The Effects of Genetic Ancestry on Elite Sprint Athlete Status in the West African Diaspora

Michael Leo Deason

B.A. Hons Magna Cum Laude, University of New Mexico

Submitted in fulfilment of the requirements for
the Degree of Doctor of Philosophy

June 2017

Institute of Cardiovascular and Medical Sciences

College of Medical, Veterinary and Life Sciences

Abstract

Elite athletic performance is widely acknowledged to result from the exposure of a favourable genetic endowment to a favourable combination of environmental factors including culture, diet, training regime and socioeconomic status. Athletes from West African descendant populations in North America and Western Europe have long been prominent in elite sprint running, constituting 63% of the top 100 performers in each sprint discipline, outperforming athletes from Europe (23%), West Africa (8%) and the rest of the world (6%). These members of the West African diaspora are genetically admixed, resulting in detectable levels of both African and European ancestry because of gene flow between African slaves and Europeans during chattel slavery in the 16th to 19th centuries.

The overall aim of this thesis was to investigate the effect that ancestral genetic composition may have had on the likelihood of becoming a top-class sprint athlete amongst samples of African-Americans and Jamaicans. It was hoped that these findings would add to the existing research in attempting to understand the unique combination of factors that produce elite sprint athletes.

Ancestrally informative genetic data from elite African-American and Jamaican sprint athletes and ethnically-matched controls were used to explore how genetic ancestry affects sprint athlete status in these populations. These data are also vital when investigating the putative origins of an admixed population, and relatively little research has investigated the genetic ancestry of modern Jamaicans when compared to African-Americans.

To bring the two groups to comparable levels of insight, the population history of the Jamaican people was estimated by comparing the observed matrilineal gene pool to the gene pools of known source regions of Africa. By simulating a stable population with the observed population dynamics from slave-era Jamaica, it was possible to draw conclusions about selection acting on the Jamaican slave population from the colonisation of the island by England in 1655 until the abolition of the slave trade in 1807. In addition to the Jamaican maternal lineages already genotyped, paternal lineages in both African-Americans and Jamaicans, as well as maternal lineages in African-Americans were genotyped to assess any

association these lineages had with elite sprint athlete status. These lineages were also compared between the cohorts to assess any differences in lineage composition across both groups of athletes and controls. Finally, locus-specific genetic ancestry was calculated to map loci associated with elite athlete status to regions of the genome with a greater amount of African or European ancestry than would be expected under the null hypothesis of no association with ancestry. Assuming a difference in the likelihood of sprint athletes originating from either Africa or Europe, detected associations between locus-specific ancestry and sprint status may indicate specific genomic regions of interest.

The main findings of this thesis are:

a) Modern Jamaicans are mostly descended from slaves originating from the Gold Coast of Africa, despite large influxes of slaves from the Bight of Biafra and West-central Africa before the end of the slave trade.
b) There appears to have been selective pressure acting on the slave population of Jamaica. Differences between the presumptive origins of the observed lineages and the outcome of the stable population model suggested varying levels of mortality and fecundity within the slave population, consistent with earlier ethnographic and linguistic studies.
c) The distribution of maternal lineages in the African-American athletes were significantly different from that of African-American controls. Maternal lineage distributions between Jamaican athletes and Jamaican controls were not significantly different. There was insufficient statistical power to infer any differences between the paternal lineages of African-American athletes and controls or the Jamaican athletes and controls. This suggests that either maternal ancestry may be a factor in elite sprint athlete status for African-Americans or it could simply be a false positive, inherent to the methodology used. Jamaican maternal lineages are homogeneous with regards to elite sprint athlete status. There was insufficient statistical power to arrive at similar conclusions regarding the paternal lineages of athletes and controls in either group.
d) The maternal lineages of African-American athletes and Jamaican athletes were significantly different, although there was insufficient statistical power to determine if there were any differences between the paternal lineages of African-American athletes and Jamaican athletes. This suggests

that the same maternal lineage distribution is not associated with sprint athlete status in the two populations, while there is insufficient evidence to make a similar claim regarding paternal lineages.

e) The maternal lineages of African-American controls and Jamaican controls were also significantly different, although there was insufficient statistical power to conclude whether significant difference exists in the paternal lineages of African-American controls and Jamaican controls. These results suggest that there is some evidence that the population histories of African-Americans and Jamaicans are significantly different despite the lack of evidence from the paternal lineages.

f) The proportion of genome-wide African ancestry did not differ significantly between either African-American athletes and controls or Jamaican athletes and controls. This suggests that environmental factors typically associated with higher levels of African ancestry in these populations (e.g. lower socioeconomic status, diminished access to healthcare) are not directly linked with elite athlete status.

g) The estimated number of generations since admixture occurred did not differ significantly between athletes and controls for either African-Americans or Jamaicans. This suggests that athletes were not more likely than controls to have had European ancestors in the recent past, thereby potentially having greater access to resources.

h) Admixture mapping was used to detect an enrichment of European ancestry at chromosome 4q13.1 significantly associated with athlete status in African-Americans. There were no significant loci associated with athlete status in Jamaicans. This suggests that the regions of the genome influencing sprint athlete status may be different in the two populations, although there was insufficient statistical power to draw any meaningful conclusions from the Jamaican data.

This thesis has potential implications for future work not only explaining the disproportionate success of West African descendant sprint athletes but also for advancing the basic understanding of the genetic influences on the limits of human performance.

Table of Contents

Table of Figures

Table of Tables

Dedication

This thesis is dedicated to the memory of my grandfather, John Davidson. The gumption, hard work and insight he displayed during his life serves as constant motivation.

I am very proud to carry on his legacy.

Acknowledgements

This thesis and the work within would not have been possible without the help and support from family members, senior academics, support staff, collaborators, co-workers, and friends.

I would first like to acknowledge my parents. Without their love and support, I would never have made it this far.

I would especially like to acknowledge my supervisor Yannis Pitsiladis. He took a tremendous risk accepting a trained biological anthropologist into sports science, but our contrasting backgrounds and differences of opinion always led to improved understanding about the interaction between science and athletics.

I am also greatly indebted to Mark Bailey, Jason Gill and Vincent Macaulay for their wisdom, guidance and patience during the last 12 months of my thesis preparation. Their complimentary fields of expertise have helped me develop my ideas and to focus my research.

I am also grateful for the guidance of Avril Conacher and the Student Representative Council. Without their assistance, this thesis may never have been completed.

I would like to acknowledge those individuals responsible for securing access to, collecting samples, and processing the data I came to use in the pages below. In particular, I would like to thank Krista Austin for collecting samples from the elite athletes in the United States, as well as Prof. Errol Morrison, Rachael Irving and Vilma Charlton for sampling the elite athletes in Jamaica and Neki and Balbir Sumal for funding and contributing to the data collection in Jamaica. I would also like to thank Noriyuki Fuku and Masashi Tanaka for contributing time, expertise, and consumables.

Thanks are also due to Antonio Salas for his help interpreting mitochondrial phylogeography in the context of the Trans-Atlantic slave trade and Simon Newman for his assistance in explaining the often daunting historical record. Thanks as well to Will Goodwin for his time, expertise, and bench space at the University of Central Lancashire and Sharizah Alimat for her patience and help

with what proved to be very trying multiplex genotyping. Additionally, I would especially like to thank Sheena MacDougall, Adam Kirk-Smith, Laureen Walker, Minnie Parmiter and Robin Lupton for their help with proofreading this thesis.

Lastly, I would like to acknowledge the members of the Glasgow University Hares and Hounds, Shettleston Harriers, Glasgow City Athletics Club, as well as the Scottish Athletics community at large. Running provided me the opportunity to meet countless amazing people, as well as an all-important sense of community and opportunity to see parts of the UK that many people do not.

Author's declaration

I declare, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature: _____

Printed name: Michael Leo Deason

Glossary

Admixture	The interbreeding of two or more genetically distinct and historically separate populations.
AIMs	Ancestry Informative Markers. Genetic markers that appear at highly different frequencies across populations from different geographical regions of the world. A set of AIMs can be selected to infer the genetic ancestry of an admixed population given its presumptive source populations.
Autosome	Any of the 22 pairs of chromosomes that are not sex chromosomes (X and Y).
BLAST	Basic Local Alignment Search Tool. Found on the website for the United States' National Center for Biotechnology Information and used to isolate or compare a nucleotide or protein sequence to an organism's reference sequence to facilitate primer design (see below).
bp	Base pair. Two nitrogenous bases paired together in double-stranded DNA. Oftentimes used as a unit of measurement.
CI	Confidence Interval. An interval estimate of actual population parameters used to indicate reliability.
GWAS	Genome-Wide Association Study. These studies test for associations between many elements of whole genome sequence data and a phenotype, trait or disease.
Haplogroup	A grouping of genealogically related haplotypes (see below).
Haplotype	A combination of the words haploid and genotype. Two or more alleles at different loci which are inherited from a single parent together.
HVS-1	The first of two highly polymorphic regions of the non-coding displacement loop of the mitochondrial genome. Commonly assumed to range roughly between base pairs 16000 and 16400.
mtDNA	Mitochondrial genome, ~16,529 base pairs in length. Assumed to pass strictly along the maternal line without recombination, though point mutations do occur.
NRY	Non-recombining Region of the Y chromosome, ~58 million base pairs in length. Passed directly along the paternal line without recombination, though point mutations do occur.

OR	Odds Ratio. A statistic used to determine the overrepresentation or underrepresentation in the counts of discreet categorical variables. Often used in a case-control study design.
PCR	Polymerase Chain Reaction, a process by which a predesignated segment of DNA is amplified exponentially using a forward and reverse primer (see below).
Primer	A strand of nucleic acid that serves as a starting point for DNA synthesis in PCR.
Primer Dimer	Primers that have hybridised to one other as the result of strings of complementary bases. As a result, the DNA polymerase amplifies the dimer, leading to competition for PCR reagents, thus inhibiting amplification of the DNA sequence targeted for PCR amplification.
RFLP	Restriction Fragment Length Polymorphism, a process by which a carefully chosen cleaving enzyme cuts one allele of an amplified segment of DNA.
RefSeq	A Reference Sequence open-source database built by the United States' National Center for Biotechnology Information.
Sahel	Semi-arid region between the coastal forests of Sub-Saharan Africa and the Saharan desert.
SBE	Single Base Extension, a primer designed to assign a pre-specified molecular weight to a PCR product.
SNP	Single Nucleotide Polymorphism. A DNA marker type constituting a difference within a species.
SWGDAM	United States Federal Bureau of Investigation's Scientific Working Group on DNA analysis Methods.
UEP	Unique Event Polymorphism. An allele assumed to have resulted from a rare single mutation event. All individuals sharing the allele are assumed to have inherited it from a common ancestor.
$\dot{V}O_2max$	Rate of maximal oxygen consumption. This value reflects the aerobic physical fitness of the individual. It is an important determinant of endurance capacity during prolonged, sub-maximal exercise and is expressed as millilitres of oxygen per kilogram of body mass per minute (mL/(kg·min)).

Publications and Presentations

Publications

Deason, M., Scott, R., Irwin, L., Macaulay, V., Fuku, N., Tanaka, M., Irving, R., Charlton, V., Morrison, E., Austin, K. & Pitsiladis, Y. 2012. Importance of mitochondrial haplotypes and maternal lineage in sprint performance among individuals of West African ancestry. Scandinavian Journal of Medicine & Science in Sports, 22, 217-223.

Deason, M., Salas, A., Newman, S., Macaulay, V., Morrison, E. & Pitsiladis, Y. 2012. Interdisciplinary approach to the demography of Jamaica. BMC Evolutionary Biology, 12, 24.

Presentations

Deason, M., Macaulay, V., Pitsiladis, Y. Assessing fertility and fecundity in slave groups of the Caribbean from mtDNA sequence variation. 2010. Annual Meeting of the American Association of Physical Anthropologists. Albuquerque, New Mexico, USA.

Deason M., Wang G., Fuku N., Mikami E., Scott R., Irwin L., Irving R., Charlton V., Morrison E., Austin K., Tladi D., Headley S., Kolkhhorst F., Yamada Y., Tanaka M., Pitsiladis Y. 2012. Genetic Ancestry and Elite Sprinting in groups of West African descent. International Convention on Science, Education and Medicine in Sport. Glasgow, UK.

1

General Introduction

Elite athletic performance is widely acknowledged to result from the exposure of favourable genetic endowment to a favourable environment, though the contribution both sets of factors make to overall performance is debateable (Tucker & Collins, 2012). Previous research has identified numerous environmental variables influencing elite athletic performance, including culture (Onywera *et al.*, 2006), diet (Grandjean, 1997), formal training (Davids & Baker, 2007) and socioeconomic status (Sohi & Yusuff, 1987). Candidate gene research has thus far shown that any genetic factors influencing elite athlete status are likely to be population-specific (Rankinen *et al.*, 2010). Sprint athletes from the West African diaspora have been prominent in elite sprint running since the 1930s and often outperform athletes from other parts of the world, averaging 63% of the top 100 performers in the 100m, 200m, 400m, 100m/110m hurdles and 400m hurdles (International Association of Athletics Federations, 2016a). Sprint running requires a great amount of formal training to achieve the strength and coordination necessary for elite performance (Daley *et al.*, 2007; Coh *et al.*, 2010), although inherent genetic factors affecting an individual's maximum performance potential are likely to be extremely influential (Tucker *et al.*, 2013; Lombardo & Deaner, 2014).

Populations of West African descent such as African-Americans and Jamaicans commonly have detectable amounts of African and European ancestry because of gene flow during the slave era between the 16th and 19th centuries. Any direct influence of slavery and the Trans-Atlantic slave trade on athletic performance has never been addressed, although a possible connection is often acknowledged by the popular media (Kane, 1971; Entine, 2000; Hughey, 2014). An improved understanding of the genetic make-up of populations with mixed ancestry has already shown promise in the field of biomedical genetics. By calculating

associations between genetic ancestry at a region of the genome and a complex disease, it is possible to identify regions harbouring influential tracts of DNA based on overrepresentation of a given ancestry (Winkler *et al.*, 2010).

The overall aim of this thesis is to investigate the association between ancestral genetic composition and sprint performance using samples from African-Americans and Jamaicans. It is hoped that these findings will add to the existing research in attempting to understand the unique combination of factors that produce elite sprint athletes.

1.1 The West African diaspora in sprint athletics

At the recent 2016 Olympic Games held in Rio de Janeiro, individuals of the West African diaspora won 25 of the 30 available individual sprint medals. The remaining five medals were won by three Europeans, a South African and an East African (International Olympic Committee, 2016).

African-Americans and Jamaicans have shown an increased dominance in sprint events recently. In 2016, all the United States' 14 Olympic medals in sprint events and relays were won by athletes from the African-American community, and Jamaica won 11 Olympic medals in sprint events. Combined, the two groups won 25 of 42 available sprint medals. Both groups have histories of sprint athletics, although both nations developed their strong sprint heritage in separate ways.

1.1.1 History of elite sprinters from the African-American community

The United States has a long tradition of athletics. The Intercollegiate Association of Amateur Athletes of America began organising competitions between universities in 1873. In 1888, the Amateur Athletic Union held the first national championship in athletics (Sears, 2001). Although the United States as a nation has been a dominant force in world athletics since the inception of the first modern Olympic Games in 1896 (USA Track and Field, 2016b), the first African-American athlete to be deemed the fastest man in the world was Eddie Tolan, setting the world record for 100 metres in 10.25 seconds in 1930 (Sears, 2001). The African-American community have dominated sprint running events not only within the United States, but also abroad, with 361 Olympic medals in events at

or below 400 metres (USA Track and Field, 2016b) and 179 medals across the sprint disciplines since 1983 in the World Championship of Athletics (USA Track and Field, 2016a). Despite the heritage of success, many gifted male sprint athletes in the United States opt instead for more financially lucrative team sports, including American football. Although Olympic sprinters attract a lot of attention prior to and during the Olympic Games, interest quickly wanes afterwards, resuming only during the build-up to the next Olympics (Pitsiladis *et al.*, 2011).

1.1.2 History of elite sprinters from Jamaica

Jamaica has a shorter history of developing sprint athletes. Dating back to 1910, the times recorded by athletes at the prestigious Boys' "Champs" for school-aged boys were already competitive with the leading times in the world (Bertram, 2012). Without international affiliation, however, there was no opportunity for international competition and enthusiasm to train and compete among senior athletes quickly vanished after graduation. This changed in 1930 when Jamaicans began to race internationally, and by 1948 they entered their first Olympic games (Bertram, 2012). Since then, Jamaica has won 77 medals at the Olympics in athletics (Jamaica Olympic Association, 2016) and 110 medals at the World Championship of Athletics, the fifth most internationally (International Association of Athletics Federations, 2016b). Sprint athletics continues to be a popular pastime with the "Champs" regularly drawing crowds of over 30,000 (Robinson, 2008). An excellent system is now in place to identify and nurture young athletes through to the senior level. The greatest international success has been found by encouraging exceptional home-grown talent to live and train in Jamaica, thereby influencing the next generation of elite sprinters (Robinson, 2008).

1.2 Functional elements of sprint running

The primary objective of a sprint race is to cover the designated distance in the shortest time possible. While technique, technology and weather can play a substantial role in the outcome of a sprint performance (Majumdar & Robergs, 2011a), there are many physiological factors influencing performance. These can be briefly introduced as physiological and biochemical factors.

1.2.1 Physiological factors of sprint running

Sprint running is a complex exercise where the entire body engages in movement. Efficient interactions between agonist, antagonist and synergist muscles during movement are key to optimal sprint performance (Mero *et al.*, 1992). Movement coordination also has a substantial impact on overall efficiency and power generated during running (Kiely & Collins, 2016). Skeletal muscle fibres are divided into two main types that are determined early in development based on gene expression programmes (Schiaffino & Reggiani, 2011) and the distribution of these types is minimally affected by training (Costill *et al.*, 1976). Type I fibres, or oxidative muscle fibres, are rich in energy-producing mitochondria and very resistant to fatigue, capable of producing repeated low power contractions. Type II fibres, or fast twitch fibres, have fewer mitochondria and myoglobin to allow for faster adenosine triphosphate (ATP) breakdown and thus faster, more forceful contractions, though only for a short period. Type II fibres can be further subdivided into moderate fatigue resistant (IIa) and low fatigue resistant (IIb/x) subtypes, and while targeted training does not affect the composition of type I or type II muscle fibre types an individual possesses, these type II fibre subtypes are relatively plastic and can change in their composition (di Prampero *et al.*, 2005). Muscle fibre size is greatly affected by age and training, and while larger fibres correspond to greater strength, the ability to generate force does not grow linearly with cross-sectional muscle area. This is explained by improvements in the neuromuscular system as the result of training to recruit and activate a greater number of muscle fibres in an optimal sequence to maximise performance (Ross *et al.*, 2001).

