J. A. & Harrap, S. B. (2003). Tall stories: the devilish detail of genetic association studies. *Clin.Endocrinol.(Oxf)* **59**, 278-279.

Ellis, J. A., Stebbing, M., & Harrap, S. B. (2001). Significant population variation in adult male height associated with the Y chromosome and the aromatase gene. *J.Clin.Endocrinol.Metab* **86**, 4147-4150.

Entine, J. (2001a). Race and Sport. The race to the swift - if the swift have the right ancestry. *Peak Performance Special Issue: Race And Genetics* 1-6.

Entine, J. (2001b). *Taboo: Why Black Athletes Dominate Sports And Why We're Afraid To Talk About It* PublicAffairs, New York.

Fagard, R., Bielen, E., & Amery, A. (1991). Heritability of aerobic power and anaerobic energy generation during exercise. *J.Appl.Physiol* **70**, 357-362.

Favier, R., Spielvogel, H., Desplanches, D., Ferretti, G., Kayser, B., & Hoppeler, H. (1995). Maximal exercise performance in chronic hypoxia and acute normoxia in high-altitude natives. *J.Appl.Physiol* **78**, 1868-1874.

Folland, J., Leach, B., Little, T., Hawker, K., Myerson, S., Montgomery, H., & Jones, D. (2000). Angiotensin-converting enzyme genotype affects the response of human skeletal muscle to functional overload. *Exp.Physiol* **85**, 575-579.

Fournier, M., Ricci, J., Taylor, A. W., Ferguson, R. J., Montpetit, R. R., & Chaitman, B. R. (1982). Skeletal muscle adaptation in adolescent boys: sprint and endurance training and detraining. *Med.Sci.Sports Exerc.* **14**, 453-456.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science* **296**, 2225-2229.

Gaskill, S. E., Rice, T., Bouchard, C., Gagnon, J., Rao, D. C., Skinner, J. S., Wilmore, J. H., & Leon, A. S. (2001). Familial resemblance in ventilatory threshold: the HERITAGE Family Study. *Med.Sci.Sports Exerc.* **33**, 1832-1840.

Gastin, P. B. (2001). Energy system interaction and relative contribution during maximal exercise. *Sports Med.* **31**, 725-741.

Gayagay, G., Yu, B., Hambly, B., Boston, T., Hahn, A., Celermajer, D. S., & Trent, R. J. (1998). Elite endurance athletes and the ACE I allele--the role of genes in athletic performance. *Hum.Genet.* **103**, 48-50.

Gennarelli, M., Pavoni, M., Cruciani, F., De Stefano, G., Dallapiccola, B., & Novelli, G. (1999). CTG repeats distribution and Alu insertion polymorphism at myotonic dystrophy (DM) gene in Amhara and Oromo populations of Ethiopia. *Hum.Genet.* **105**, 165-167.

Grafakou, O., Hol, F. A., Otfried, S. K., Siers, M. H., ter Laak, H., Trijbels, F., Ensenauer, R., Boelen, C., & Smeitink, J. (2003). Exercise intolerance, muscle pain and lactic acidaemia associated with a 7497G>A mutation in the tRNASer(UCN) gene. *J.Inherit.Metab Dis.* **26**, 593-600.

Hagberg, J. M., Ferrell, R. E., McCole, S. D., Wilund, K. R., & Moore, G. E. (1998). VO2 max is associated with ACE genotype in postmenopausal women. *J.Appl.Physiol* **85**, 1842-1846.

Hagberg, J. M., McCole, S. D., Brown, M. D., Ferrell, R. E., Wilund, K. R., Huberty, A., Douglass, L. W., & Moore, G. E. (2002). ACE insertion/deletion polymorphism and submaximal exercise hemodynamics in postmenopausal women. *J.Appl.Physiol* **92**, 1083-1088.

Hagberg, J. M., Moore, G. E., & Ferrell, R. E. (2001). Specific genetic markers of endurance performance and VO2max. *Exerc.Sport Sci.Rev.* **29**, 15-19.