1.2.2 Biochemical factors of sprint running

All cells use the molecular compound ATP to generate energy. The total amount of freely available ATP in a cell produces only approximately three seconds of work (Baker *et al.*, 2010). For a cell to work continuously, there are three systems used to regenerate ATP: mitochondrial respiration, the glycolytic system and the phosphagen system. While mitochondrial respiration is responsible for meeting longer duration demands on muscle cell ATP, the glycolic system and phosphagen system are most relevant to sprint performance. Together, these systems meet the energy demands of a 10-second maximal sprint by utilising 53% phosphagen,

44% glycolysis and 3% mitochondrial respiration, and a 30-second sprint with 23% phosphagen, 49% glycolysis and 28% mitochondrial respiration (Van Someren, 2006).

Mitochondrial respiration involves the use of oxygen in mitochondria to burn of different fuel sources for energy, a process referred to as oxidative phosphorylation. These fuel sources include free fatty acids and glycogen from the muscle, free fatty acids from adipose tissue and glucose from the blood or liver (Baker *et al.,* 2010). Mitochondrial respiration can resynthesise ATP if fuel and sufficient oxygen are available, although the amounts of energy generated are not sufficient for the explosive demands produced by actions such as sprint running. The efficiency of mitochondria to generate ATP has been suggested to be an important determinant for performance in elite athletes (Ahmetov & Fedotovskaya, 2015).

When exercise last longer than two seconds, the glycolytic system is utilised to regenerate ATP (Pilegaard *et al.,* 1999). The glycolytic system relies on blood glucose or muscle glycogen as a fuel source and regenerates ATP much more quickly than mitochondrial respiration (Robergs *et al.*, 2004). The glycolytic system can be broken into two phases: the first is ATP costly, though the second takes the product of the first and generates a net ATP gain. The number of reactions involved in this metabolic pathway means that ATP is not generated as quickly in the glycolytic system as in the phosphagen system (Baker *et al.,* 2010), though as stated above, the glycolytic system is vital to ATP regeneration during a sprint performance.

The dominant metabolic energy system for sprint performance relies on energy-storing compounds found in muscle and nervous tissue known as phosphagens. The most common phosphagen in humans is creatine phosphate (Baker *et al.*, 2010). To replenish the levels of ATP, high energy compounds known as creatine phosphates are broken down by the enzyme creatine kinase into creatine and phosphate molecules. The phosphate group is transferred to freely available adenosine diphosphate molecules to form ATP once again. Thus, if creatine phosphate remains available, ATP is regenerated at a very high rate and muscle ATP is maintained at a moderately constant level. However, the phosphagen system can only meet the energy demands of intensely contracting muscle for a

short time, reaching maximum energy output after 1.3 seconds (Maughan *et al.*, 1997) and only sustaining energy production to approximately 10 seconds before muscle cells deplete (Cheetham *et al.*, 1986).

1.3 Genetics of sport

Genetics play a key role in how effective athletes are at generating and utilising ATP (Rankinen *et al.,* 2000). Unsurprisingly, there has been a substantial amount of association-based research conducted into the genetics behind these and other factors influencing sprint performance (Ahmetov & Fedotovskaya, 2015). Success in sport is commonly believed to be an interaction between genetic and environmental determinants (Tucker & Collins, 2012). While some researchers have suggested that performance-intensive and deliberate practice can activate dormant performance-related genes in almost every person (Ericsson *et al.*, 2009), the consensus amongst sports scientists is that a variable genetic endowment among athletes results in a genetically predetermined ceiling to performance for every individual (Tucker *et al.*, 2013). The speculative contribution of genetic in the form of innate ability and training for six hypothetical athletes with varying levels of performance at first exposure is shown in Figure 1.1.

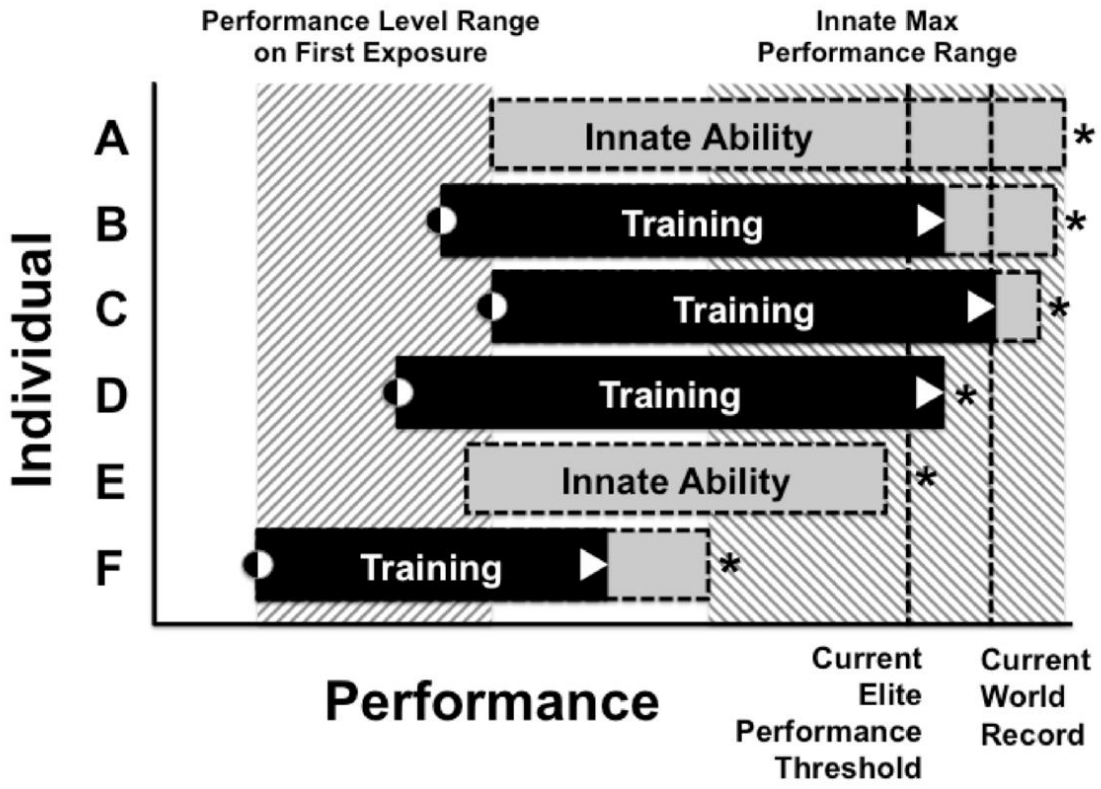


Figure 1.1 Theoretical combined effects of genetics and environment on the level of performance attainment for six hypothetical individuals. Figure reproduced from Tucker and Collins (2012). Individuals A and E are not athletes and never took up the sport, while individuals B, C, D and F are all actively training. Performance level at first exposure is indicated as a shaded region to the left of the figure, while the innate maximum performance range is to the right of the figure. The latter indicates the range at which an individual, were they ever to have taken up the sport, would attain given their genetic potential. Black and white circles indicate initial performance level, while white triangles indicate current performance level. The asterisks indicate the maximum performance level attainable given their genetic potential. Dashed vertical lines indicate the current elite performance threshold and the current world record.

Early attempts to identify genetic variants associated with athlete status were conducted as candidate gene studies. While these studies reported statistically significant associations between sampled athletes and controls across a variety of genetic markers, many did not account for the inflation of type I error due to unacknowledged multiple testing. By not taking this multiple testing error into account, these studies were unable to state with confidence whether any detected associations were true associations or merely a false positive. The genes associated with sprint and power athlete status with three or more published studies are *ACE*, *ACTN3*, *AMPD1* and *HIF1A* (Ahmetov & Fedotovskaya, 2015). However, as the number of investigations began included a broader sampling of world populations, these gene associations came under scrutiny. In addition to issues with statistical power, these associations were not found in a study of elite

African-American and Jamaican athletes, two of the most competitive groups in the world (Scott *et al.*, 2010).

Most pertinent to the genetic aspect of muscular strength and power is the alpha-actinin skeletal muscle isoform 3 gene, *ACTN3*. Expression of the alpha-actinin-3 protein is limited to fast-twitch muscle fibres responsible for generating force at high velocity. A common variant in this gene is the replacement of an arginine (R) with a stop codon at position 577 (X). The X/X genotype is associated with deficiency of alpha-actinin-3, and was first reported as having a disadvantage for sprint and power-related athletes (Yang *et al.*, 2003). The amount of alpha-actinin-3 produced is directly related to the number of R alleles present in the gene (MacArthur *et al.,* 2007). Many studies have been able to show a dearth of X/X individuals in elite power and sprint-related sports across a wide range of ethnicities in Europe (Niemi & Majamaa, 2005; Druzhevskaya *et al.*, 2008; Papadimitriou *et al.*, 2008; Eynon *et al.*, 2009; Wang *et al.*, 2013) and Asian (Mikami *et al.*, 2014). The systematic review of 18 studies invesitaging the *ACTN3* gene and sprint/power athlete status found a siginificant overrepresentation of the R/R genotype (OR = 1.21; 95% CI [1.03-1.42]; Ma *et al,* 2013).

In 2008, Roth *et al.* compared the frequency of *ACTN3* genotypes in a group of African-American and European-American elite-level bodybuilders and strength athletes with ethnically-matched controls to test for an association between the X/X genotype and athlete status. The authors used a one-sided χ^2 test to compare the total genotype distributions between groups. There was a statistically significant difference between the European-American athletes (X/X = 9.6%; *N* = 52) and controls (X/X = 19.9%; *N* = 668), although there was no significant difference between African-American athletes (X/X = 0.00%; *N* = 23) and controls (X/X = 4.8%, *N* = 208). The authors conclude that the X genotype is underrepresented in elite strength athletes and consistent with the theory that a deficit of alpha-actinin-3 impairs muscle performance, although admit that additional work is necessary to better address the influence of alpha-actinin-3 on specific muscle phenotypes that are most affected by the deficiency.

Similarly, there was no association between athlete status and *ACTN3* genotype in samples of elite African-American and Jamaican sprint athletes that also appear in this thesis (Scott *et al.*, 2010). The frequency of the X/X genotype in both

sample populations was very low in athletes and controls: 2 of 113 African-American athletes, 7 of 190 African-American controls, 3 of 114 Jamaican athletes and 6 of 390 Jamaican controls. The authors found no significant differences between the genotype distributions in African-American athletes and controls (P = 0.63) or Jamaican athletes and controls (P = 0.81). The authors also found no evidence for any difference between the sex of the athletes and controls in African-Americans (male, P = 0.88; female, P = 0.45) or Jamaicans (male, P = 0.49; female, P = 0.84). Neither associations nor differences were found because of poor statistical power, given the relatively small sample sizes and very low X/X genotype frequencies.

While these studies sought to test the hypothesis of an association between elite athlete status and genotype at candidate loci, an alternative approach removes the assumptions about the functional causality of the chosen polymorphism and instead compares the entire genome for a hypothesis-free test of genetic association. This genome-wide association study (GWAS) technique has proven enormously successful in identifying novel genetic variants in a wide variety of sport-related and other traits including total energy expenditure and maximal heart rate (Comuzzie *et al.*, 2012), lung function (Ong *et al.*, 2013) and comprehensive strength and appendicular lean mass (Han *et al.*, 2012).

An important step before beginning a GWAS is to quantify the proportion of total observed variation in a quantitative phenotype in a population that can be attributed to genetic variation, an estimate known as heritability. Heritability can be calculated by a parent-offspring regression, a model which calculates the amount of the observed variation in the offspring by regressing the observed phenotype of the offspring against the average of the observed phenotype of the parents. Similar calculations can also be made by regressing the quantitative phenotype of monozygotic and dizygotic twins against one another (Wray & Visscher, 2008). Heritability can also be estimated in a group of unrelated individuals by calculating the total variance in a phenotype that can be described by all single nucleotide polymorphisms (SNPs) under an additive model controlling for confounding variables such as sex and age (Yang *et al.*, 2010).

Heritability can range from 0 to 1. Heritability is 0 in traits that are entirely attributed to environment, such as adherence to a specific religion (Eaves *et al.*,

1990) and any traits that have reached genetic fixation, such as bipedalism in humans. Heritability is 1 when any variation in the observed traits is entirely caused by genetics. Though no observable phenotype is entirely free from environmental factors (e.g. trauma), traits such as fingerprint ridge count have been shown to have heritability estimates between 90 and 95% (Medland *et al.*, 2007). Heritability estimates for athlete status in a study of 4,488 British adult monozygotic female twins suggested that heritability of athlete status was 66% across a broad range of sports (95% CI [59 – 71%] (De Moor *et al.*, 2007). With these estimates, it is possible to compare the amount of heritability explained by common variation in the genome with individual SNP heritability and observed phenotypic variation.

For a GWAS to compensate for the inherent rise in false-positive detection rates introduced by potentially conducting over one million statistical tests, sample sizes need to be very large and significance thresholds set very low (P-value < 5×10^{-8}) to reliably detect any significant effects. Despite identifying over 700 significant associations, GWAS conducted on highly heritable quantitative traits such as height can only explain roughly 45% of the variance when all genotyped SNPs are taken together (Visscher *et al.*, 2010; Yang *et al.*, 2010). A GWAS by Bouchard *et al.* (2010) identified a set of 21 SNPs accounting for 49% of the variance in trainability to improve the maximum rate of oxygen uptake. The missing heritability in these and other traits is thought to result from not only common variants with very low and practically undetectable effect sizes, but also rarer variants with potentially larger effect sizes that were not genotyped (Gibson, 2012). Missing heritability can partially be overcome by creating custom chips homing in on candidate sites highlighted during an initial GWAS, however unfeasibly large sample sizes are often needed to detect SNPs with very small effect sizes.

There is only a single recorded GWAS of elite sprint athlete status. In that study, the author found no evidence that any loci were associated with athlete status at the strict genome-wide significance levels when comparing African-American, Jamaican and Japanese sprint athletes with ethnically-matched controls (Wang, 2013). The African-American and Jamaican samples used in that study are also used in this thesis. Unfortunately, the author did not estimate the genetic

component of athlete status by calculating the heritability in these populations, although such calculations would be possible given the available data.

1.4 Environmental aspects of sport

An athlete's developmental environment has a great impact on their eventual peak level of performance. As shown in Figure 1.1 above, optimal environmental factors are required for elite performance. Formal training makes up a substantial portion of that environment. Formal training is a form of deliberate practice, defined as the engagement with full concentration in a training activity designed to improve an aspect of performance with immediate feedback, opportunities for gradual refinement by repetition and problem solving (Ericsson, 2007). While commencing with formal training earlier in life has been shown to result in a greater level of attainment (Vaeyens *et al.*, 2009), this can vary greatly from individual to individual (Elferink-Gemser *et al.*, 2011). Sprint athletes later classed as elite did not begin formal training significantly earlier than athletes later classed as non-elite (Macnamara et al., 2016). The same research also found that variation in the number of years in formal training only influences 1% of the performance outcome in these elite sprint athletes. A study of former Olympic sprint champions found that they were already faster than 95%-99% of their cohorts when began formal training and only took a median of 3 years deliberate practice to reach world-class levels of performance (Lombardo & Deaner, 2014).

Sport is seen by some as offering a level playing field, where those with sufficient talent and commitment can succeed, regardless of their race or socioeconomic status (Kahn, 2000). Evidence that socioeconomic status influences elite performance is mixed. In Nigeria, elite athletes from 15 disciplines were shown to have had a lower socioeconomic status than non-elites (Sohi & Yusuff, 1987). By collecting interview data from these elite athletes, the authors concluded that the potential for upward social mobility was a major driving force in these athletes' pursuit of elite-level sport. However, in a similar study most elite English athletes were shown to be of higher socioeconomic status than the general population, though relationship was largely sport-dependent (Collins & Buller, 2003). These results suggest parallel systems incentivising the attainment of elite athlete status. While concentrating on sports performances may be more

financially lucrative than pursuing work for young athletes in areas of lower socioeconomic status, young athletes from more wealthy areas can exploit their additional resources for access to better coaches, training facilities and competitions.

Societal behaviour is another facet of the environment that may influence the development of elite athletes. Stereotype threat is a concept from sociology that refers to situations where people feel they are at risk of conforming to a stereotype about their social group. This effect is measurable whether the individual subscribes to that belief or not, since the mechanism appears to work by increasing anxiety and depleting working memory (Beilock *et al.*, 2007). For example, in a study assessing stereotype and sports performance, European-Americans performed a golf-related task significantly better than African-Americans when the task was framed by the experimenter in the context of sports intelligence, defined as one's ability to think strategically while playing a sport; however, when the task was framed by the experimenter as a diagnostic of natural athletic ability, defined as one's genetically determined physical gifts, European-Americans performed the same task significantly worse than their African-American counterparts (Stone *et al.*, 1999). A related concept, stereotype boost is the observable boost in performance when an individual is primed with a positive stereotype for their social group (Shih *et al.*, 2002) and could potentially explain ethnic differences in elite sprint performance. Because of stereotype boost, more young African-Americans and Jamaicans may take up sprint running because they believe they are inherently good at it.

The observed genetics of a population reflect their population history. Slavery is part of the population histories in African-Americans and Jamaicans. By understanding the motivations for relocating Africans to the new world, the conditions in which they lived, the environmental pressures they were exposed to and the gene flow that would have occurred, clues may be found to help explain the genetics of these unique populations that produce a disproportionate number of elite-calibre sprint athletes. In the next section, a brief history of slavery in the United States and Jamaica is presented.

1.5 Slavery and Africa

It could be argued that the Trans-Atlantic slave trade and the environment of chattel slavery may have contributed to a unique genetic make-up resulting in a higher sport-dependent genetic ceiling for sprint athletics of African-Americans and Jamaicans. To understand the implications of the Trans-Atlantic slave trade on the West African diaspora, it is important to introduce not only Africa between the 15th and the 19th centuries, but also the slave trade itself. To understand the populations these athletes arise from, certain key factors need to be considered. These include the history and motivation behind the enforced movement of their African ancestors to the Americas and the Caribbean, the potential demographic selection pressures involved in slavery and any subsequent interbreeding resulting in genetic admixture.

1.5.1 Trans-Atlantic slave trade

An intra-African slave trade was present and commonplace throughout much of the continent prior to the onset of the fourteenth century. Slaves were often taken as the result of military conquests, and a bustling trans-Saharan slave trade sprang up running along busy trade routes between the pastoral groups of the Mediterranean, the Sahel and the West African coast (Lovejoy, 1994). Due to the intra-continental trade networks and few naturally sheltered river ports (Hilling, 1969), coastal West Africa was sparsely inhabited at the time of European contact. The introduction of large sea-based vessels created the incentive to establish port towns to exploit these new trade opportunities and cater to ever-larger ships (Lovejoy, 1983). Slaves were a common trading commodity, although the discovery of the Western Hemisphere and the rise in demand for luxury agricultural produce such as sugar and tobacco in Europe led to the largest systematic relocation of people in world history (Eltis, 1983).