Hamilton, B. (2000). East African running dominance: what is behind it? *Br.J.Sports Med.* **34**, 391-394.

Hamilton, B. & Weston, A. (2000). Perspectives on East African middle and long distance running. *J.Sci.Med.Sport* **3**, vi-viii.

Hedrick, P. & Kumar, S. (2001). Mutation and linkage disequilibrium in human mtDNA. *Eur.J.Hum.Genet.* **9**, 969-972.

Henze PB (2000). *Layers of Time: A History of Ethiopia*, pp. pp. xxiii. C.Hurst & Co Ltd..

Herrnstadt, C. & Howell, N. (2004). An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. *Mitochondrion.* **4**, 791-798.

Hirschhorn, J. N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genet.Med.* **4**, 45-61.

Hoberman, J. (1997). *Darwin's Athletes. How sport has damaged black America and preserved the myth of race* Houghton Mifflin Company, New York.

Holloszy, J. O. & Coyle, E. F. (1984). Adaptations of skeletal muscle to endurance exercise and their metabolic consequences. *J.Appl.Physiol* **56**, 831-838.

Hoppeler, H. & Vogt, M. (2001). Hypoxia training for sea-level performance. Training high-living low. *Adv.Exp.Med.Biol.* **502**, 61-73.

Huang, H. & Manton, K. G. (2004). The role of oxidative damage in mitochondria during aging: a review. *Front Biosci.* **9**, 1100-1117.

Ingman, M., Kaessmann, H., Paabo, S., & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713.

Ioannidis, J. P. (2003). Genetic associations: false or true? *Trends Mol.Med.* **9**, 135-138.

Jobling, M. A., Hurles, M. E., & Tyler-Smith, C. (2004). *Human Evolutionary Genetics* Garland Publishing.

Jobling, M. A. & Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nat.Rev.Genet.* **4**, 598-612.

Jones, A., Montgomery, H. E., & Woods, D. R. (2002). Human performance: a role for the ACE genotype? *Exerc.Sport Sci.Rev.* **30**, 184-190.

Kenyan Central Bureau of Statistics. Statistical Abstract. Ministry of Planning and National Development . 2003. Nairobi, Kenya, Government Press.
Ref Type: Generic

Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., Parik, J., Geberhiwot, T., Usanga, E., & Villems, R. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am.J.Hum.Genet.* **75**, 752-770.

Klissouras, V. (1971). Heritability of adaptive variation. *J.Appl.Physiol* **31**, 338-344.

Klissouras, V. (1997). Heritability of adaptive variation: an old problem revisited. *J.Sports Med.Phys.Fitness* **37**, 1-6.

Kuroki, Y., Iwamoto, T., Lee, J., Yoshiike, M., Nozawa, S., Nishida, T., Ewis, A. A., Nakamura, H., Toda, T., Tokunaga, K., Kotliarova, S. E., Kondoh, N., Koh, E., Namiki, M., Shinka, T., & Nakahori, Y. (1999). Spermatogenic ability is different among males in different Y chromosome lineage. *J.Hum.Genet.* **44**, 289-292.

Larsen, H. B. (2003). Kenyan dominance in distance running. *Comp Biochem.Physiol A Mol.Integr.Physiol* **136**, 161-170.

Larsen, H. B., Christensen, D. L., Nolan, T., & Sondergaard, H. (2004). Body dimensions, exercise capacity and physical activity level of adolescent Nandi boys in western Kenya. *Ann.Hum.Biol.* **31**, 159-173.

Larsen, H. B., Nolan, T., Borch, C., & Sondergaard, H. (2005). Training response of adolescent Kenyan town and village boys to endurance running. *Scand.J.Med.Sci.Sports* **15**, 48-57.

Leonard, W. R., Sorensen, M. V., Galloway, V. A., Spencer, G. J., Mosher, M. J., Osipova, L., & Spitsyn, V. A. (2002). Climatic influences on basal metabolic rates among circumpolar populations. *Am.J.Hum.Biol.* **14**, 609-620.

Lesage, R., Simoneau, J. A., Jobin, J., Leblanc, J., & Bouchard, C. (1985). Familial resemblance in maximal heart rate, blood lactate and aerobic power. *Hum.Hered.* **35**, 182-189.

Levine, B. D. & Stray-Gundersen, J. (1997). "Living high-training low": effect of moderate-altitude acclimatization with low-altitude training on performance. *J.Appl.Physiol* **83**, 102-112.

Maca-Meyer, N., Gonzalez, A. M., Larruga, J. M., Flores, C., & Cabrera, V. M. (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC.Genet.* **2**, 13.

MacArthur, D. G. & North, K. N. (2005). Genes and human elite athletic performance. *Hum.Genet.* **116**, 331-339.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B., & Torroni, A. (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control- region sequences and RFLPs. *Am.J.Hum.Genet.* **64**, 232-249.

Maldonado, S., Mujika, I., & Padilla, S. (2002). Influence of body mass and height on the energy cost of running in highly trained middle- and long-distance runners. *Int.J.Sports Med.* **23**, 268-272.

Manners, J. (1997). Kenya's Running Tribe. *The Sports Historian* **17**, 14-27.

Marconi, C., Marzorati, M., Grassi, B., Basnyat, B., Colombini, A., Kayser, B., & Cerretelli, P. (2004). Second generation Tibetan lowlanders acclimatize to high altitude more quickly than Caucasians. *J.Physiol* **556**, 661-671.

Marconi, C., Marzorati, M., Sciuto, D., Ferri, A., & Cerretelli, P. (2005). Economy of locomotion in high-altitude Tibetan migrants exposed to normoxia. *J.Physiol* **569**, 667-675.

Marcuello, A., Gonzalez-Alonso, J., Calbet, J. A., Damsgaard, R., Lopez-Perez, M. J., & Diez-Sanchez, C. (2005). Skeletal muscle mitochondrial DNA content in exercising humans. *J.Appl.Physiol* **99**, 1372-1377.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M. D., Sukernik, R. I., Olckers, A., & Wallace, D. C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc.Natl.Acad.Sci.U.S.A* **100**, 171-176.

Montgomery, H. E., Marshall, R., Hemingway, H., Myerson, S., Clarkson, P., Dollery, C., Hayward, M., Holliman, D. E., Jubb, M., World, M., Thomas, E. L., Brynes, A. E., Saeed, N., Barnard, M., Bell, J. D., Prasad, K., Rayson, M., Talmud, P. J., & Humphries, S. E. (1998). Human gene for physical performance. *Nature* **393**, 221-222.

Mortimer, H., Patel, S., & Peacock, A. J. (2004). The genetic basis of high-altitude pulmonary oedema. *Pharmacol.Ther.* **101**, 183-192.

Murakami, H., Ota, A., Simojo, H., Okada, M., Ajisaka, R., & Kuno, S. (2002). Polymorphisms in control region of mtDNA relates to individual differences in endurance capacity or trainability. *Jpn.J.Physiol* **52**, 247-256.

Murakami, H., Soma, R., Hayashi, J., Katsuta, S., Matsuda, M., Ajisaka, R., Okada, M., & Kuno, S. (2001). Relationship between mitochondrial DNA polymorphism and the individual differences in aerobic performance. *Jpn.J.Physiol* **51**, 563-568.

Myburgh, K. H. (2003). What makes an endurance athlete world-class? Not simply a physiological conundrum. *Comp Biochem.Physiol A Mol.Integr.Physiol* **136**, 171-190.

Myerson, S., Hemingway, H., Budget, R., Martin, J., Humphries, S., & Montgomery, H. (1999). Human angiotensin I-converting enzyme gene and endurance performance. *J.Appl.Physiol* **87**, 1313-1316.

Myerson, S. G., Montgomery, H. E., Whittingham, M., Jubb, M., World, M. J., Humphries, S. E., & Pennell, D. J. (2001). Left ventricular hypertrophy with exercise and ACE gene insertion/deletion polymorphism: a randomized controlled trial with losartan. *Circulation* **103**, 226-230.