The political climate of the African mainland made the establishment of plantations reliant on slave labour unfeasible. Sugar plantations were initially established on sparsely population islands along the coast of Africa including the Canary Islands and modern-day São Tomé to minimise the disruption from local people (Behrendt *et al.*, 2001). The climate and soil composition of these volcanic islands were not ideal for growing sugar, and as the demand for sugar rose in

Europe, it became more economically feasible to ship African labour to American and Caribbean plantations. Additionally, new plantations possessed a clean-slate advantage and new planters were not only able to utilise nutrient-rich soil and resources like timber, but also to plan their plantations to specifications that would maximise profit utilising modern technology (Moore, 2003). An Atlantic triangular trade network soon developed seeing sundry trade items such as textiles, weapons and rum shipped from the Northeastern United States and Europe to Africa. These goods would then be bartered for African slaves who were then shipped to the plantations of the Americas and the Caribbean. Sailors then unloaded the slaves from the ships and replaced them with agricultural staples to send to Europe and the Northeastern United States to be processed into trade goods, many which would ultimately head to African ports, completing the trade cycle (Williams, 1994).

1.5.2 The Middle Passage

The Middle Passage refers to the transportation of African slaves from Africa to the Americas. Depending on the destination and departure locations and environmental factors such as weather conditions, the median journey lengths were between 33 days to sail from West-central Africa to Rio de Janeiro and 107 days to sail between the Bight of Benin to modern day Haiti, then the French colony Saint-Domingue (Haines *et al.*, 2001). Of the estimated 12.5 million Africans to have been transported during the Trans-Atlantic slave trade, only 85.5% are recorded as having reached the New World (Eltis, 2009). Describing the conditions of the slave ships as uniformly deplorable by assigning a fixed mortality rate, however, would be ignoring variability in trading practices between nations and through time. The combined effect of conditions immediately before leaving Africa and conditions on ships heavily influenced slave mortality. These conditions varied greatly through time, dependent on a host of aggravating factors such as provision miscalculations, disease outbreaks, number of stopovers and levels of overcrowding (Haines *et al.*, 2001).

Captured Africans, whether taken as prisoners during wartime or in slaving raids, were highly valued in the slave trade. Regardless of their destination, these people were often housed in what came to be known as slave factories or coastal

barracoons, built to house the slaves before they were sold. Even after their sale, people were often kept on board the ship for weeks before leaving Africa, sailing between neighbouring ports to fill slave quotas in a process known as coasting (McGowan, 1990). It has been suggested that a large proportion of slave mortality during the Middle Passage was attributable to diseases contracted before the slaves had left Africa as opposed to the squalid conditions encountered in the slave ships (Haines et al., 2001). Conditions inside these ships, however, would have been conducive to the spread of infectious diseases and increased mortality risk (Kiple & Higgins, 1989). The variability between ships from the European slaving nations exemplifies the stark differences in policy each nation adopted regarding the shipment of slaves. As the trade matured, treatment improved and amelioratory measures were put in place to maximise the investment of the ships' owners. Slaves that died during the journey were deemed a commercial loss, and individuals arriving in a poor state of health would often be sold for less than market price (Burnard & Morgan, 2001).

Given the accounts above, the ports of embarkation and disembarkation played a significant role in determining the mortality of the slaves aboard. The length of time slaves spent in a physiologically taxing and deadly environment was directly linked to the overall distance and route taken between the ports of embarkation and disembarkation (Haines *et al.*, 2001). Once slaves had been loaded, British ships could not simply head westwards from ports in modern-day Sierra Leone or Ghana to the Caribbean and North America, but rather had to travel east along the West African coastline before heading south to catch the northern equatorial trade winds to sail to across the Atlantic. Though this route was the most time efficient, the long journey would have greatly impacted the slave mortality rate. Portuguese ships, on the other hand, could leave from Angola and sail directly to Brazil in much less time by taking advantage of the southern equatorial trade winds, ultimately reducing the slave mortality on their ships (Haines *et al.*, 2001).

1.5.3 Acclimatisation of newly arrived Africans to slavery

The acclimatisation process for the survivors of the Middle Passage was largely informed by the belief that keeping one's humours in equilibrium was vital to good health (Smith, 2015). Both newly arrived Africans and Europeans were thought to

require a period to acclimatise to a new environment to bring the humours back into alignment by contemporary medical doctors. However, the process for newly arrived African slaves also involved the physical and psychological acclimatisation to life in slavery and was characterised by high levels of mortality (Sheridan, 1985; Wood, 1996). Acclimatisation was assumed to be a fixed period, between one to three years, where slaves would be given a relatively lighter workload to help in the transition. Newly arrived slaves were typically given clothes and assigned an elder slave from their native country when possible to ensure they knew how to grow their own provisions, build a dwelling and generally to make the acclimatisation process as smooth as possible (Smith, 2015).

African-born slaves were especially vulnerable to new diseases. In Collins' guide to slave and plantation management (Collins, 1971), originally published in 1803, he identified six causes for over 25% of all acclimatisation deaths: infectious disease originally contracted in Africa including dysentery, yaws, and Guinea worms; change of climate resulting in fluxes (diarrhoea and dysentery) and dropsies (oedema); food that was defective in quality, deficient in quantity, or differed from dietary patterns in Africa; labour demands; complications from punishment and suicide as a result of despondency. Overall, one in three slaves was thought to have died during the acclimatisation process (Sheridan, 1985). Those slaves born into slavery were relatively immune to the American disease environment (Smith, 2015).

1.5.4 Slavery and its aftermath in the United States

The development and adoption of slavery in the North American colonies came gradually. Initially, indentured servants from Western Europe were coerced into immigrating to alleviate the labour shortage present in the colonies (Galenson, 1984). The first Africans introduced to the British North American colonies were treated as indentured servants, given their freedom after a fixed period of labour and given a parcel of land (Vaughan, 1972). While the first slaves were traded in Virginia in 1619 (Sluiter, 1997), it was not until 1644 that the first servant was recognised as chattel, and in 1662, Virginia became the first colony to adopt legal measures establishing the heritable nature of slavery by dictating that the status of newborn children was to be the same as that of their mother (Berlin, 2003).

The more temperate climate of the Mid-Atlantic regions of Chesapeake Bay, Virginia and North Carolina were ideal for the large-scale farming of tobacco, a staple crop that the indigenous Americans had grown for centuries (Pego *et al.*, 1995). Indentured labourers were brought in to cultivate this luxury good, and as demand for tobacco grew overseas, so did the demand for labour. Slaves were purchased to fill the labour shortage around Chesapeake Bay and its hinterland as the number of plantations grew (Galenson, 1984). Preparing the land for the crop required relatively little effort, and after the plants matured, individual leaves were cut, bundled, and transported to a drying house. Despite employing gang labour, the foreman dedicated to setting the pace was aware that cultivation was a delicate art and treated his labourers with due diligence which also encouraged quality control (Kulikoff, 1986).

While the area around Chesapeake Bay had few large urban markets, smaller local markets were more prominent because their proximity enabled owners to keep their transportation and operational costs low (Walsh, 1989). This presumably also fostered increased contact between slave owners and their slaves. In South Carolina, Georgia and westward to Texas, the cultivation of rice and cotton were more standard. Slaves cultivating both crops were managed under a system known as "tasking" in which an individual was given a parcel of land to manage themselves with very little supervision (Berlin, 2003). Quality was measured in terms of total weight, devoid of any craftsmanship. As opposed to the area around Chesapeake Bay, the Deep South consisted of few local markets and instead gravitated to urban centres for the sale of the crops (Genovese, 1989). Rice and cotton plantations were built on a very large scale and often developed self-sufficient institutions to maximise profits in the cash crop markets (Egerton, 1996). As a result, these vast plantations led to a great deal of disconnection between the slave owner and slaves (Berlin, 2003).

North American planters purchased an estimated 300,000 slaves before the abolition of the slave trade in 1808. Though still suffering from high levels of mortality, slaves in North America reproduced at levels significantly above replacement. In 1850, the slave population grew more quickly than did either Europeans or European-Americans (Tadman, 2000), and by 1860, the North American slave population had reached 2,762,398 (United States Bureau of the Census, 1975). After the emancipation of the slaves in 1865 in the wake of the US

Civil War, African-Americans found themselves left with few necessary life-skills and inculcated with resentment towards their former owners. Many African-Americans began their new lives as sharecroppers, farming small plots of land to generate produce for sale at local markets. Many also left the South for the more industrialised cities in the North-East, Midwest and California in a period termed the Great Migration (Farley, 1968; Johnson & Campbell, 1981). While Europeans had been migrating to the northern towns and cities of the United States since the seventeenth century, the rate of influx exploded during the Industrial Revolution (Tanner & Reiff, 1995) putting them in direct competition with the migrating African-Americans (Eyerman, 2001).

Despite the rise in the overall wealth of the United States and social advances in the past 50 years, the African-American community is still largely marginalised. While making up 13.6% of the total population in 2010 (Rastogi & United States Bureau of the Census, 2011), they made up 27.4% of those living in poverty (DeNavas-Walt *et al.*, 2012). Although African-Americans are overrepresented in the world of sprint athletics, they are 50% more likely to be obese than non-Hispanic Whites. African-American women are also 80% more likely to be obese than non-Hispanic White women (National Center for Health Statistics, 2013). Likely a result of the high prevalence of obesity, African-Americans are disproportionately affected by many largely preventable non-communicable diseases such as cardiovascular disease, hypertension and stroke (Cooper *et al.*, 2000). African-American women are also 60% more likely than non-Hispanic White women to have high blood pressure (National Center for Health Statistics, 2013). The 2010 Summary Health Statistics for US Adults (Schiller *et al.*, 2012) report that African-Americans were 30 percent more likely to die from heart disease when compared to non-Hispanic Whites.

1.5.5 Slavery and its aftermath in Jamaica

Jamaica was captured by the English in 1655, and the conquering soldiers became the first English colonisers. Finding the island to be poor in precious metals, early settlers began pursuing a diversified agricultural economy. Available labour was lacking and early attempts at importing sustainable European labour to the Caribbean proved to be a failure (Beckles, 1989). Rising social unrest in England

and the draw of cultivating profitable crops for sale in Europe drew the initial English settlers to Jamaica (Burnard, 1996), and planters from Barbados were also incentivised to colonise the island (Dunn, 2000). Luxury agricultural produce such as tobacco, indigo, cotton, cocoa and ginger could all be grown by the settlers themselves with only modest investment in both equipment and labour. Most planters, however, eventually opted to cultivate sugar cane. Jamaica quickly adopted chattel slavery and sugar production, a trend that was already proving profitable in the Caribbean and firmly established in the British stronghold of Barbados. By the beginning of the eighteenth century, Jamaica was the largest sugar producer in the world (Higman, 1976).

Such a feat was only made possible by the exploitation of slave labour, although Jamaica was always seen as peripheral for wealth production in the British Empire. As a direct effect of this, Jamaica never experienced the broad demographic movement seen in colonies in North America or Barbados (Burnard, 1996). A culture of absenteeism was established by the investing planters by simply employing an attorney to manage their estates and agreeing to pay him a percentage of the total profits. These hired men were often only interested in maximising the profit for themselves and had no wider interest in the sustainability or treatment of land, equipment or slaves (Burnard, 2004). European overseers, managers and other plantation personnel also exploited the slaves sexually, resulting in a modest number of mixed race slaves (Burnard, 1998; Morgan, 2006).

The British Trans-Atlantic slave trade ceased in 1807, although slavery was permitted to continue with the estimated 350,000 slaves remaining on the island (Higman, 1976). The squalid conditions that the slaves were forced to live and work in resulted in very high rates of mortality (Sheridan, 1985). Because of this, low levels of reproduction and a large influx of slaves in the final years of the slave trade, roughly 35% were African-born in 1807 (Higman, 1976). Slavery was abolished in 1834, though the most economically productive years for the island were directly before the cessation of the slave trade. Without a constant influx of new labour, production began to decline sharply (Higman, 1976), a trend that continued long after slaves had been freed.

Post-emancipation Jamaica was not typical for the Caribbean. While Barbadian planters found the system of wage labour much more effective for producing sugar cane through lower operating costs and improved technology, Jamaican planters resisted the new system adamantly. They instead relied on the implementation of coercive tactics through unfair tax codes and inconsistent wages to keep their plantations operating. This proved disastrous as many former slaves across the island embraced a life of subsistence agriculture, working only when the wages were guaranteed (Sewell, 1968). The economically backwards Jamaican planters could no longer compete for the British market against cheaper and more reliable sugar from Cuba and Brazil. As a result, capital was quickly withdrawn from the island by absentee planters as hundreds of European planters left the island. The already suffering sugar industry fell derelict. Without access to basic healthcare, mortality in Jamaica increased drastically after emancipation (Emmer, 2000).

Jamaicans have been leaving Jamaica in search of better economic opportunities since work began on the Panama Canal in 1853 (Hall, 1971). Such large-scale population movements have been the result of high unemployment rates from falling sugar prices and the opportunity to use remittance to purchase land or support family members in Jamaica (Marshall, 1982). By the end of the Second World War, many Jamaicans had left for Canada, the United States and United Kingdom to find work in their now depleted work forces. According to the World Bank's Migration and Remittances Factbook (Worldbank, 2011), in 2010 there were 985,500 emigrants from Jamaica living predominantly in the United States, the United Kingdom and Canada. Remittances are second only to tourism in contributing to their economy with an estimated 2.02 billion US dollars received in 2010 (Worldbank, 2011).

Occurring coincidently with the rise in prominence of its sprint athletes, lifestyle-related diseases such as cardiovascular diseases, type 2 diabetes mellitus and cancer have become the leading cause of death in Jamaica, accounting for 60% of deaths in men and 75% of deaths in women (Wilks *et al.*, 2008). Rise in tobacco use, unhealthy diets, physical inactivity and alcohol abuse are largely responsible for the development of these non-communicable disease (Lee *et al.*, 2012a). According to the Jamaica Health and Lifestyle Survey (Wilks *et al.*, 2008) almost a half of the adult population was classified as having low physical activity or being inactive, a stark contrast with the success of the nation in athletics. Over 90% of

persons who were diagnosed as being obese, having a high blood pressure and having high cholesterol were not on a specific diet for their condition and about 99% of Jamaicans currently consume below the daily recommended portions of fruits and vegetables (Wilks *et al.*, 2008).

1.6 Proposed Selection Pressures Imposed by Slavery

It has been hypothesised that the physiological pressures of coerced migration and slavery in a new environment were strong enough to lead to a selective sweep in West African slave-descendant populations. To explain the high levels of hypertension observed in African-Americans when compared with European-Americans, Wilson and Grimm (1991) hypothesised that the pressures of dehydration during the Middle Passage and chattel slavery were strong enough to select for individuals with a greater ability to retain sodium. The authors argued that enslaved Africans were biologically adapted to hot, humid environments with very low access to salt and that slaves with a greater ability to conserve salt were more likely to survive dehydration induced by perspiration, diarrhoea and nausea during the Middle Passage. Finally, the high mortality rates associated with working as plantation labourers in hot and humid conditions would have selected for those able to survive complications of dehydration.

Despite being praised by the popular media (Diamond, 1991; Dubner, 2005), the "slavery hypothesis" for hypertension has received a critical response from other investigators. Coastal African populations would have evaporated sea water to create salt for trade with the interior of Africa (Curtin, 1992). As the sources of slaves in Africa were presumed to be predominantly within 100 to 200 miles of the coast along trade routes possibly dating back thousands of years (Posnansky, 1973; Nunn, 2008), there seems to be little evidence that these African groups were adapted to environments without access to salt. Modern African populations from these regions have been shown to have lower levels of hypertension than certain European groups (Cooper & Rotimi, 1994). Despite compelling accounts that dehydration contributed greatly to mortality during the Middle Passage (Kiple & Higgins, 1989), historical research also suggests dehydration did not lead to as many deaths during the Middle Passage as suggested (Steckel & Jensen, 1986). As the price of slaves rose during the slaving period (Bean, 1972), it was in the ship

owner's best interest to keep as many slaves as possible alive and in reasonable health for trade in the Americas, so purposefully withholding water would have been extremely illogical (Curtin, 1992). In conclusion, it seems unlikely that slave dehydration during the Middle Passage would have been a strong enough selective force to manifest itself as the high prevalence for hypertension seen in modern African-Americans.

After arriving, slaves died much more frequently from pneumonia and tuberculosis than gastrointestinal complications arising from dehydration (Sheridan, 1985; Kiple & King, 2003). The treatment of slaves varied greatly depending on their location. While mortality was high for slaves working on Caribbean sugar plantations, slaves in North America experienced relatively lower levels of mortality (Engerman, 1976). Despite evidence suggesting that the proposed selection pressures would have been stronger for Caribbean slave populations than North American slave populations, the prevalence of hypertension is higher in African-Americans than Afro-Caribbean populations (Cooper *et al.*, 1997). Additionally, it is unlikely that sodium retention and excretion are both controlled by the same genetic mechanism (Weder & Schork, 1994). It was also suggested that there was a high mortality pressure from complications associated with dehydration. However, if these complications were driving selection, it is more likely that individuals would be selected for resistance to those maladies as opposed to moderation of any associated symptoms by selecting for improved water retention (James & Baker, 1995). The candidate genes associated with hypertension in African-Americans do not show a reduction in genetic variability when compared to Africans, as would be expected were a selection event to have occurred (Poston *et al.*, 2001). Finally, it is unlikely that the effects of a single negative selection event could persist for hundreds of years. Once the slaves arrived in the New World, continued gene flow with Europeans would increase the genetic diversity of slave-descendant populations and gradually dilute any signal of the selection event (Jackson, 1991).

More recently, several Jamaican medical researchers have hypothesised that the slave trade is directly responsible for the overrepresentation of the West African diaspora in elite sprint athletics. Morrison and Cooper (2006) hypothesised that the disproportionate success of the West African diaspora in sprint athletics was the result of natural selection on biochemical metabolic pathways, pulmonary

physiology and muscle-fibre biology originating from environmental pressure from the Trans-Atlantic slave trade.

The first trait suggested by Morrison and Cooper (2006) subject to selection by slavery was biochemical metabolic pathways. Greater activity in the phosphagenic, glycolytic and lactate dehydrogenase metabolic pathways has been shown in African-Canadians when compared to European-Canadians (Ama *et al.*, 1990). The increase in enzyme activity would produce ATP more quickly and convert lactic acid back into glucose more efficiently.

Next, Morrison and Cooper (2006) suggest that pulmonary physiological differences were under selection by slavery. West African descendant populations have also been shown to have smaller lung volume than European-Americans (Hankinson *et al.*, 1996), potentially leading to greater rates of ventilation and ultimately oxygen consumption in every phase of exercise due to a higher provision of oxygen to the total area of the lungs in African-Americans (Damon, 1966). Unfortunately, Morrison and Cooper (2006) do not offer any testable hypotheses as to why West African descendant populations outperform native West Africans in sprint performance, relying on unspecified selection mechanisms to explain this observable difference.