Nazarov, I. B., Woods, D. R., Montgomery, H. E., Shneider, O. V., Kazakov, V. I., Tomilin, N. V., & Rogozkin, V. A. (2001). The angiotensin converting enzyme I/D polymorphism in Russian athletes. *Eur.J.Hum.Genet.* **9**, 797-801.

Niemi, A. K. & Majamaa, K. (2005). Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *Eur.J.Hum.Genet.* **13**, 965-969.

North, K. N., Yang, N., Wattanasirichaigoon, D., Mills, M., Easteal, S., & Beggs, A. H. (1999). A common nonsense mutation results in alpha-actinin-3 deficiency in the general population. *Nat.Genet.* **21**, 353-354.

Nummela, A. & Rusko, H. (2000). Acclimatization to altitude and normoxic training improve 400-m running performance at sea level. *J.Sports Sci.* **18**, 411-419.

Paracchini, S., Arredi, B., Chalk, R., & Tyler-Smith, C. (2002). Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucleic Acids Res.* **30**, e27.

Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M., & Santachiara-Benerecetti, A. S. (1998). Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am.J.Hum.Genet.* **62**, 420-434.

Perusse, L., Rankinen, T., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2003). The human gene map for performance and health-related fitness phenotypes: the 2002 update. *Med.Sci.Sports Exerc.* **35**, 1248-1264.

Phillips, J. C. (1976). Racial variation in sports participation. *Int.Rev.Sports.Sociol* **3**, 39-53.

Pitsiladis, Y. P. & Scott, R. (2005). Essay: The makings of the perfect athlete. *Lancet* **366 Suppl 1**, S16-S17.

Pkalya, P. & Aden, M. (2003). *Conflicts in Kenya*, pp. 8-19. ITDG, Nairobi, Kenya.

Predel, H. G., Rohden, C., Heine, O., Prinz, U., & Rost, R. E. (1994). ACE inhibition and physical exercise: studies on physical work capacity, energy metabolism, and maximum oxygen uptake in well-trained, healthy subjects. *J.Cardiovasc.Pharmacol.* **23 Suppl 1**, S25-S28.

Rajput, C., Najib, S., Norboo, T., Afrin, F., & Qadar Pasha, M. A. (2006). Endothelin-1 gene variants and levels associate with adaptation to hypobaric hypoxia in high-altitude natives. *Biochem.Biophys.Res.Commun.* **341**, 1218-1224.

Rando, J. C., Pinto, F., Gonzalez, A. M., Hernandez, M., Larruga, J. M., Cabrera, V. M., & Bandelt, H. J. (1998). Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann.Hum.Genet.* **62 (Pt 6)**, 531-550.

Rankinen, T., Perusse, L., Gagnon, J., Chagnon, Y. C., Leon, A. S., Skinner, J. S., Wilmore, J. H., Rao, D. C., & Bouchard, C. (2000a). Angiotensin-converting enzyme ID polymorphism and fitness phenotype in the HERITAGE Family Study. *J.Appl.Physiol* **88**, 1029-1035.

Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2001). The human gene map for performance and health-related fitness phenotypes. *Med.Sci.Sports Exerc.* **33**, 855-867.

Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2002). The human gene map for performance and health-related fitness phenotypes: the 2001 update. *Med.Sci.Sports Exerc.* **34**, 1219-1233.

Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2004). The human gene map for performance and health-related fitness phenotypes: the 2003 update. *Med.Sci.Sports Exerc.* **36**, 1451-1469.

Rankinen, T., Wolfarth, B., Simoneau, J. A., Maier-Lenz, D., Rauramaa, R., Rivera, M. A., Boulay, M. R., Chagnon, Y. C., Perusse, L., Keul, J., & Bouchard, C. (2000b). No association between the angiotensin-converting enzyme ID polymorphism and elite endurance athlete status. *J.Appl.Physiol* **88**, 1571-1575.

Raymond, M. & Rousset, F. An exact test for population differentiation. Evolution 49, 1280-1283. 1995.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., & Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature* **411**, 199-204.

Reilly, J. J., Jackson, D. M., Montgomery, C., Kelly, L. A., Slater, C., Grant, S., & Paton, J. Y. (2004). Total energy expenditure and physical activity in young Scottish children: mixed longitudinal study. *Lancet* **363**, 211-212.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al Zaheri, N., Santachiara-Benerecetti, S., Scozari, R., Torroni, A., & Bandelt, H. J. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am.J.Hum.Genet.* **67**, 1251-1276.

Richards, M., Rengo, C., Cruciani, F., Gratrix, F., Wilson, J. F., Scozzari, R., Macaulay, V., & Torroni, A. (2003). Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am.J.Hum.Genet.* **72**, 1058-1064.

Rieder, M. J., Taylor, S. L., Clark, A. G., & Nickerson, D. A. (1999). Sequence variation in the human angiotensin converting enzyme. *Nat.Genet.* **22**, 59-62.

Rigat, B., Hubert, C., Alhenc-Gelas, F., Cambien, F., Corvol, P., & Soubrier, F. (1990). An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J.Clin.Invest* **86**, 1343-1346.

Rivera, M. A., Wolfarth, B., Dionne, F. T., Chagnon, M., Simoneau, J. A., Boulay, M. R., Song, T. M., Perusse, L., Gagnon, J., Leon, A. S., Rao, D. C., Skinner, J. S., Wilmore, J. H., Keul, J., & Bouchard, C. (1998). Three mitochondrial DNA restriction polymorphisms in elite endurance athletes and sedentary controls. *Med.Sci.Sports Exerc.* **30**, 687-690.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., & Wallace, D. C. (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**, 223-226.

Rupert, J. L. (2003). The search for genotypes that underlie human performance phenotypes. *Comp Biochem.Physiol A Mol.Integr.Physiol* **136**, 191-203.

Salas, A., Richards, M., De la, F. T., Lareu, M. V., Sobrino, B., Sanchez-Diz, P., Macaulay, V., & Carracedo, A. (2002). The making of the African mtDNA landscape. *Am.J.Hum.Genet.* **71**, 1082-1111.

Salas, A., Richards, M., Lareu, M. V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., & Carracedo, A. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am.J.Hum.Genet.* **74**, 454-465.

Saltin, B. (1996). Exercise and the environment: focus on altitude. *Res.Q.Exerc.Sport* **67**, S1-10.

Saltin, B., Kim, C. K., Terrados, N., Larsen, H., Svedenhag, J., & Rolf, C. J. (1995a). Morphology, enzyme activities and buffer capacity in leg muscles of Kenyan and Scandinavian runners. *Scand.J.Med.Sci.Sports* **5**, 222-230.

Saltin, B., Larsen, H., Terrados, N., Bangsbo, J., Bak, T., Kim, C. K., Svedenhag, J., & Rolf, C. J. (1995b). Aerobic exercise capacity at sea level and at altitude in Kenyan boys, junior and senior runners compared with Scandinavian runners. *Scand.J.Med.Sci.Sports* **5**, 209-221.

Scanavini, D., Bernardi, F., Castoldi, E., Conconi, F., & Mazzoni, G. (2002). Increased frequency of the homozygous II ACE genotype in Italian Olympic endurance athletes. *Eur.J.Hum.Genet.* **10**, 576-577.

Schmidt, W., Heinicke, K., Rojas, J., Manuel, G. J., Serrato, M., Mora, M., Wolfarth, B., Schmid, A., & Keul, J. (2002). Blood volume and hemoglobin mass in endurance athletes from moderate altitude. *Med.Sci.Sports Exerc.* **34**, 1934-1940.

Schneider, S., Roessli, D., & Excoffier, L. Arlequin: A software for population genetics data analysis. Genetics and Biometry Lab, Dept.of Anthropology, University of Geneva Version 2.0.0.0. 2000.

Semino, O., Santachiara-Benerecetti, A. S., Falaschi, F., Cavalli-Sforza, L. L., & Underhill, P. A. (2002). Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am.J.Hum.Genet.* **70**, 265-268.