The final trait suggested by Morrison and Cooper (2006) to be acted upon by natural selection was muscle-fibre biology, specifically its relationship to the sickle cell trait. Populations originating from environments with endemic malaria, such as West Africans, have developed resistance to falciparum malaria over thousands of years through selection for sickling haemoglobin alleles, thus preventing the pathogen from spreading through the host (Lopez *et al.*, 2010). While the heterozygous *HbA*/*HbS* genotype of the *HBB* gene confers a resistance for malaria – referred to here as the sickle cell trait, the homozygous *HbS*/*HbS* genotype leads to many health complications generally referred to as sickle cell disease. There are compensatory mechanisms exhibited by individuals with the sickle-cell trait to combat an elevated risk of anaemia including increased cardiac output, increased blood viscosity and cardiac enlargement (Mozos, 2015). Additionally, a study into the potential compensatory mechanisms for the selection of the sickle-cell trait has led to evidence of higher cross-sectional surface area of the IIx muscle fibres that are involved in generating explosive

power (Vincent *et al.*, 2010). Evidence of an association between the sickle cell trait and athletic performance is mixed. There is limited evidence that heterozygotes are overrepresented in elite sprint athletes from the Ivory Coast (Le Gallais *et al.*, 1991) and a small sample of elite French sprint athletes of Afro-Caribbean descent (Marlin *et al.*, 2005), although there was no evidence for overrepresentation in elite sprint athletes from Tunisia (Touhami *et al.*, 2010).

The sickle cell trait provides resistance against falciparum malaria, a disease not endemic to North America or the Caribbean, apart from Haiti and the Dominican Republic (Gething *et al.*, 2011). A recent study found that approximately 8% of African-Americans (Nelson *et al.*, 2016) and 10% of Jamaicans possess the sickle cell trait (Hanchard *et al.*, 2006). On average, approximately 0.3% of African-Americans (Therrell & Hannon, 2006) and 0.6% of Jamaicans (Serjeant, 2013) are homozygous for the sickle allele. Nevertheless, despite the removal of malaria from the environment, there does not appear to be a reduction in the prevalence of sickle cell disease brought about by relaxed selection and gene drift (Hanchard *et al.*, 2006; Hassell, 2010). With the improvement of neonatal screening and disease treatments, selection against homozygotes has been suppressed, while at the same time the sickle cell trait is no longer selectively advantageous. Individuals with the sickle cell trait generally live normal lifespans without any complications arising from their sickle cell trait (Thompson, 2013). Under extreme conditions, however, these individuals are at increased risk of exertional rhabdomyolysis, splenic infarction and renal papillary necrosis (Key & Derebail, 2010).

After a succession of highly publicised sudden deaths in collegiate athletes with the sickle cell trait, the National Collegiate Athletics Association began testing all incoming student athletes for the presence and amount of sickled haemoglobin (Hosick, 2010). This, however, was countered by a statement from the American Society of Haematology, suggesting that individuals with the sickle cell trait could be discriminated against and universal guidelines to reduce exertion-related injuries regardless of genotype would be more effective (American Society of Hematology, 2012), an approach shown to be effective in the US military (Kark *et al.*, 2010).

In an investigation to determine the frequency, epidemiology and clinical profile of deaths related to the sickle cell trait in United States Sudden Death in Athletes Registry (maintained by Minneapolis Heart Institute Foundation), 23 of 2462 athletes from youth to professional level were shown to carry the sickle cell trait (Harris *et al.*, 2012). All 23 of these athletes were African-American. Interestingly, when considering there were 699 African-American athletes in the registry, these data suggest that only 3.3% of these athletes had the sickle cell trait. This value is far below the estimate of 8% for the general population of African-Americans (Nelson *et al.*, 2016), suggesting that the sickle cell trait is actully underrepresented in athletes. The association between elite sprint performance and the sickle cell trait in the hypothesis posited by Morrison and Cooper (2006) is puzzling in light of these findings. However, caution must be paid before assuming these athletes respresent all African-American athletes. In conclusion, while the sickle allele has been associated with greater cross-sectional surface area of type IIx muscle fibre types (Vincent *et al.*, 2010), there is not enough evidence to draw any firm conclusions about the role of the sickle allele in sprint performance.

It has also been hypothesised that selection for increased levels of testosterone during slavery may have improved the reproductive success of slaves, thus passing this trait to their progeny and leading ultimately to greater sprint performance (Aiken, 2011). The author draws evidence for this selection event from the enhanced testosterone responsiveness in the West African diaspora, potentially leading not only to higher instance rates of prostate cancer in African-Americans and Afro-Caribbeans (Rebbeck *et al.*, 2013), but also to elite sprint performance. Increased levels of serum testosterone have been associated with better sprint performance and explosive power (Bosco *et al.*, 1996). Aiken goes further, hypothesising that the higher levels of testosterone linked to greater levels of sebum production and increased skin thickness (Giacomoni *et al.*, 2009) would allow slaves to better combat any macerating effect of lying in potentially infectious bodily fluids during the Middle Passage. Increased testosterone and aggression are suggested as beneficial traits for a slave on the plantations, increasing the chances of survival and reproduction. Aiken (2011) also postulates that more responsive androgen receptors would have a survival advantage during the Middle Passage. Additionally, more responsive androgen receptors have been associated with leaner muscle mass, greater oxygen capacity and the ability to

deliver oxygen more effectively (Herbst & Bhasin, 2004) which may have allowed coerced Africans to remain cool in the stifling conditions below the deck of the slave ships.

The hypothesis posited by Aiken (2011) suggests that increased levels of testosterone would have been beneficial not only during the Middle Passage, but also during life as a slave. In this hypothesis, the effects are assumed to be so pronounced that the descendants of the slave trade would have inherited increased androgen receptor responsiveness, leading to an overrepresentation of prostate cancer and elite sprint performance. Testosterone has been shown to be associated with elevated levels of aggression, social dominance and sensitivity to status threats in both men and women (Mehta & Beer, 2010). While it may be possible that heightened aggression and resulting social dominance could be advantageous within the slave communities where resources were scarce, aggressive reactions to social provocation in the strict social hierarchy of racialised slavery seems to be very maladaptive.

Both hypotheses for the slave trade directly influencing sprint performance suffer the same weaknesses as the "slavery hypothesis" of hypertension due to evidence of varying treatment of slaves across the Americas and the Caribbean. Ultimately, any of the selection pressures would have had to be passed on to the descendants of slaves. While there is evidence that the effects of past stresses may be heritable through epigenetic inheritance (Yehuda *et al.*, 2015), the mechanisms behind such inheritance in humans are not fully understood (Heard & Martienssen, 2014).

While the previously stated hypotheses rely on selection as the result of environmental pressures imposed by slavery to explain the overrepresentation of the West African diaspora in hypertension, prostate cancer and elite sprint performance, there is the possibility that directional selection could be mediated by genetic admixture introducing novel alleles into a population adapting to a new environment. Theoretically, alleles advantageous into the new environment could spread within the admixed population, assuming a strong enough selection pressure and enough generations for the trait to permeate through the population. This demographic mechanism has been used to test for selection by the environment in a sample of African-Americans, Jin *et al.* (2012) claimed to have located regions of the genome that were overrepresented for both African and

European genetic ancestries, concluding these regions may have been under selection in the new environment. However, in response to this research, Bhatia *et al.* (2014) ran a similar analysis with a sample over 10 times larger and determined that the results observed by Jin *et al.* (2012) were expected purely by chance.

1.7 Genetic Ancestry of African-Americans and Jamaicans

The many signatures of African and European DNA in the genomes of African-Americans and Jamaicans permit the application of methods from phylogeography, the principles and processes behind the distribution of genealogical lineages across geographical regions. The relationships between these lineages allows inferences to be made about the ancestry of test individuals. Phylogeographic techniques have been used to better understand the origin of modern humans (Ingman *et al.,* 2000) and the relatedness of populations around the world (Li *et al.,* 2008). Thus, it is possible to investigate the relationship between African-Americans, Jamaicans and other populations around the world by determining their genealogical relationships using these molecular data.

1.7.1 Mitochondrial DNA

As described in Section 1.2.2, mitochondria are critical for cellular energy production in every eukaryotic cell. The origin of mitochondria is best explained by the endosymbiont hypothesis (Margulis, 1981). This hypothesis purposes that mitochondria descend from alpha-proteobacterial endosymbionts that were ultimately integrated into host archaeal cells which provided the nuclear genome to form early eukaryotes. This hypothesis explains the origin of the mitochondrial genome as well as the structural and functional complexities of mitochondria. It assumes that nuclear and mitochondrial genomes (mtDNA) were derived from evolutionary distinct lineages before they became united and that eukaryotic cells are a remnant of the original endosymbiotic event. Additionally, the hypothesis assumes that there has been a substantial transfer of genetic information between the two genomes over the course of evolution because neither nucleated cells or mitochondria can exist independently (Gray, 1989).

The mtDNA is passed from mother to offspring and is assumed not to recombine, though there is some limited evidence of rare paternal inheritance (Schwartz & Vissing, 2002; Piganeau & Eyre-Walker, 2004) despite none being detectable at the population level (White *et al.*, 2013). Even assuming paternal mtDNA leakage, at most it must be less than 1 in 1,000 as there are approximately 100,000 mitochondria in the human egg and only approximately 100 in the sperm (Satoh & Kuroiwa, 1991), making any noticeable signal at the population level practically indiscernible with current molecular techniques.

Mutations along the entirety of the mitochondrial genome occur more frequently than the nuclear genome (Ballard & Whitlock, 2004) and are regular enough that fine-scale trees of global genetic diversity can be created (van Oven & Kayser, 2009). These mutational differences can then be used to create a genealogical tree of relatedness, with branches forming clusters of haploid genotypes (haplotypes), referred to as haplogroups (Torroni *et al.*, 1993). An example of a low-resolution tree is shown in Figure 1.2. The distribution of mitochondrial lineages across geographical regions and their genealogical relationship to each other allows individuals of uncertain origin to match their maternal lineage back to its most likely origin. However, caution is needed in making over-specific claims about the geographical origin of a haplotype due to the movement of people within modern times (Rando *et al.*, 1998; Salas *et al.*, 2004; Cerezo *et al.*, 2012). A summary of population-level mtDNA ancestry inferences presented in this chapter are summarised in Table 1.1, presented at the end of Section 1.7.1.2.

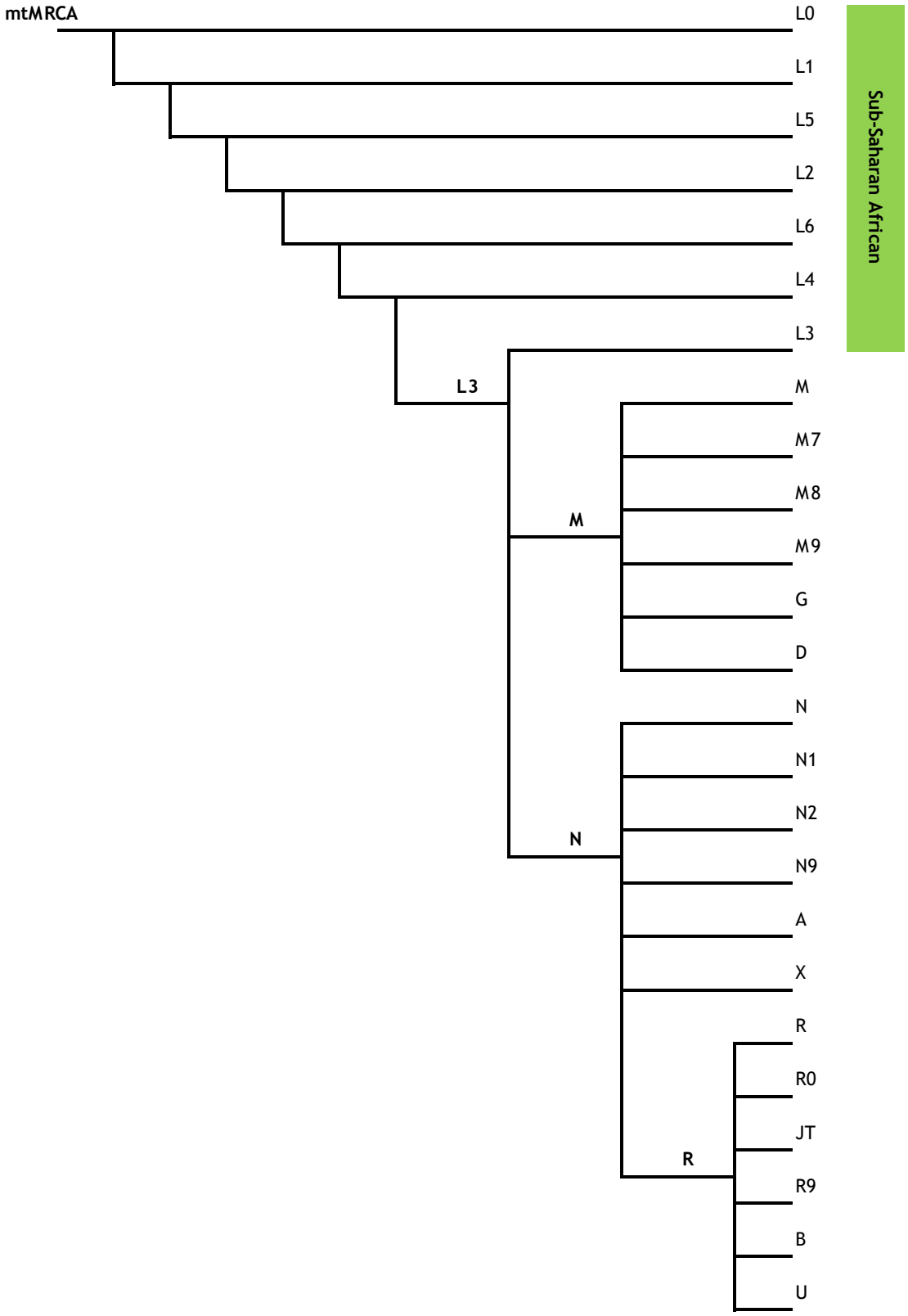


Figure 1.2 Phylogenetic tree of major mitochondrial haplogroups adapted from Phylotree.org (van Oven & Kayser, 2009). The root of the tree (mtMRCA) refers to the most recent maternal common ancester for all living people. Sub-Saharan African haplogroups begin with "L" and are indicated in green, while all mitochondrial diversity outside of Africa descends from either haplogroups M or N.

1.7.1.1 mtDNA research in African-Americans

After the seminal work by Cann *et al*. (1987b) demonstrating the matrilineal connectedness of all living people, numerous criticisms were levied at the authors for using a sample of African-Americans as proxy for native Sub-Saharan African mtDNA in their worldwide phylogenetic study (Darlu & Tassy, 1987; Spuhler, 1988; Excoffier & Langaney, 1989). Though the work did provide an explanation for this, the published response to the criticism failed to address the rather large assumption that African-American mtDNAs can be used to represent Sub-Saharan African mtDNAs (Cann *et al*., 1987a). Not until native African individuals had their mtDNA genotyped by Vigilant *et al*. (1991) was the assignment of African-Americans as Sub-Saharan assessed at the level of mtDNA. The study revealed that the mtDNA of the African-Americans used in Cann *et al*. (1987b) linked them closely to native Africans, though one individual's mtDNA was more closely related to that of Europeans.

One of the earliest and more rigorous attempts at qualifying the ancestral make-up of the West African diaspora was by Parra *et al*. (1998). As part of a larger study assessing uniparental and genome-wide population-specific markers, the researchers estimated the genetic contributions of European and African ancestry in 10 populations spread across the United States (Maywood, Illinois; Detroit, Michigan; New York, New York; Philadelphia, Pennsylvania; Pittsburgh, Pennsylvania; Baltimore, Maryland; Charleston, South Carolina; New Orleans, Louisiana; and Houston, Texas) as well as Jamaica. Using six restriction fragment length polymorphisms (RFLPs), each corresponding to a basal SNP defining a particular haplogroup, the authors were able to infer the continental origin of the maternal lineages. The Detroit sample was completely devoid of European maternal contribution while the Baltimore sample was estimated at 15.0% European. Although molecular methodology improved greatly since Parra *et al*. published this work, the predominately Sub-Saharan origin of the African-American matrilineal gene pool continued to be observed in the more sophisticated studies that followed.

An early large-scale study sampling 1,148 anonymous African-Americans as part of the United States' Federal Bureau of Investigation's Scientific Working Group on DNA Analysis Methods (SWGDAM) reference database for mtDNA was able to highlight the high frequency of Sub-Saharan African mtDNA haplogroups present

in a comprehensive sample of African-Americans (Monson *et al.*, 2002; Allard *et al.*, 2005). Sequencing the first hypervariable segment of the mtDNA's non-coding control region (HVS-1), the authors observed 13 of 18 haplogroups previously identified as originating in Sub-Saharan Africa (Salas *et al.*, 2002). Of these, L2a was the most commonly observed (18.8%), followed by L1c (11.0%), L1b (9.1%), L3e2 (9.0%) and L3b (8.1%). Roughly 8% of the haplogroups observed within African-Americans were of European, Native American or East Asian origin.

Using a set of samples collected from sites in both Tennessee and Florida, Diegoli *et al.* (2009) used HVS-1 to characterise the mtDNA profile of 248 self-identified African-Americans. As was expected, the majority (94%) of observed haplogroups are of Sub-Saharan origin, specifically a high frequency of haplogroups L1b, L1c, L2b, L2c, L2d, L3b, L3d, and L3e with a low frequency of other L3 or non-L3 groups. Haplogroups L1c (11.3%), L2a (23.0%), and L3e (16.5%) were found to be the most common, while other L3 and non-L groups combined represent only 12.5%. East Asian or Native American-specific haplogroups (A, B, C, D, and M) have been shown previously to be a minority component of the African-American mtDNA pool in the African-American population and the Diegoli *et al.* (2009) data set only contained five instances (2%). European representation is marginally greater, accounting for approximately 4.5% of the observed haplogroups. These results were consistent with other published reports of African-American mtDNA variation.

Within an effort to characterise the phylogenetic landscape of four broad ethnic groups in the United States (Native-, European-, Asian-, and African-American), Lao *et al.* (2010) used a sample of 246 self-declared African-American males collected from six locations across the nation to better understand not only the genetic demography of the United States, but also assess any spatial differences. The authors found mtDNA haplotypes in African-Americans to be 92.7% Sub-Saharan African, in line with previous reports, with the non-African lineages being of European (H1a, H1c, H5, HV0, J1b, J1c, T2, U5a, U5b) or of Native American (A2, B4a) or East Asian origin (F3b, G, D/E/G, M10). Importantly, they found no evidence of geographic substructure in the mtDNA ancestry. This meant that there was no statistically significant difference between the mtDNA haplogroup profile distributions of the six locations.