Simoneau, J. A. & Bouchard, C. (1995). Genetic determinism of fiber type proportion in human skeletal muscle. *FASEB J.* **9**, 1091-1095.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S. F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S. P., Waterston, R. H., Wilson, R. K., Rozen, S., & Page, D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825-837.

Soma, R., Murakami, H., Hayashi, J., Ito, S., Nagai, J., Sugita, M., Kuno, S., & Okada, M. (2001). The effects of cytoplasmic transfer of mtDNA in relation to whole-body endurance performance. *Jpn.J.Physiol* **51**, 475-480.

Soubrier, F., Martin, S., Alonso, A., Visvikis, S., Tiret, L., Matsuda, F., Lathrop, G. M., & Farrall, M. (2002). High-resolution genetic mapping of the ACE-linked QTL influencing circulating ACE activity. *Eur.J.Hum.Genet.* **10**, 553-561.

Speakman, J. R., Talbot, D. A., Selman, C., Snart, S., McLaren, J. S., Redman, P., Krol, E., Jackson, D. M., Johnson, M. S., & Brand, M. D. (2004). Uncoupled and surviving: individual mice with high metabolism have greater mitochondrial uncoupling and live longer. *Aging Cell* **3**, 87-95.

Stone, J., Lynch, C., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *J.Pers.Soc.Psychol* **77**, 1213-1227.

Stray-Gundersen, J., Chapman, R. F., & Levine, B. D. (2001). "Living high-training low" altitude training improves sea level performance in male and female elite runners. *J.Appl.Physiol* **91**, 1113-1120.

Sugimura, K., Tian, X. L., Hoffmann, S., Ganten, D., & Bader, M. (1998). Alternative splicing of the mRNA coding for the human endothelial angiotensin-converting enzyme: a new mechanism for solubilization. *Biochem.Biophys.Res.Commun.* **247**, 466-472.

Taylor, R. R., Mamotte, C. D., Fallon, K., & van Bockxmeer, F. M. (1999). Elite athletes and the gene for angiotensin-converting enzyme. **87**, 1035-1037.

Tesfaghiorghis, H. (1986). The growth of urbanization in Ethiopia, 1966-1984. *East Afr.Econ.Rev.* **2**, 157-167.

Tsai, H. J., Kho, J. Y., Shaikh, N., Choudhry, S., Naqvi, M., Navarro, D., Matallana, H., Castro, R., Lilly, C. M., Watson, H. G., Meade, K., Lenoir, M., Thyne, S., Ziv, E., & Burchard, E. G. (2005). Admixture-matched case-control study: a practical approach for genetic association studies in admixed populations. *Hum.Genet.* 1-14.

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W.,

Cavalli-Sforza, L. L., & Oefner, P. J. (2000). Y chromosome sequence variation and the history of human populations. *Nat.Genet.* **26**, 358-361.

van der Walt, J. M., Nicodemus, K. K., Martin, E. R., Scott, W. K., Nance, M. A., Watts, R. L., Hubble, J. P., Haines, J. L., Koller, W. C., Lyons, K., Pahwa, R., Stern, M. B., Colcher, A., Hiner, B. C., Jankovic, J., Ondo, W. G., Allen, F. H., Jr., Goetz, C. G., Small, G. W., Mastaglia, F., Stajich, J. M., McLaurin, A. C., Middleton, L. T., Scott, B. L., Schmechel, D. E., Pericak-Vance, M. A., & Vance, J. M. (2003). Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am.J.Hum.Genet.* **72**, 804-811.

Watson, E., Bauer, K., Aman, R., Weiss, G., von Haeseler, A., & Paabo, S. (1996). mtDNA sequence diversity in Africa. *Am J Hum Genet.* **59**, 437-444.

Watson, E., Forster, P., Richards, M., & Bandelt, H. J. (1997). Mitochondrial footprints of human expansions in Africa. *Am.J.Hum.Genet.* **61**, 691-704.

Weston, A. R., Karamizrak, O., Smith, A., Noakes, T. D., & Myburgh, K. H. (1999). African runners exhibit greater fatigue resistance, lower lactate accumulation, and higher oxidative enzyme activity. *J.Appl.Physiol* **86**, 915-923.

Weston, A. R., Mbambo, Z., & Myburgh, K. H. (2000). Running economy of African and Caucasian distance runners. *Med.Sci.Sports Exerc.* **32**, 1130-1134.

Wilber, R. L. (2001). Current trends in altitude training. *Sports Med.* **31**, 249-265.

Wolfarth, B., Bray, M. S., Hagberg, J. M., Perusse, L., Rauramaa, R., Rivera, M. A., Roth, S. M., Rankinen, T., & Bouchard, C. (2005). The human gene map for performance and health-related fitness phenotypes: the 2004 update. *Med.Sci.Sports Exerc.* **37**, 881-903.

Woods, D., Hickman, M., Jamshidi, Y., Brull, D., Vassiliou, V., Jones, A., Humphries, S., & Montgomery, H. (2001). Elite swimmers and the D allele of the ACE I/D polymorphism. *Hum.Genet.* **108**, 230-232.

Woods, D., Sanders, J., Jones, A., Hawe, E., Gohlke, P., Humphries, S. E., Payne, J., & Montgomery, H. (2004). The serum angiotensin-converting enzyme and angiotensin II response to altered posture and acute exercise, and the influence of ACE genotype. *Eur.J.Appl.Physiol* **91**, 342-348.

Woods, D. R., Humphries, S. E., & Montgomery, H. E. (2000). The ACE I/D polymorphism and human physical performance. *Trends Endocrinol.Metab* **11**, 416-420.

Woods, D. R. & Montgomery, H. E. (2001). Angiotensin-converting enzyme and genetics at high altitude. *High Alt.Med.Biol.* **2**, 201-210.

Woods, D. R., World, M., Rayson, M. P., Williams, A. G., Jubb, M., Jamshidi, Y., Hayward, M., Mary, D. A., Humphries, S. E., & Montgomery, H. E. (2002). Endurance enhancement related to the human angiotensin I-converting enzyme I-D polymorphism is not due to differences in the cardiorespiratory response to training. *Eur.J.Appl.Physiol* **86**, 240-244.

Y Chromosome Consortium (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339-348.

Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Easteal, S., & North, K. (2003). ACTN3 genotype is associated with human elite athletic performance. *Am.J.Hum.Genet.* **73**, 627-631.

Yu, N., Chen, F. C., Ota, S., Jorde, L. B., Pamilo, P., Patthy, L., Ramsay, M., Jenkins, T., Shyue, S. K., & Li, W. H. (2002). Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269-274.

Zhang, B., Tanaka, H., Shono, N., Miura, S., Kiyonaga, A., Shindo, M., & Saku, K. (2003). The I allele of the angiotensin-converting enzyme gene is associated with an increased percentage of slow-twitch type I fibers in human skeletal muscle. *Clin.Genet.* **63**, 139-144.

Zhao, B., Moochhala, S. M., Tham, S., Lu, J., Chia, M., Byrne, C., Hu, Q., & Lee, L. K. (2003). Relationship between angiotensin-converting enzyme ID polymorphism and VO(2max) of Chinese males. *Life Sci.* **73**, 2625-2630.

Zhu, X., Bouzekri, N., Southam, L., Cooper, R. S., Adeyemo, A., McKenzie, C. A., Luke, A., Chen, G., Elston, R. C., & Ward, R. (2001). Linkage and association analysis of angiotensin I-converting enzyme (ACE)-gene polymorphisms with ACE concentration and blood pressure. *Am.J.Hum.Genet.* **68**, 1139-1148.

Zhu, X., McKenzie, C. A., Forrester, T., Nickerson, D. A., Broeckel, U., Schunkert, H., Doering, A., Jacob, H. J., Cooper, R. S., & Rieder, M. J. (2000). Localization of a small genomic region associated with elevated ACE. *Am.J.Hum.Genet.* **67**, 1144-1153.