In another study investigating geographical population substructure of mtDNA in African-Americans, Johnson *et al.* (2013) genotyped and later haplogrouped the mtDNA of 343 individuals participating in the multi-cohort Reaching for Excellence in Adolescent Care and Health (REACH) study (Rogers *et al.*, 1998). Enrolment into REACH took place at 15 clinics in 13 cities throughout the United States including four recruitment sites in the Northeast, three in the Mid-Atlantic, six in the Southeast, one in the Mid-West, and one in the South-West. The authors found comparable results to those in previously published studies. The REACH cohort appears, however, to have a relatively high proportion of non-L haplogroups (12.0% compared to 8.2% in the SWGDAM data set). This may be an artefact of the smaller sample size, or indeed a previously underreported level of matrilineal diversity. Location specific results were not reported, so any spatial variation remains unevaluated.

1.7.1.2 mtDNA research in Jamaicans

The published literature on the genetic ancestry of African-Americans is much larger than that of populations of the West African diaspora in the Caribbean. Parra *et al.* (1998) estimated the contribution of European and African ancestry by genotyping mtDNA RFLPs corresponding to particular haplogroups in 102 Jamaicans. Using the weighted least squares method described in Long (1991), the authors found evidence of a 12.9% matrilineal contribution from Europeans, suggesting moderate gene flow from Europeans into Africans.

In another study of mtDNA in Jamaicans, McLean *et al.* (2003) genotyped three HVS restriction site polymorphisms found to be attributable to different continental regions. One hundred and eighteen Jamaicans were sampled as part of a larger sampling with individuals from Sierra Leone and African-Americans from both rural and urban South Carolina. The authors found no indication of any genetic introgression from European maternal lineages in their sample. In a follow-up to their previous work, McLean *et al.* (2005) genotyped seven mtDNA restriction sites within HVS-1 in 104 of the 118 Jamaicans genotyped in their previous study. With additional genetic data, the authors found only a negligible European contribution, and using an admixture calculation presented by Long (1991) found the European matrilineal contribution to be 2%. These results were noticeably

smaller than the Parra *et al.* (1998) estimates, clearly suggesting either sampling bias in one of the cohorts. Very little has been done to assess the mitochondrial gene pool of Jamaicans despite the opportunity to examine how the practices associated with slavery, namely the Trans-Atlantic slave trade, influenced on the genetic constitution in a historically well-characterised population. This was a large motivation behind the work presented in Chapter 3.

Table 1.1 Summary of European ancestry inferred from mtDNA in African-Americans and Jamaicans for studies introduced in this thesis. The percentage European denotes the estimated proportion of individuals with a European maternal origin. The asterisk denotes the weighted mean of the European admixture estimates for the African-American sample. The dagger indicates the weighted mean of the two McLean *et al.* studies (0.9%) was used before calculating a conservative weighted mean for 104 individuals from McLean *et al.* and 102 from Parra *et al.*

Population	**Study**	***N***	**% European**
African-American	Parra *et al.*, 1998	1020	7.6
	Monson *et al.*, 2002	1148	4.3
	Lind *et al.*, 2007	93	8.5
	Diegoli *et al.*, 2009	248	4.5
	Lao *et al.*, 2010	246	7.3
	Johnson *et al.*, 2013	343	10.8
	Total	3098	6.5*
Jamaican	Parra *et al.*, 1998	102	12.9
	McLean *et al.*, 2003	118	0.0
	McLean *et al.*, 2005	104	2.0
	Total	206	6.9†

1.7.2 Y Chromosome

One of the smallest human chromosomes, the Y chromosome is a sex chromosome containing 61 protein-coding genes, 25 protein-coding genes and transmitted largely without recombination from fathers to sons (Periyasamy *et al.*, 2009). Of the roughly 59 million base pairs that make up the Y chromosome, about 5% recombine with the X chromosome (Rappold, 1993). The remaining portion of the non-recombining Y chromosome (NRY) contains 78 protein-coding genes (Ginalski *et al.*, 2004) and is used in genealogical research and is important for creating genealogical trees (Tilford *et al.*, 2001). Like mtDNA, the distribution of NRY haplotypes show strong spatial patterns due to its low effective population and higher genetic drift (Hammer & Zegura, 1996). A summary of population-level NRY ancestry inferences presented in this chapter are summarised in Table 1.2, presented at the end of Section 1.7.2.2.

1.7.2.1 Y chromosomal research in African-Americans

Parra *et al.* (1998) used ancestrally specific NRY markers to estimate the genetic contributions of European and African ancestries in 10 populations spread across the United States. The paternal component in all cases was much larger than that of mtDNA, ranging from 8.6% in Houston to 46.9% in New Orleans.

Analysing the same 246 African-American men from across the United States mentioned above in Section 1.7.1.1, Lao *et al.* (2010) genotyped 42 haplotype-specific NRY markers to assess genetic ancestry in African-Americans. To evaluate fine-scale stratification among the common Sub-Saharan haplogroup E, and potentially detect known haplotypes of haplogroup E now present throughout Europe and Eurasia (Underhill *et al.*, 2001; Cruciani *et al.*, 2004), 18 additional NRY markers were genotyped. The authors determined that 69.5% of the sample's paternal gene pool was of Sub-Saharan origin, with the next largest being European at 29.7%. The remaining 0.8% came from a single individual belonging to Southeast Asian basal paraphyletic group (paragroup) K*. As with mtDNA, no statistically significant geographic sub-structuring was detected in the sample population.

In a more recent effort to characterise the Y chromosome lineages in the African diaspora, Torres *et al.* (2012a) genotyped eight Y single tandem repeat (STR) loci, a Y-specific Alu insert (DYS287), and a Y-SNP (M89) in 426 men from South Carolina and 106 men from Washington, DC (Kittles *et al.*, 2001; Bonilla *et al.*, 2006; Kidd *et al.*, 2006). Self-declared African-Americans from South Carolina were found to have a 32.3% European contribution, while the contribution was slightly higher for Washington DC at 36.6%. The discrepancy between the observed European mtDNA contribution and observed European NRY contribution is consistent with the directional gene flow between European men and African women.

1.7.2.2 Y chromosomal research in Jamaicans

Parra *et al.* (1998) estimated the genetic contributions of European and African ancestry in 102 Jamaicans. The European paternal component of Jamaica was inferred to be 17.9%. This study was the first to highlight the significance of the European introgression despite Europeans only constituting roughly 10% of the island's entire population during the slave era (Engerman, 1976). Perhaps as an

artefact of sampling or molecular methodology, the results for the continental contribution to paternal lineages in Jamaica vary much more between publications than in the mtDNA literature. Torres *et al.* (2007) genotyped a sample of 53 Jamaicans at 12 NRY loci. These loci consisted of 11 short tandem repeats (DYS 19, DYS 385a/b, DYS 389I, DYS 398II, DYS 390, DYS 391, DYS 392, DYS 393, DYS 437, DYS 438, DYS 439) and one Alu insertion (DYS 287). The authors determined that roughly 58.3% of male lineages in their sample were of Sub-Saharan descent. As part of a larger study into the Y chromosomal patterns present in a wider variety of Caribbean populations, Torres *et al.* (2012a) estimated the paternal lineages of 75 Jamaicans. Genotyping eight Y-STR loci and additional markers (DYS287 (YAP), M89 SNP) they estimated the European paternal contribution as 41.1%.

Simms *et al.* (2012) sampled 159 Jamaicans at 177 high-resolution Y-chromosome binary markers and 17 Y-STR loci as part of larger study into the paternal lineages found in the Haiti and Jamaica. The authors used the coalescent-based approach described in Dupanloup and Bertorelle (2001) and the PASW analysis software version 18 (SPSS, Chicago, IL, 2009) to generate admixture coefficients. Both methods found the African component to be the largest contributor to Jamaican paternal lineages (78.6% and 68.7%, respectively), followed by European (19.3% and 12.0%, respectively) and East Asian (2.1% and 5.8%, respectively). The weighted least squares method found negligible contributions of Native American and South Asian ancestry as well, while the coalescent-based approach found non-negligible amounts (at 0.043 and 0.091, respectively). Interestingly, signatures of Asian haplogroups were also found at low levels (East Asian haplogroup O-M175 (3.8%) and South Asian haplogroups H-M69 (0.6%) and L-M20 (0.6%) perhaps indicative of post-abolition labour movement within the British Empire (Emmer, 1995). More recently, Simms *et al.* (2013) reported the allele and haplotype frequencies of the commonly cited Y-filer genotyping kit (Applied Biosystems, 2006) to assess the paternal genetic structure of the Bahamas, Haiti, and Jamaica. After genotyping 17 STR loci and calculating admixture fractions for 140 of the 159 Jamaicans describe in Simms *et al.* (2012) using PASW, the authors concluded that the largest component was African (71.1%), followed by European (23.5%). As seen in the other works above, there was also detectable ancestry from East Asians (3.5%) and Native Americans (2.0%). Even when considering discrepancies in

molecular methodology and statistical analyses, there seems to be a substantial variability in the European contribution to the Jamaican paternal gene pool.

Uniparental markers are invaluable in understanding the origins of populations. However, the application of these findings to estimate the genome-wide ancestry of a population is less informative. Genotype data from the autosomes permits ancestry detection at a much finer scale by utilising observable allele frequency differences from populations around the world. This avenue of research is complementary to uniparental lineage analyses and can provide a much fuller account of the genetic ancestry of population without relying on a single realisation of the evolutionary process.

Table 1.2 Summary of European ancestry inferred from the Y chromosome in African-Americans and Jamaicans for studies introduced in this thesis. The percentage European denotes the estimated proportion of the paternal gene pool with a European origin. The asterisk denotes the weighted mean of the European admixture estimates for the African-American sample. The dagger indicates the weighted mean of the two Simms *et al.* studies (19.4%) was used before calculating a conservative weighted mean for 140 individuals from Simms *et al.* with the remaining Jamaican studies.

Population	**Study**	***N***	**% European**
African-American	Parra *et al.*, 1998	1020	24.6
	Lao *et al.*, 2010	246	30.5
	Lind *et al.*, 2007	93	28.5
	Torres *et al.*, 2012	532	34.5
	Total	1891	28.3*
Jamaica	Parra *et al.*, 1998	102	17.9
	Torres *et al.*, 2012	75	41.1
	Simms *et al.*, 2012	159	15.7
	Simms *et al.*, 2013	140	23.5
	Total	317	24.0†

1.7.3 The autosomal genome

While uniparental markers can trace the likely origins of both maternal and paternal lineages, a recent study by Emery *et al.* (2015) found evidence that there was little correspondence between mitochondrial lineages and genome-wide ancestry in worldwide sample of population. The researchers determined that only limited information about either continental ancestry or continental region of origin can be gained from mtDNA haplogroup data. Rough bi-parental autosomal population-level admixture coefficients, however, can be estimated solely using observed mtDNA and NRY genotype data (Bedoya *et al.*, 2006; Lind *et al.*, 2007). Since the Y chromosome admixture fraction m_y is equal to the male contribution

and the mtDNA admixture fraction m_{mt} is equal to the female contribution, the bi-parental admixture fraction m_{auto} for a population can be formulated as $m_{\text{auto}} = \frac{1}{2} m_{\text{y}} + \frac{1}{2} m_{\text{mt}}$, however the utility of this estimate is limited compared to more accurate methodology. Geographically distributed alleles along the autosome with sufficient allele frequency differences are known as ancestrally informative markers (AIMs). Comparing the markers in admixed populations to presumptive source populations enables the ancestry of an admixed population to be more accurately estimated (Weiss & Long, 2009; Novembre & Ramachandran, 2011). A summary of population-level autosomal ancestry inferences presented in this chapter are summarised in Table 1.3, presented at the end of Section 1.7.3.2.

1.7.3.1 Genome-wide ancestry in African-Americans

To assess the genome-wide ancestry of African-Americans, Parra *et al.* (1998) used nine ancestrally informative markers on the autosomes to estimate the genetic contributions of European and African ancestry in 10 African-American populations across the United States. Bi-parental African and European admixture was calculated using both the weighted least squares regression and gene identity methods, utilising only nine autosomal DNA markers. They estimated average European ancestry of African-Americans ranging from 11.6% in Charleston to 22.5% in New Orleans. Although nine markers are far below the 2,000 – 3,000 markers suggested to tag the entire African-Americans genome (Seldin *et al.*, 2011), this early study was the first to attempt to characterise genome-wide ancestry in geographically distinct groups of African-Americans.

In an attempt to compare the levels of admixture in African-Americans over the four regions of the genome (autosome, Y chromosome, X chromosome and mtDNA), Lind *et al.* (2007) collected 93 samples from men in Chicago, Baltimore, Pittsburgh, and North Carolina and genotyped them at 2,018 previously identified ancestrally informative SNPs (Smith *et al.*, 2004). The broad geographic sampling was chosen to mitigate the effect of any sampling bias on the results. The authors also examined 121 X chromosome SNP genotypes, ten Y chromosome SNPs and six mitochondrial SNPs. The latter two markers allowed the authors to identify 11 Y haplogroups and six mtDNA haplogroups, respectively. Bi-parental admixture coefficients were calculated at the population level using the weighted least

squares method. At the individual level, a maximum likelihood estimation was used for mtDNA and Y chromosome and the genetic clustering algorithm STRUCTURE (Pritchard *et al.*, 2000; Falush *et al.*, 2003) was used for the autosomes and X chromosome. The authors found no significant difference between sampling locations. The average European genetic contribution was largest in the Y chromosome (28.5%) followed by the autosome (20.0%), X chromosome (12.1%) and mtDNA (8.5%), consistent with the evidence for reasonably strong sex-biased gene flow during the founding of the African-American population.

More recently, Guo *et al.* (2013), examined how well the self-declared ancestry of two large sets of young Americans from African, European, South Asian, East Asian and Native American backgrounds correlated with their genetic ancestry using question-dependent responses to both university accommodation applications (*N* = 2,065) and a longitudinal health survey (*N* = 2,281). The authors used a set of 186 ancestrally informative markers to assess racial fluidity and sociological influences on biological ancestry. The authors found strong concurrence between reported ethnicity and biological ancestry, with self-declared African-Americans having on average 8.7% (*N* = 279) and 5.9% (*N* = 378) European ancestry in the respective studies. Their analysis also revealed that there were far fewer individuals with African ancestry of 10-50% than individuals with African ancestry of 50-90%. They conclude that this imbalanced distribution is likely a result of the 'one-drop' rule and/or the minimal miscegenation between African-Americans and European-Americans since 1865 (Williamson, 1995). Mixed-race individuals with one African-American parent and one European-American parent were treated as African-American rather than mixed-race individuals. Under such racial exclusion, these mixed-race individuals may have partnered predominantly with other mixed-race or African-American individuals rather than European-Americans. These patterns of gene flow redistributed the European ancestry in the original mixed-race individuals, allowing those alleles to permeate through the general African-American population and yielding few individuals of more than 50% European ancestry.

The research techniques utilising AIMs were important for estimating ancestry, however as whole-genome SNP arrays became cheaper, a greater amount of data was now available. With these new forms of data came new analytical techniques

to explain observable genetic variation (PCA) and clustering (e.g. STRUCTURE). Along with high-quality reference data from HapMap and 1000 Genomes Project, these new methods ushered in a new era of ancestry research.

In the first use of genome-wide SNP data, Zakharia *et al.* (2009) characterised autosomal admixture using genotype data from over 450,000 SNPs in 136 self-declared African-Americans genotyped elsewhere (Assimes *et al.*, 2008). Using the Bayesian ancestry estimation program FRAPPE (Tang *et al.*, 2005), levels of European ancestry in the sample were 22.0% with a standard deviation of 12.2%, in line with previous works. After removing eight outliers with substantial amounts of European ancestry, the authors assessed the African component of the genome, determining that among reference populations included in this study, African-Americans are more like the Yoruba of West-central Africa (63.7%), followed by the Mandenka of West Africa (19.2%) and Bantu of East Africa (13.8%). There were negligible contributions from the Biaka pygmies of Central Africa (1.0%), the Mbuti pygmies of Central Africa (0.2%) and the San of southern Africa (2.0%). The larger than expected similarity to the Bantu sample is assumed to be an artefact of the Bantu people's dispersal from Northern Cameroon and Southern Nigeria throughout eastern and southern Africa roughly 4,000 years ago (Beleza *et al.*, 2005; Coelho *et al.*, 2009).

Later, Bryc *et al.* (2010) sought to quantify patterns of population structure in West Africans and African-Americans by obtaining a genome-wide perspective of ancestry. After genotyping 500,000 SNPs, the authors used a principal component analysis (PCA) to demonstrate how their cohort of 365 African-Americans fit within observable global genetic variation by comparing them to 225 individuals from 11 African populations (Tishkoff *et al.,* 2009) and a subset of 400 individuals from the 6,000 person POPRES database containing individuals of European, North American and Asian descent. The authors also used the algorithm FRAPPE to examine individual and population clustering to investigate population structure. West Africa's genetic structure was shown to reflect language groupings and geography. The median estimated European ancestry in the African-American cohort was 18.5%, although there was considerable variation between individuals. The patterns of genetic similarity suggest that African-Americans are most like non-Bantu Niger-Kordofanian-speaking populations, an observation consistent with the Trans-Atlantic slave trade. Some regions of the African-American genome showed

excess African ancestry, though the authors are cautious in stating whether these regions are under directional selection, as conclusions may be biased due to poor statistical power.

Using a much larger sample of genetic data from the direct-to-consumer genomics company 23andMe, Bryc *et al.* (2015) sought to characterise the geographic variability of genetic ancestry in 5,269 self-declared African-American, 8,663 Latino-American and 148,789 European-American customers. These individuals were genotyped using one of four different whole-genome SNP arrays, though each genotyped at least 550,000 SNPs. Using the proprietary fine-scale ancestry deduction pipeline Ancestry Composition (Durant *et al.*, 2014), the authors observed notably lower mean African ancestry (73.2%) and elevated levels of European ancestry (24.0%) in African-Americans than has been observed in other studies (Table 1.3). They also observed 108 African-Americans with less than 2% African ancestry, though the authors attribute these observations to survey error. Excluding these individuals raises the mean level of African ancestry to 74.8%. The authors also found evidence of regional differences in ancestry, reflecting either differences in admixture, self-identity or degree of assortative mating.

Lastly, Baharian *et al.* (2016) sought to provide a comprehensive assessment of genetic diversity of 3,726 African-Americans using a geographically representative sample. Using a combination of three different cohorts with 553,527 variants in common across the data sets, they estimated a national means of 82.1% African ancestry, 16.7% European ancestry, 1.2% Native American ancestry. The proportion of African ancestry was elevated in the South when compared with the rest of the United States because of its greater concentration of slaves during the slave era. By estimating the ancestry at each locus in the genome, the authors found evidence that admixture in the South occurred before the beginning of the Civil War. The authors also found evidence of long-range relatedness, providing evidence of African-Americans dispersal from the South to the rest of the country in search for employment and social betterment during the early twentieth century, an event known as the Great Migration (Section 1.5.4).

1.7.3.2 Genome-wide ancestry in Jamaicans

As part of the larger study mentioned above, Parra *et al.* (1998) used nine AIMs to estimate the genetic contributions of European and African ancestry in 102 Jamaicans. African and European admixture was calculated with both the weighted least squared regression and gene identity methods and estimated that the sampled Jamaicans had 6.8% European ancestry. More recently, Torres *et al.* (2008) genotyped 28 AIMs chosen specifically to differentiate between West African groups (Central African Republic, Nigeria and Sierra Leone), European groups (Spain, Germany, England and Ireland) and Native American groups (Maya, Pima, Cheyenne and Pueblo). The unsupervised genetic clustering program STRUCTURE was used initially to assess the number of clusters that the data split into, finding the best fit with two clusters. Nevertheless, using the weighted least squares method, the mean and standard error for the tri-parental admixture coefficients for African, European and Native American were 84.4% ± 3.1, 12.4% ± 3.5, and 3.2% ± 3.1, respectively. Using the individual-based maximum likelihood approach (Hanis *et al.*, 1986; Chakraborty *et al.*, 1986), mean admixture coefficients and their ranges were 80.8% (46.8–97.0), 13.3% (0.7–47.1), and 5.9% (0.9–33.9), respectively.

Simms *et al.* (2010) genotyped 119 Jamaicans and 111 Haitians at 15 hypervariable autosomal STR loci to infer divergent demographic histories. The authors explored three separate admixture algorithms to calculate tri-parental contribution of lineages found in Africa, Europe, and East Asian using a weighted least squares method, a coalescent-based approach (Bertorelle & Excoffier, 1998) and finally a least square regression method (Roberts & Hiorns, 1965). All three methods found the African component to be the largest contribution to Jamaican ancestry, ranging from 76.5% to 82.1%, followed by European (14.2% to 17.9%) and a negligible Native American or East Asian component (3.7% to 5.7%). Using the same three methods, the Haitian samples were estimated as predominately African (77.8% to 100%), followed by European (4.3% to 29.8%) and negligible amounts of East Asian ancestry (0.0% to 0.3%). The authors found evidence of divergent population histories between the two island populations, however they highlight their inability to detect more accurate and finer-scale population substructure because they genotyped too few markers by an order of magnitude to differentiate between admixed populations (Ardlie *et al.*, 2002).

Published alongside their Y chromosomal analysis mentioned in Section 1.7.2, Simms *et al.* (2012) used the genotyping data from 15 autosomal STR loci in their previous publication (Simms *et al.*, 2010) and recalculated autosomal admixture coefficients using five reference populations with the software PASW 18 (SPSS, Chicago, IL, 2009). The authors concluded that the African component was the largest (81.4%), with European the next largest (12.3%). There were also detectable, albeit negligible, introgression from Native American (1.6%), East Asian (4.8%) and South Asian (0.0%)[1] gene pools.

As part of a larger cross-disciplinary effort to assess the effect of differences in European rule of the Caribbean, Torres *et al.* (2013) genotyped 44 Jamaicans at 105 AIMs. Using STRUCTURE, average individual admixture estimates were calculated assuming three source populations. In accordance with previously published results, the Jamaican sample was determined to be 81.4% African, although the authors found 8.3% Native American and 10.3% European components. This peculiar result is thought to be the result of several samples containing very high proportions of Native American ancestry, signified by the standard deviation of 13.5. This is perhaps a result of more recent migration from regions with demonstrably elevated levels of Native American ancestry such as Central America (Wang *et al.,* 2008) rather than admixture with indigenous Jamaicans.

While not genotyping Jamaica specifically, Moreno-Estrada *et al.* (2013) investigated the genomes of various Caribbean populations to characterise the effect of colonisation on the demographic histories of these populations. The authors genotype 330 individuals from Greater Antilles (Cuba, Puerto Rico, Hispaniola), mainland (Honduras, Columbia), and Native South America (Yukpa, Bari, Waro) at over 389,225 SNPs. Using these data, the authors found evidence of two pulses of African immigration: the first pulse represented by short ancestry tracts more like populations from the western African coast exploited earlier in the slave trade, and the second pulse represented by long ancestry tracts more like populations from the west central African coast exploited in the latter stages of the slave trade. The authors also found evidence for extensive pre-colonisation era gene flow across the Caribbean basin. Interestingly, they also determined that

[1] The standard deviation of the South Asian component was 4.0%.

the Latino-specific European component is distinct from modern Iberians, possibly reflecting the small European founder population in the Caribbean.

There have been very few genome-wide ancestry estimates reporting comparisons between the ancestry estimates of African-Americans and Jamaicans. Chiang *et al.* (2010) investigated the genome-wide ancestry in two cohorts of 521 and 321 African-Americans and two cohorts of 688 and 480 Jamaicans. After quality control, the African-American cohorts had pooled genotype data for 378,337 and 476,847 SNPs, respectively, while the Jamaican cohorts had pooled genotype data for 353,320 and 303,269 SNPs, respectively. Allele frequency data from each cohort were pooled together and compared to the YRI and CEU populations from HapMap to estimate the cohort-specific genome-wide ancestry coefficients. These estimates were generated using a linear regression by modelling the estimated pooled allele frequencies of each cohort as a linear combination of the known allele frequencies in the reference HapMap populations. The results suggested that the African-American samples had 71.3%-82.4% African ancestry and 17.5%-24.1% European ancestry, while the Jamaican sample had approximately 82.2%-86.8% African ancestry and 10.1%-12.2% European ancestry. The authors conclude that using pooled samples is a cost-effective method to estimate genome-wide ancestry in samples, although as the cost of whole-genome SNP arrays continued to fall, one would expect pooling DNA isolates to generate these data to become a less desirable method of evaluating population structure.

The recent work by Mathias *et al.* (2016) summarises the genome-wide ancestry of African-Americans and Jamaicans well. Characterising data as part of the multinational Consortium on Asthma among African-ancestry Population in the Americas (CAAPA, 2012), the authors genotyped 642 individuals from the African-Diaspora (including 328 from eight locations in the United States as well as 45 individuals from Jamaica), observing 43.2 million single nucleotide variants, more than the 38 million genotyped as part of the 1000 genomes project. Using the program ADMIXTURE (Alexander *et al.*, 2009), ancestry estimates for African-Americans across the different regions of the US ranged from 76%-84% African, 15%-23% European and 1%-2% Native American. The ancestry estimates for Jamaicans were 89% African, 11% European and 1% Native American. African-Americans cluster with Jamaicans and Barbadian and Nigerians, while Brazilians,

Puerto Ricans, Columbians and Dominicans form a second cluster, largely because of elevated Native American ancestry in the latter cluster.

Autosomal admixture coefficients have been shown to vary greatly between individuals within a population. These ancestry coefficients presented above provide evidence that the genetic structure of both African-Americans and Jamaicans is a unique combination of their population histories and the genes of their source populations. Both populations show signals for directional gene flow from European men to African women. This pattern manifests itself in mtDNAs of predominantly African origin and large proportions of Y chromosomes of European origin. Considering these findings, African-Americans are estimated to have approximately 80% African ancestry while Jamaicans are estimated to have approximately 90% African ancestry, suggesting different rates of these admixture events. The causative genetic factors behind the overrepresentation of both populations in sprint athletics will have their roots in the ancestry of these athletes. A review of the work in genetics and sprint running is necessary before any associations between ancestry and elite sprint athlete status can be made.

Table 1.3 Summary of European ancestry inferred from the autosomal genome in African-Americans and Jamaicans for studies introduced in this thesis. The asterisk denotes the weighted mean of the European admixture estimates for the African-American sample.

Population	**Study**	***N***	**% European**
African-American	Parra *et al.*, 1998	1020	16.2
	Lind *et al.*, 2007	93	20.0
	Zakharia *et al.*, 2009	136	22.0
	Bryc *et al.*, 2010	365	18.5
	Chiang *et al.*, 2010	521	17.5
	Chiang *et al.*, 2010	321	24.1
	Lao *et al.*, 2010	246	13.8
	Guo *et al.*, 2013	279	8.7
	Guo *et al.*, 2013	378	5.9
	Bryc *et al.*, 2015	5269	24.0
	Baharian *et al.*, 2016	3726	16.7
	Mathias *et al.*, 2016	328	18.6
	Total	12682	19.5*
Jamaica	Parra *et al.*, 1998	102	6.8
	Chiang *et al.*, 2010	688	12.2
	Chiang *et al.*, 2010	480	10.1
	Simms *et al.*, 2012	159	12.3
	Torres *et al.*, 2013	44	10.3
	Mathias *et al.*, 2016	45	11.0
	Total	2856	11.1*

1.8 Genetics of sport: ancestry based approaches

Genetic endowment is known to play a favourable role in elite athletic success (Section 1.3 above) and past work has observed athlete status to be approximately 66% heritable (de Moor *et al.*, 2007). Uniparental markers can be used to assess population stratification by sports performance. There is evidence to suggest that people choose mates on both polygenic traits such as eye colour and height as well as behavioural traits such as generosity and environmental factors such as education level (Creanza & Feldman, 2016). This process of assortment where people choose mates similar to themselves is referred to as homophily. Homophily may also directly affect lineage fitness. Evidence suggests more similar mates tend to have higher fertility (Thiessen & Gregg, 1980). Lineages significantly associated with elite athlete status may act as markers for an extended history of athletically gifted individuals producing athletically gifted children for repeated generations. Testing for associations between lineages and elite athlete status complements the more traditional functional genetic research model by allowing researchers to home in on candidate lineages. This process may uncover any atypical evolutionary events that may account for these observable discrepancies in athletic ability.

1.8.1 Associations between mtDNA genotypes and athlete status

The mitochondrial genome encodes for 13 structural proteins in the respiratory chain and 22 transfer RNA genes (DiMauro & Schon, 2003). The process of oxidative phosphorylation, described in Section 1.2.2, is encoded within mtDNA. Variations in mtDNA haplotype have been associated with climatic adaptation, high altitude tolerance and obesity, as well as changes in susceptibility or severity of maladies including diabetes, sepsis, stroke, Alzheimer's and Parkinson's disease and hypertension (Wallace, 2010). Lineage-based analyses have also observed that phenotypes such as aerobic capacity and trainability (Murakami *et al.*, 2002) as well as maximal oxygen capacity (Murcuello *et al.*, 2009*)* were influenced by mitochondrial haplotype.

There has been evidence of selective forces acting on mitochondria. Mishmar *et al.* (2003) noted that the variable pattern of mtDNA sequences in the *ATP6* gene differ by temperature zones, providing evidence for the involvement of selection in shaping regional mtDNA variants. Wallace *et al.* (1999) first suggested that

different mtDNA haplotypes might be influencing the physiology of individuals by predisposing or protecting them from certain diseases involved in modulating oxidative phosphorylation. Suggestive associations in small studies with specific mtDNA haplotypes have been observed for several diseases such as cardiomyopathy (OR = 10.1; 95% CI [1.0–110.0]; Shin *et al.*, 2000), transient ischaemic attack and ischaemic stroke (OR = 0.54; 95% CI [0.39-0.75]; Chinnery *et al.*, 2010), and coronary artery disease (OR = 2.36; 95% CI [1.52–3.65]; Kofler *et al.*, 2009).

Extending the concept of clinical assessments of mtDNA to examine elite endurance athletes, Rivera *et al.* (1998) sampled 125 elite male endurance athletes of European heritage from various countries and a range of sports with a $\dot{V}O_2max$ of 78.9 ± 3.8 mL/(kg·min). Sixty-five controls were selected from the multigenerational HERITAGE study of young sedentary North Americans with a $\dot{V}O_2max$ of 39.8 ± 8.2 mL/(kg·min) for comparison. After examining two coding RFLPs and one non-coding RFLP, the authors determined that there was no difference in the frequency of these polymorphisms, and thus no association between elite endurance athlete status and these mtDNA lineages at this level of power.

Scott *et al.* (2005) were the first to test the association between elite athlete status and mitochondrial haplogroupings in a population not of European origin. Using HVS-1 data gathered from a sample of 76 elite male and female endurance athletes from Ethiopia and 108 ethnically-matched controls, the authors were unable to find evidence that the haplogroup distribution of the athletes differed significantly from controls (P = 0.63). They concluded that the elite athletes are not a mitochondrially distinct group relative to the general control population. Castro *et al.* (2007) re-examined the link between elite endurance athletes and mitochondrial haplogroupings. Comparing 95 elite male Spanish endurance athletes ($\dot{V}O_2max$: 56-84 mL/(kg·min) to 250 healthy adult male controls, Castro *et al.* could conclude that haplogroup T, and notably a transversion at base pair 13368, was negatively associated with elite endurance status (OR = 0.13; 95% CI [0.02-0.97]). There was only a single athlete belonging to haplogroup T, potentially skewing the results if the haplogroup distribution in the population of elite Spanish endurance athletes was not accurately represented in this sample of 95.

Additionally, Scott *et al.* (2009) genotyped 291 elite Kenyan athletes and divided them into national and international calibre. There were significant differences between the total haplogroup distribution of national-calibre athletes and controls ($P = 0.023$) and international-calibre athletes and controls ($P < 0.001$). International-calibre athletes also showed an overrepresentation of haplogroup L0 relative to controls (OR = 2.37; 95% CI [1.1 - 5.18]) while national-calibre athletes displaying an overrepresentation of M haplogroups (OR = 4.36; 95% CI [1–19.01]). Despite the authors concluding that mtDNA haplogroups are influential in Kenyan running success, population stratification and the opaque methodology in assigning athlete calibre may have confounded the results.

Very few attempts have been made to assess the association between mtDNA haplogroupings and elite sprint athlete status. In 2005, Niemi and Majamaa compared the haplogroup distributions of 89 elite Finnish sprinters to 52 elite Finnish endurance runners (Niemi & Majamaa, 2005). The authors found a significant difference between the total haplogroup frequency distributions of sprint and endurance athletes ($P = 0.039$). The authors also noted greater frequencies of haplogroups J and K in sprint athletes when compared to endurance athletes, though they did not report the results of any statistical tests comparing these frequencies. When the haplogroup counts of the Finnish sprint athletes were compared to the sample taken from ethnically-matched controls ($N = 1060$), a Fisher's exact test using a Monte Carlo simulation with 10,000 replicates to compute the *P*-value indicated that there was no significant difference between Finnish sprint athletes and controls ($P < 1 \times 10^{-4}$).

In the first attempt to assess the relationship between mtDNA and performance in East Asian populations, Mikami *et al.* (2010) examined a broad range of disciplines in elite Japanese athletes. Of these, 79 were elite endurance/middle-power athletes and 60 elite sprint/power athletes. When compared to an ethnically-matched control population of 672, they inferred a significant association between the two sets of athletes and two distinct haplogroups. Associations were found between haplogroups G1 and elite endurance/middle power athletes (OR = 2.52; 95% CI [1.05-6.02]) and haplogroup F and sprint/power athletes (OR = 2.79; 95% CI [1.28-6.07]).

The mitochondrial genomes of elite Spanish athletes were re-examined in 2011 by Nogales-Dagea *et al*. Testing 102 elite endurance athletes, 51 elite power athletes and 478 non-athletic controls, the authors found a significant difference between the endurance cohort and controls. However, no significant differences were found between power athletes and controls or between power athletes and endurance athletes. They were unable to replicate earlier findings in the same population where haplogroup T was underrepresented (Castro *et al*., 2007), and though this found only haplogroup V was overrepresented in endurance athletes. The authors concluded that while variants in mtDNA may influence elite endurance athletes, no association was drawn between mtDNA and elite power athlete status. The failure to replicate the work of Castro *et al*. (2007) in a group of athletes from a similar population highlighted the sensitivity of mtDNA association studies to the chosen methodology. More broadly, these results emphasise the necessity of very large sample sizes to detect a robust association when considering the small effect sizes expected.

1.8.2 Associations between NRY genotypes and athlete status

The NRY makes up 95% of the Y chromosome and contains 63 protein-coding genes largely associated with spermatogenesis (Ensembl, 2014). With a very low rate of recurrent mutation, the NRY provides a valuable system for the reconstruction of population history. Like mtDNA, the NRY does not recombine during meiosis, and a genealogical tree can be created using point polymorphisms. Several associations have been reported in small studies between specific NRY haplotypes and several phenotypes including infertility (Kuroda-Kawaguchi *et al*., 2001), low sperm count (Kuroki *et al*., 1999), prostate cancer (Ewis *et al*., 2006) and high blood pressure (Charchar *et al*., 2012).

The Y chromosome has also been utilised to assess differences in paternal lineages in a case-control framework as a proxy for athlete status, though only appearing once in the literature to date. In 2004, Moran *et al*. analysed the Y chromosome in 62 male Ethiopian athletes from the national senior or junior team, 95 controls from the general Ethiopian population and 85 controls from the Arsi region of the country where a disproportionate number of the athletes originate. After separating the athletes into their preferred events and using contingency exact

test for population differentiation, the authors found significant differences between the haplogroup distributions of general Ethiopian controls and the 5,000m-10,000m runners ($P = 0.048$) and combined endurance runners ($P = 0.042$), though there was no significant difference between track and field athletes and general Ethiopian controls. There were also significant differences between the Arsi controls and the 5,000m-10,000m runners ($P = 0.029$), marathon runners ($P = 0.009$) and combined endurance runners ($P = 0.009$).

Odds ratios were also calculated to assess any overrepresentation or underrepresentation of athlete status for each lineage when compared to the general Ethiopian population and the controls gathered from the Arsi region. The authors concluded that East African paragroup E3* (OR = 5.2; 95% CI [1.3–20.8]) occurred at significantly greater frequency in the 5-10k runners than in the general controls, though there were no significant differences in the individual haplogroup frequencies between 5-10k runners and Arsi controls. Eurasian paragroup K*(xP) was also found to be significantly overrepresented in marathon runners when compared to Arsi controls (OR = 12.1; 95% CI [1.2–122.7]), though a similar relationship was not found when marathon runners were compared to general Ethiopian controls. The authors conclude that certain Y chromosome haplogroups are associated with athlete status, although they are cautious in attributing these findings to either possible functional genetic pathways or simply population stratification. It is very likely that there was not enough statistical power to substantiate any findings given the complex study design and relatively low sample numbers.

1.8.3 Associations between autosomal ancestry and athlete status

We can assume that the African and European ancestors of admixed African-Americans and Jamaicans were adapted to the environments in which they originated. During an admixture event, the genomes of the previously isolated Africans and Europeans recombine, generating long tracts of the admixed genome that can be attributed to one parental population or the other, and because these parental populations originated in different disease environments, they may have different susceptibility to certain diseases (Smith *et al.*, 2004). By calculating the non-random associations of alleles at different loci co-inherited on an intact

chromosomal block (linkage disequilibrium; Chakraborty & Weiss, 1988), it is possible to identify regions harbouring causative gene variants (Winkler *et al.*, 2010). This technique of disease mapping is referred to as admixture mapping and has been used to identify regions of the genome associated with an increased risk of hypertension (Zhu *et al.,* 2005), type 2 diabetes mellitus (Elbein *et al.*, 2009), prostate cancer (Freedman *et al.*, 2006) and renal disease (Kao *et al.,* 2008), though this technique was largely surpassed by GWAS as it became more economically viable to genotype very large sample sizes.

An alternative to admixture mapping is testing for associations between estimates of genome-wide genetic ancestry and a phenotype. By testing for associations in this fashion, it is possible to address more nuanced effects ancestry may have on a phenotype (Flores *et al.,* 2012). The use of genome-wide admixture coefficients in an experimental case-control framework has been used to identify increased risk of asthma (Vergara *et al.,* 2009; Flores *et al.,* 2012), tuberculosis (Serpa *et al.,* 2009), food allergen sensitivity (Kumar *et al.,* 2011) and cardiac arrhythmia (Delaney *et al.,* 2012).

A limited amount of research had investigated any association between ancestry and sport-related phenotypes in admixed populations. Brutsaert *et al.* (2003) found an association between an increase in $\dot{V}O_2max$ and an increase of Spanish admixture in individuals from the high-altitude dwelling Quechua people of Peru. Roy *et al.* (2006) compared $\dot{V}O_2max$ in a sample of European-American and African-American women matched for weight and fat-free mass to explore any influence ancestry has on cardiovascular performance, concluding that African-American women had significantly reduced $\dot{V}O_2max$, haemoglobin levels and muscle oxidative capacity than their European-American peers.

Mapping of admixture disequilibrium can be extrapolated by assuming some of these regions of the admixed genome overrepresented for a given ancestry are actively under selection. In a sample of admixed Puerto Ricans, Tang *et al.* (2007) found three regions of their composite genome to have a deficit in European ancestry when compared to the genomic average. Two of the three regions were identified as harbouring olfactory receptor genes, leading the authors to speculate about potential selective pressures during the slave trade in the development of modern Puerto Ricans. Bryc *et al.* (2010) use a similar method to identify three

different regions in a sample of African-Americans, though they were wary to make claims regarding selection in these regions. More recently, Jin *et al.* (2012) identified regions linked with previously identified selection-candidate genes for hypertension and prostate cancer in a sample of African-Americans (*N* = 1,890), although these loci were not replicated in a much larger data set (*N* = 29,141; Bhatia *et al.,* 2014). The methodology was extrapolated further by creating a synthetic African population from the African ancestral component of the sampled African-Americans and comparing the artificial sample to contemporary Yorubans from Nigeria. The authors notably found a gene associated with reduced malaria resistance in the artificial sample, arriving at the conclusion that the reduced stress for the disease in North America has led to a weakened selection pressure. Although use of admixture mapping for identification of candidate disease-causing genes has been well-established since the 1990s (McKeigue, 1998; Winkler *et al.*, 2010), implying that these regions are not only under selective pressure, but also observable as differences in locus-specific ancestry has been met with resistance (Price *et al.*, 2008).

1.9 Summary

In summary, the literature reviewed in this chapter explored not only the genetic and environmental factors impacting elite sprint athlete status, but also the historical background of the Trans-Atlantic slave trade and slavery in the seventeenth to nineteenth centuries. Together, these factors may help explain the overrepresentation of the West African diaspora in the world of sprint athletics. This chapter focussed predominantly on African-Americans and Jamaicans, the two populations studied in this thesis. Gene and lineage association studies were largely underpowered by sample sizes insufficient to detect what was likely a small effect. These studies were also unable to reproduce significant associations within the same population, suggesting methodological issues such as unaccounted population stratification. African-Americans and Jamaicans were shown to both have varying levels of African and European ancestry and both populations show signs of directional gene flow from Europeans into Africans, an artefact of their shared history of enslavement. It was hoped that introducing the history of the slave trade and slavery would provide a background to the environment and potential selection pressures that both populations may

have experienced. The literature presented here serves as a guide and justification for researching the effects of genetic ancestry on elite sprint athlete status in African-Americans and Jamaicans.

1.10 Aims and Objectives

The overall aim of this thesis is to investigate the effect the Trans-Atlantic slave trade had on elite sprint athlete status in populations of the West African diaspora. The focus will be on ancestrally informative DNA markers and locus-specific ancestry inferences in elite African-American and Jamaican sprint athletes and their respective controls. These findings add to existing efforts aimed at not only investigating any influence the Trans-Atlantic slave trade and chattel slavery had on extreme phenotypes such as elite sprint performance, but also increasing the understanding of the genetic ancestry of African-Americans and Jamaicans.

To achieve this overall aim, this thesis will address the following objectives:

a) To infer the likely origins of modern Jamaicans by comparing their mtDNA haplogroup profile to that from representative regions of Africa. (Chapter 3)
b) To investigate how past demographic pressures during slavery may have shaped modern Jamaicans by comparing their likely origins to a simulated Jamaican population under similar demographic pressures. (Chapter 3)
c) To determine if the distribution of maternal or paternal lineages is different in athletes and controls, and if so, which lineages are overrepresentation or underrepresented. (Chapter 4)
d) To determine if there is a distribution of maternal or paternal lineages associated with athlete status shared between African-Americans and Jamaicans. (Chapter 4)
e) To determine if the distribution of maternal or paternal lineages is different between African-American controls and Jamaican controls. (Chapter 4)
f) To assess if there are any significant differences in the proportions of genome-wide African ancestry between African-American and Jamaican sprint athletes and their respective controls. (Chapter 5)

g) To investigate if there are any significant differences in the time since admixture is estimated to have begun for African-American and Jamaican sprint athletes and their respective controls. (Chapter 5)

h) To assess if any regions of the genome are significantly enriched for either African or European ancestry in African-American and Jamaican athletes and their respective controls. (Chapter 5)

2

Materials and Methods

2.1 Sample collection and classification in case and control cohorts

After providing informed consent, African-American and Jamaican participants were asked to complete a simple questionnaire detailing athletic achievements (if applicable), birthplace, the birthplaces of their parents and the birthplaces of both sets of grandparents. Individuals born outside of their respective country or with any reported relatives born outside the country were excluded from any downstream analyses. One hundred and nineteen self-declared African-American elite athletes from around the United States were sampled during the 2005 National Outdoor Championship and USA Track and Field Annual Meeting and 107 elite athletes from around the entire island of Jamaica were sampled either at training or at their homes. All biological samples collected from African-American athletes were collected by Krista Austin and Jamaican athletes and controls by Yannis Pitsiladis in Jamaica. Sample information for the African-American athlete cohort was anonymised before it was received, while complete sample information was available for the Jamaican athlete cohort. All athletes who took part in this study were competitive athletes and participated in the following events: 100 metres, 200 metres, 400 metres, 100/110[2] metre hurdles, 400 metre hurdles, long jump and triple jump.

Competitive appearances from both groups of athletes ranged from national-level competitions to the Olympics. If personal bests were listed as an athletic achievement in the African-American cohort or searchable online in the Jamaican cohort, these performances were recoded as per the IAAF performance scoring table for athletics (Spiriev, 2011). The IAAF scoring tables contain values out of a potential 1400 and include performances beyond that of current world records.

[2] Women compete over the 100 metre hurdles, while men compete over the 110 metre hurdles

An athlete's highest score was recorded if multiple personal bests were provided. The sample mean for African-American athletes (*N* = 56) was 1241.61 with a standard deviation of 41.04, while the sample mean for the Jamaicans athletes (*N* = 91) was 1154.90 with a standard deviation 84.05.

For the purpose of comparison, 45 self-declared African-American control samples were gathered from around the United States, including San Diego, California; Springfield, Massachusetts and Americus, Georgia. A sample of 293 Jamaican controls representative of the Jamaican population in the geographical distribution of their places of birth were sampled at the University of the West Indies in the Kingston suburb of Mona. Genotyped samples are summarised in Table 2.1.

As explained in greater detail in Section 5.3, eleven African-American athletes were estimated to have less than 5% genome-wide African ancestry. African-Americans are a self-declared ethnic group, and while it is not uncommon for African-Americans to have very little African ancestry (Bryc *et al.*, 2015), there were not comparable controls with similarly low levels of African ancestry. These eleven athletes were excluded from the analyses presented in Chapters 4 and 5.

Table 2.1 Number of genotyped ahletes and controls from the African-American and Jamaican samples used in this thesis. African-American control data was gathered from the literature cited. The disparity in sample sizes between the mtDNA and NRY data sets and the Autosomal data is largely due to the quality of the extracted whole DNA explained in Section 2.2.2 below and the cost of the whole-genome SNP arrays, outlined in Section 5.2.1.

		mtDNA	**NRY**	**Autosome**
African-American	Athlete	105	50	108
	Control	1642*	246†	397‡
Jamaican	Athlete	107	44	93
	Control	293	47	101

* Monson *et al.* (2002); Diegoli *et al.* (2009); Lao *et al.* (2010)
† Lao *et al.* (2010)
‡ Includes 350 from Cheng *et al.* (2011)

2.2 Molecular data collection

All participating individuals were shown the appropriate sampling procedure before collection began. Subjects placed specially designed, long, slender, sterile plastic dowels with detachable fabric swabs (Puregene, Gentra Systems, Minneapolis, MN, USA) into their mouths and vigorously rubbed the insides of their

cheeks for roughly 20 seconds to collect buccal cells. These swabs were then collected from all subjects, detached from the dowel and placed into individually sealed tubes containing a stabilising cell lysis solution for storage. Samples were then appropriately labelled and kept at –20°C until needed for further analysis or at -80°C for long-term storage.

2.2.1 Extraction of whole DNA from buccal swabs

DNA was extracted from the buccal swabs using the Qiagen QIAamp DNA Mini kit (Qiagen Ltd., Crawley, UK). Once all samples and consumables had reached room temperature, 200 µl of cell lysis solution suspending the buccal swab was removed from the original container[3] and placed in a 1.5ml microcentrifuge tube. Proteinase K (20 µl) was added and mixed to ensure proper reaction between the sample and the digestive enzyme to deactivate any nuclease which may degrade the DNA during purification. Next, 200 µl Buffer AL, a lysis buffer and binding agent, was added to the solution and vortexed for a minimum of 15 seconds to ensure a homogeneous mixture.

After 10 minutes of incubation at room temperature, the solution was then briefly spun down in a centrifuge to remove any condensation that may have formed on the lid of the microfuge tube. Pure ethanol (200 µl) was then added and mixed again for 15 seconds before the solution was briefly spun down once more to remove any liquid that may have accumulated on the lid. The entirety of this solution was then carefully pipetted into a QIAamp Mini spin column filter resting atop a 2 ml collection tube. When DNA passes through this silica-containing filter, chaotropic salts found in the lysis buffer destabilise the hydrogen bonds in the solution and ultimately disrupt the suspension of DNA in water allowing it bind to the silica in the spin column. After closing the column's lid, the sample was then centrifuged at 8,000 rpm for one minute passing the solution through the silica, leaving the DNA in the filter of column while the remainder of the solution is free to pass through. The spin column was then removed and placed on to the mouth a clean 2 ml collection tube while the tube containing the filtrate was discarded.

[3] In cases where the buccal swab was suspended in ethanol, the ethanol was first evaporated off at 37°C before adding 1 ml of cell lysis solution and thoroughly vortexing. Samples were determined to be suspended in ethanol if they were not frozen upon retrieving them from the -20°C freezer. Additionally, a visual inspection was performed on thawed sampled. Cell lysis solution would readily form bubbles when shaken, while ethanol would not.

500 µl Buffer AW1 was added to the spin column to denature and digest any proteins that may have collected on the filter. The solution was forced through the filter by spinning the column for an additional minute at 8,000 rpm. After centrifugation, the new tube containing the filtrate was discarded and the column is placed on another new clean tube. Next, 500 µl Buffer AW2 was added to the spin column to remove any salts which may have collected on the filter and spun at 14,000 rpm for three minutes. Once complete, the filtrate is once again discarded and the spin column is place inside a 1.5 ml microcentrifuge tube. Finally, 200 µl Buffer AE elution buffer was added to the column and allowed to incubate for at room temperature for one minute before centrifuging at 8,000 rpm for one minute. The spin column was then discarded and the resulting microfuge tube was labelled appropriately then stored at -20°C for future use or at –80°C for long-term storage.

2.2.2 DNA concentration

The quality of the extracted samples was then assessed using the eight-channel NanoDrop 8000 (NanoDrop Technology®, Cambridge, UK), a UV micro-spectrophotometer sample retention system that calculates the concentration of DNA present in the sample. First, the sample was left to reach room temperature before being vortexed to ensure a homogeneous sampling. Next, both the top and bottom portions of the fibre optic pedestal were scrubbed vigorously with a nonabrasive, anti-static, low lint cellulose fibre wipe to remove any residue from the previous use. After opening the accompanying NanoDrop 8000 software on a connected laptop, 1 µl of deionised water was placed on the lower portion of the pedestal. The lever arm was then lowered, resting close enough to create a column of liquid between both lower and upper fibre optic receptors. After initialising the machine, the receptors were then scrubbed before adding 1 µl Buffer AE to the pedestal to blank the system by lowering the lever arm and initialising the machine again. The surface was scrubbed once more and the system made ready to measure the quality of the samples. Extracted DNA (1 µl) was pipetted onto the pedestal, the lever arm lowered and the machine initialised. After the values were recorded using the software, the fibre optic pedestals were scrubbed clean in preparation for the next batch. Each sample was

repeated at least twice to give a more accurate measure of the amount of DNA present in the sample.

The above qualification was especially important to the whole-genome sequencing mentioned in Section 5.2.1, where the amount of DNA present in each sample was critical to generate high quality output. Rather discouragingly, the values generated using the NanoDrop 8000 produced a weak relationship with the values later quantified by the PicoGreen method (Invitrogen Ltd., Paisley, UK) used by collaborators at the Tokyo Metropolitan Institute of Gerontology. This is plotted in Figure 2.1. The main disadvantage of the Nanodrop UV spectroscopy system used above was the inclusion of single-stranded DNA and other contaminants such as proteins and extraction buffers in the estimate of double-stranded DNA concentration (Keer & Birch, 2008). The PicoGreen reagent binds to double-stranded DNA and creates a maximum emission at 530 nm and has been shown to provide not only better measurement accuracy but also better consistency (English *et al.*, 2006). The observed pattern of dispersal may also indicate that the DNA was not fully dissolved into the buffering solution.

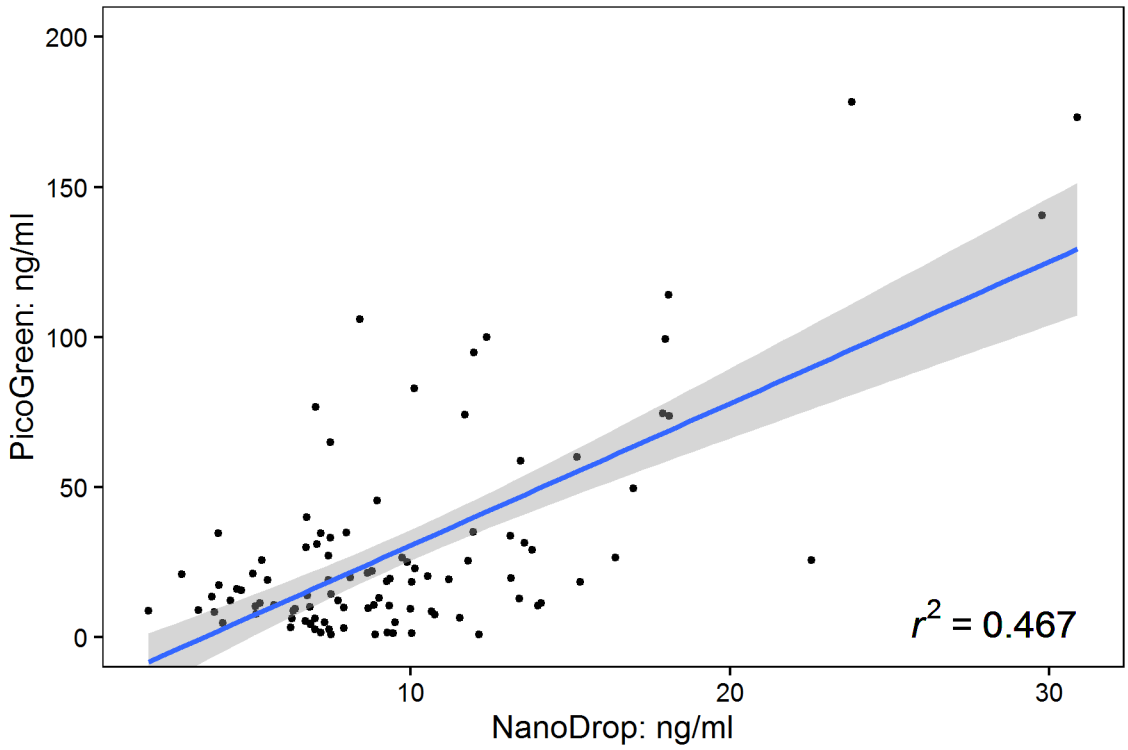


Figure 2.1 Comparison of DNA concentrations generated by the NanoDrop microspectrophotometer and PicoGreen assay. Each point represents an individual sample. The coefficient of determination (r^2) indicates that the values generated for the NanoDrop were only able to explain 46.7% of the variance. Shaded area represents a 95% confidence interval for the line of best fit, shown in blue.

2.3 Mitochondrial DNA

HVS-1 was amplified and sequenced to identify patterns of polymorphisms, known as HVS-1 motifs. The reverse strands of mtDNA were sequenced using fluorescent dideoxynucleotides (ddNTPs) in the BigDye™ terminator cycle sequencing ready reaction (Applied Biosystems, Foster City, CA, USA). A reverse primer was used to generate all sequences and when necessary, a forward primer was also used to read the sequence on the 5' end of the polycytosine tract between base pairs 16184-16193, according to the revised Cambridge reference sequence for mtDNA (rCRS (Andrews *et al.*, 1999), NCBI Reference Sequence: NC_012920.1). These primers can be found in Table 2.2. HVS-1 polymerase chain reaction (PCR) sequencing was performed from a 20 µl total volume, containing 10µL of 2x ReddyMix PCR master mix (Thermo Scientific, Surrey, UK), 1 µl of 0.2 µM primer and 2 µl of total DNA (estimated at 10 ng, based on Qiagen QIAamp DNA Mini kit documentation and tested averages). The remaining 6 µl were composed of autoclaved, deionised, ultraviolet-treated water. DNA was denatured for 5 minutes at 94ºC followed by 30 cycles of denaturation (30 seconds, 94ºC), annealing (45 seconds, 51ºC) and elongation (60 seconds, 72ºC). A final elongation step was performed (10 minutes, 72ºC) before samples were held at 4ºC. Negative controls with no DNA were included in each sample run to aid in the detection of contamination. PCR product was then run on a 2% agarose electrophoresis gel containing ultraviolet-fluorescent nucleic acid stain ethidium bromide to determine whether the reaction had been successful. Sequencing products were separated by 5% denatured Long Ranger™ gel (FMC Bio-Products, Rockland, ME, USA) and base pairs were detected using a PE Applied Biosystems 377 DNA sequencer. Sequencing was performed by the Julie Galbraith of the Sir Henry Wellcome Functional Genomics Facility at the University of Glasgow.

The resulting sequence electropherograms were read using Chromas (version 1.45, Technelysium, Australia). The length polymorphism found between base pairs 16184-16193 was controlled for by manually inserting or removing cytosines to standardise the overall sequence length. Differences from rCRS were recorded manually to establish the HVS-1 motif necessary to haplogroup each sample.

For further resolution of ambiguous haplotype motifs, RFLPs were used to genotype coding region mtSNPs. Three mtSNPs were chosen to refine African haplogroups based on the phylogenetic tree of recorded global whole-genome

mitochondrial diversity (van Oven & Kayser, 2009). The transition G10398A was used here predominately to resolve macro-haplogroup N, though it can be used to distinguish finer-resolution of Sub-Saharan haplogroups L0d1b1, L1c1a1 and L3e1a3. The transition C10400T was used to resolve macro-haplogroup M. Finally, the transition 13803 was used to resolve the Sub-Saharan haplogroup L2. Primers were designed using the NCBI Primer-BLAST (Ye *et al.*, 2012) and checked against the revised Cambridge Reference Sequence of human mtDNA (rCRS) for any typographic errors (Table 2.2). PCR cycles were identical to that of the HVS-1 protocol. Although the various primers' melting temperatures were held in consideration, the annealing temperature produced satisfactory results at 51°C. The resulting PCR products then had restriction enzymes appropriate to the mtSNP added before being left to incubate at 37°C overnight. After digestion, the product was then run on a previously described agarose electrophoresis gel to determine whether the enzyme had cleaved the PCR product, indicative of an mtSNP. This information was recorded and collated with the HVS-1 data to inform haplogroup assignment further.

The HVS-1 motif and any additional genotyping data were used to haplogroup these data in accordance with the comprehensive full mtDNA genome phylogenetic tree Phylotree Build 11 (7 Feb 2011) (van Oven & Kayser 2009) and the comprehensive mtDNA database published alongside the Genographic Project (Behar *et al.*, 2007) at a comparable resolution to that in the literature (Salas *et al.*, 2005b). To monitor errors that could arise from the sequencing procedure, motif inconsistencies were investigated by comparing potentially incorrect loci against a list of known problems and inconsistencies found in published sequence data (Bandelt *et al.*, 2002; Bandelt *et al.*, 2004a; Bandelt *et al.*, 2004b; Salas *et al.*, 2005a; Salas *et al.*, 2007; Yao *et al.*, 2009). Sequencing electropherograms were rechecked when such phylogenetic inconsistencies were observed. DNA samples were extracted again or resequenced when doubt persisted regarding the variant in question. Additional coding-region mtSNP genotyping performed as part of larger SNP array data (outlined in Section 5.2.1) also allowed for increased confidence during haplogroup assignment.

2.4 General statistical methodology

Many statistical tests were performed in this thesis, though a majority were experiment dependent. Exact tests of population differentiation, however, were used at multiple points throughout this thesis. An exact test of population differentiation tests for the random distribution of individuals between pairs of populations under the assumption of random mating. This test is analogous to a Fisher's exact test on a 2 × 2 contingency table though extended to 2 × k, where k is the number of haplotypes in a sample (Raymond & Rousset, 1995). These were performed with Arlequin 3.5 (Excoffier & Lischer, 2010) using a Markov–Chain Monte–Carlo simulation of chain length 1×10^6 with 1×10^5 dememorisation or burn-in iterations. Longer chain lengths allow for greater precision for the exact *P*-value and standard deviation. A critical α of 0.05 was chosen. Values below 0.05 then suggested that the distributions were significantly different.

Unless stated otherwise, the open-source statistical environment R (R Core Development Team, 2008) was used for all statistical analyses. All figures were generated using the R graphics package ggplot2 (Wickham, 2009), apart from those figures sourced from the literature.

Table 2.2 Primers used to generate the first hypervariable segment sequence data of mtDNA and coding region RFLPs used in this thesis

	Site	Enzyme	Primer	5'-3'	Annealing Temperature (ºC)
HVS-1 15,975 – 16,450*[4]			Forward	CTCCACCATTAGCACCCAAAGC	51
			Reverse	CGAGGAGAGTAGCACTCTTG	
Coding RFLP					
	10398	-DDeI	Forward	CCCTACCATGAGCCCTAC	51
			Reverse	CTAGGCATAGTAGGGAGGAT	
	10400	+AluI	Forward	CCCTACCATGAGCCCTAC	51
			Reverse	CTAGGCATAGTAGGGAGGAT	
	13803	+HaeIII	Forward	ATATCATACACAAACGCCTG	51
			Reverse	GTTGAGGTCTAGGGCTGTTA	

[4] Base pairs according to revised Cambridge Reference Sequence (rCRS, Andrews *et al*. 1999)

3

What are the African origins of modern Jamaicans and what does this suggest about selection pressures during slavery?

3.1 Introduction

As introduced in Section 1.6, several researchers have posited that selection pressures from slavery may have resulted in the overrepresentation of the West African diaspora in sprint athletics (Morrison & Cooper, 2006; Aiken, 2011). While their hypotheses are based on biological observations in these populations, they do not test how slavery-moderated selection may have affected the slaves and their descendants. An attempt was made in this chapter to assess the selection pressures that may have affected the slaves on Jamaica. By understanding the selection pressures in the ancestors of modern Jamaicans, it was hoped to add to the existing research attempting to identify the unique combination of factors that produce elite sprint athletes.

Detailed historical records from the Trans-Atlantic slave trade have been synthesised over the past 50 years, giving researchers valuable demographic data on numerous slave populations by year and region of origin (Eltis & Richardson, 2010). The African diaspora in the New World provides the unique opportunity to understand the likely selection pressures imposed on the Africans forcibly relocated during the Trans-Atlantic slave trade. Numerous attempts have been made to trace the likely origins of former slave groups back to regions of Africa by analysing genetic data collected from modern African populations and the African diaspora together with slave importation records.

The distribution of mitochondrial haplogroups is very sensitive to the effects of minor genetic drift and population isolation, resulting in haplogroup frequency

differences over even small geographic distances (Samuels *et al.*, 2006). Comparing haplogroup distributions between groups of interest and presumptive source populations can facilitate investigation into the likely origins of the descendant populations. The first published attempt to trace the mtDNA haplogroups of recent African ancestry in North, Central and South America back to regions of Africa was done using a multinomial sampling algorithm to calculate intra-continental admixture coefficients for broad regions of Africa based on mtDNA haplogroup distributions (Salas *et al.*, 2004). The resulting intra-continental admixture coefficients agreed with historical evidence (Eltis & Richardson, 2010). The mtDNA haplogroup profiles of the African diaspora in North and Central America resembled the groups found adjacent to the recorded West African trading ports used by the Spanish and British slave traders, while the mtDNA haplogroup profiles of South Americans of African descent largely resembled prominent Portuguese slaving areas in West-central Africa (Salas *et al.*, 2004). Probably because the estimations were based on comparing haplogroup proportions rather than direct sequence data, the authors highlight the inability to trace individual lineages to localities within the African continent due to the observably broad dispersal of individual mtDNA sequences across Africa (Salas *et al.*, 2002).

3.1.1 Inferred matrilineal origins of African-Americans

As outlined in Sections 1.5.4 and 1.5.5, the living and working conditions of slaves in the United States differed greatly from groups in the Caribbean. North American planters purchased an estimated 300,000 slaves before the abolition of the slave trade in 1808 (Figure 3.1A, based on Eltis and Richardson, 2010). Slaves in North America quickly began to reproduce above the levels of replacement (Morgan, 2007), reaching an estimated total population of 2,762,398 in 1860, five years before the abolition of slavery in 1865 (United States Bureau of the Census, 1975). In a continuation of earlier work by Salas *et al.* (2004), Salas *et al.* (2005) compared the SWGDAM database (Section 1.7.1; Allard *et al.*, 2005) to a much larger database of 4,860 mtDNAs from around Africa (Salas *et al.*, 2005b). Their results were also in agreement with the historical record, showing that greater than 55% of African-American lineages stem from modern Western Africa, with fewer than 41% coming from West-central or South-western Africa. These results

must be interpreted with caution however, as the authors' division of the African continent aggregated areas with large amounts of mtDNA variability not equally represented in the slave trade (Harich *et al.*, 2010).

In a more recent effort utilising mtDNA to trace the likely African origin of slave-descendant populations in North, Central and South America, Stefflova *et al.* (2011) attempted to identify a link between inferred intra-continental African ancestry and the contemporary colonising European nations. Using a representative sample of African-Americans from around the United States, in addition to a sample from Philadelphia, Pennsylvania, the authors found that the composite African-American population had a greater affinity with mtDNA found in modern Senegambia (a former British-controlled region) than the rest of the African coast, while African-Americans in Philadelphia showed a greater affinity to the mtDNAs from the Bights of Benin and Biafra along the West African coast. Although a connection was found between the colonial systems, the authors' sample from Philadelphia was treated as if the city were importing slaves itself. In reality, only 912 recorded slaves arrived at the port in Philadelphia between 1759 and 1800, although the regions of origin for those slaves are missing from the Trans-Atlantic slave trade database (Eltis & Richardson, 2010). African-Americans migrated into the industrialised Northeast during the reconstruction of the nation after the Civil War (Farley, 1968; Johnson & Campbell, 1981; Tanner & Reiff, 1995), and very few African-American populations can simply be assumed to have remained static since the slave era (Ely *et al.*, 2006).

3.1.2 Inferred matrilineal origins of Jamaicans

The island of Jamaica was sparsely inhabited by an indigenous population when the first Spanish settlers arrived in 1509. However, by the time of the English conquest of Jamaica in 1655, these indigenous people had either fled the island or had their populations had reduced greatly by forced labour and diseases introduced by the Spanish (Whitehead, 1999). While only an estimated 670 recorded African slaves disembarked on the island under Spanish rule between 1600 and 1613 (Eltis & Richardson, 2010), the English soon established intensive slave-labour sugar plantations like those proving profitable on Barbados (Higman, 1976). An estimated 930,500 African slaves disembarked on the island between

1655 and the abolition of the slave trade in 1807 (Figure 3.1B, Eltis & Richardson, 2010). However, a slave census in 1808 estimated the slave population only numbered 354,000. By the abolition of slavery in 1834, the population dropped further to 311,070 (Higman, 1976). The obvious disparity in these values indicates an environment not conducive to reproduction above the level of replacement in the slave population of Jamaica.

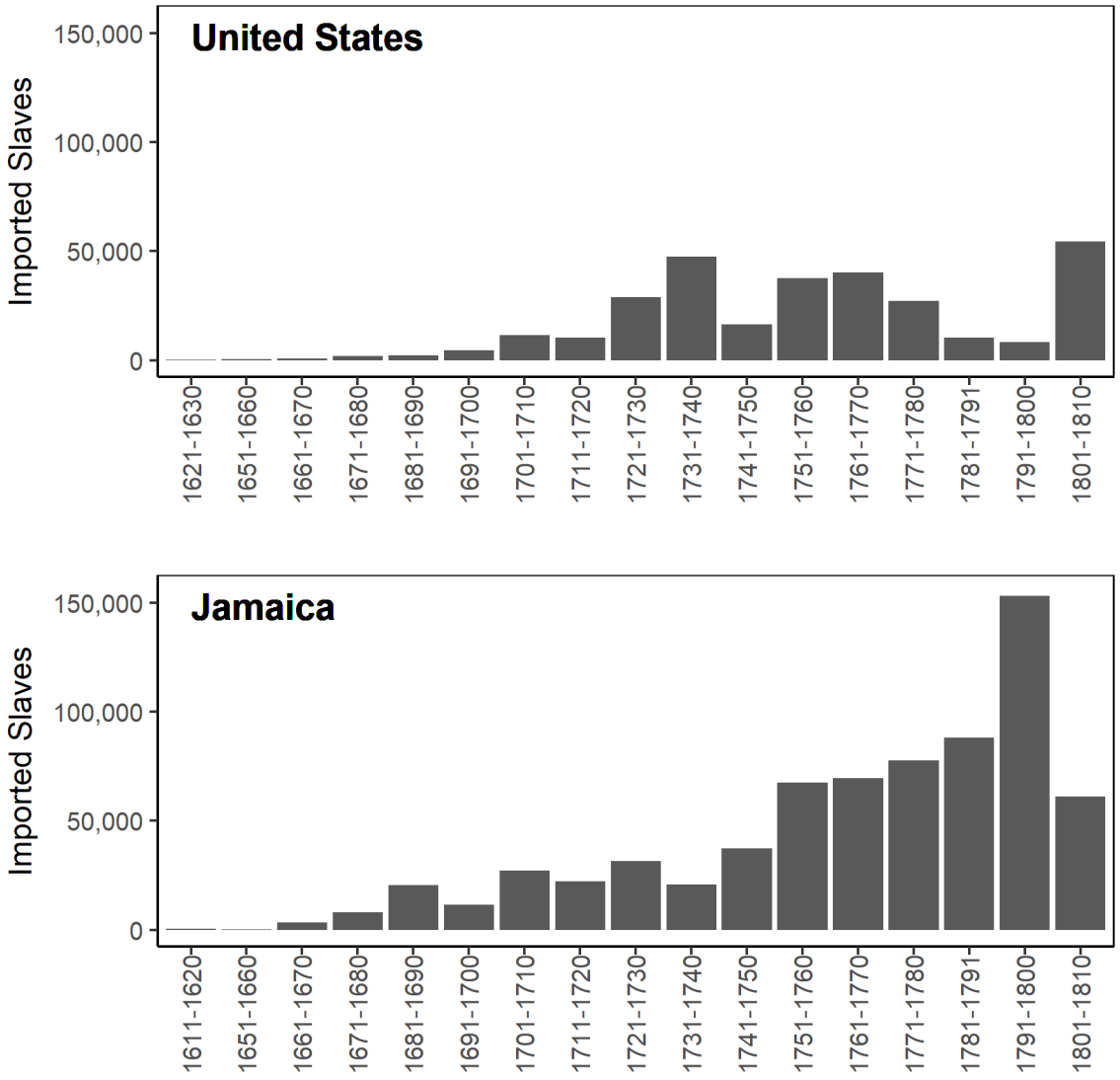


Figure 3.1 Imputed number of slaves imported to the United States and Jamaica during the Trans-Atlantic slave trade (Eltis & Richardson, 2010). Only years of legal trade are considered. The slave trade was abolished in 1808 and 1807 for the United States and Jamaica, respectively), and the number of slaves imported in the last decade only represents seven and six years of trading, respectively. As a result, the final decade is not entirely representative of the importation trend at the time and likely underestimating the number of slaves imported were the trade to continue to 1810.

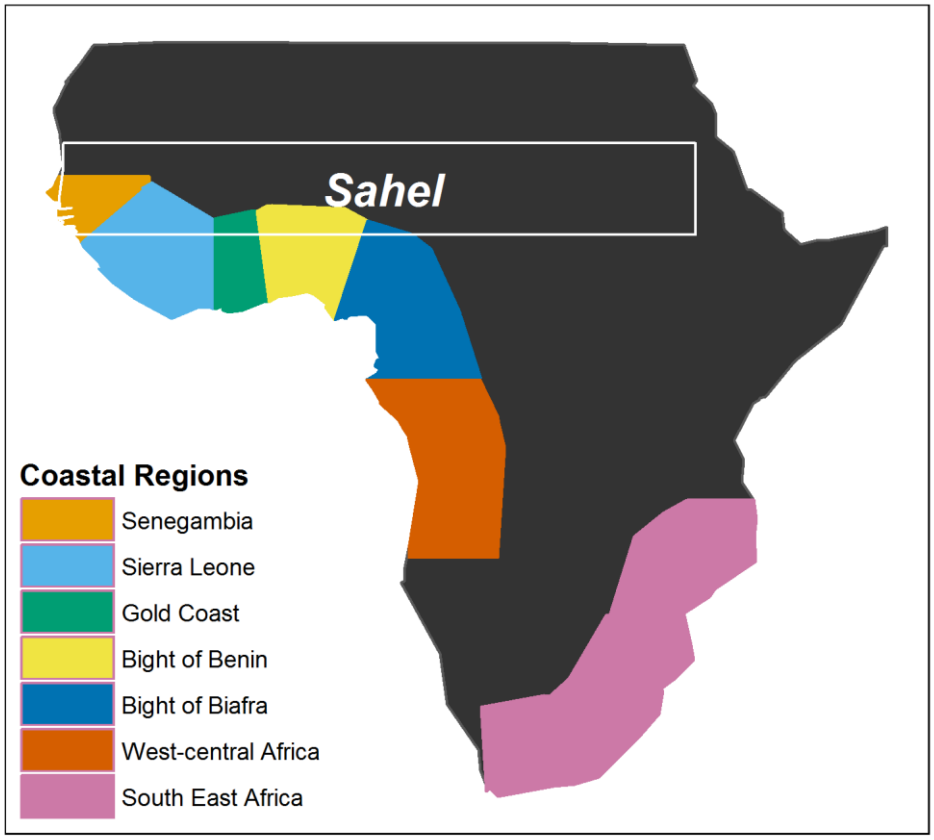


Figure 3.2 Schematic of the African coastal regions used in this analysis. The semi-arid Sahel is outlined in white. The region between the end of the West-central African coast and the beginning of the South East African coast is the desert coastline region of modern Angola and Namibia that saw no recorded trade activity (Manning, 1983).

Between 1655 and 1807, active slave trading ports could be found from Senegambia to Madagascar. The ports in these regions (Table 3.1, Figure 3.2 and Figure 3.3) were not uniformly active between 1655 and 1807 in shipping slaves to Jamaica. The observed variability in shipping locations resulted from several economic pressures. Slavery was widespread in Africa largely because people were the only form of private, revenue-producing property recognised across the continent (Thornton, 1992). The arrival of trade with Europe in the 15th century introduced an additional external market for an already common practice. Highly regulatory European trading companies established forts and slave factories to prepare the captured Africans for transport across the Atlantic that increasingly led to civil unrest among participating African nations. The increasing volume of trade with European slave traders likely led to the enslavement of people by the ruling groups. While most groups would sell prisoners acquired during conflicts, some populations went as far as undertaking specialised military campaigns to

procure more slaves for the Atlantic trade (Klein, 1999). Inter-regional political variability also led to changes in slave exportation. The most developed areas of Africa at the time of European contact were also those most likely to be drawn into the Trans-Atlantic slave trade as only the most politically unified and densely populated areas could foster the early commodity trade with Europe. When European merchants' interests turned to slaves, these developed African groups were most directly affected because they had the population capacity to respond to the explosive increase in demand during the middle and late eighteenth century (Nunn, 2008).

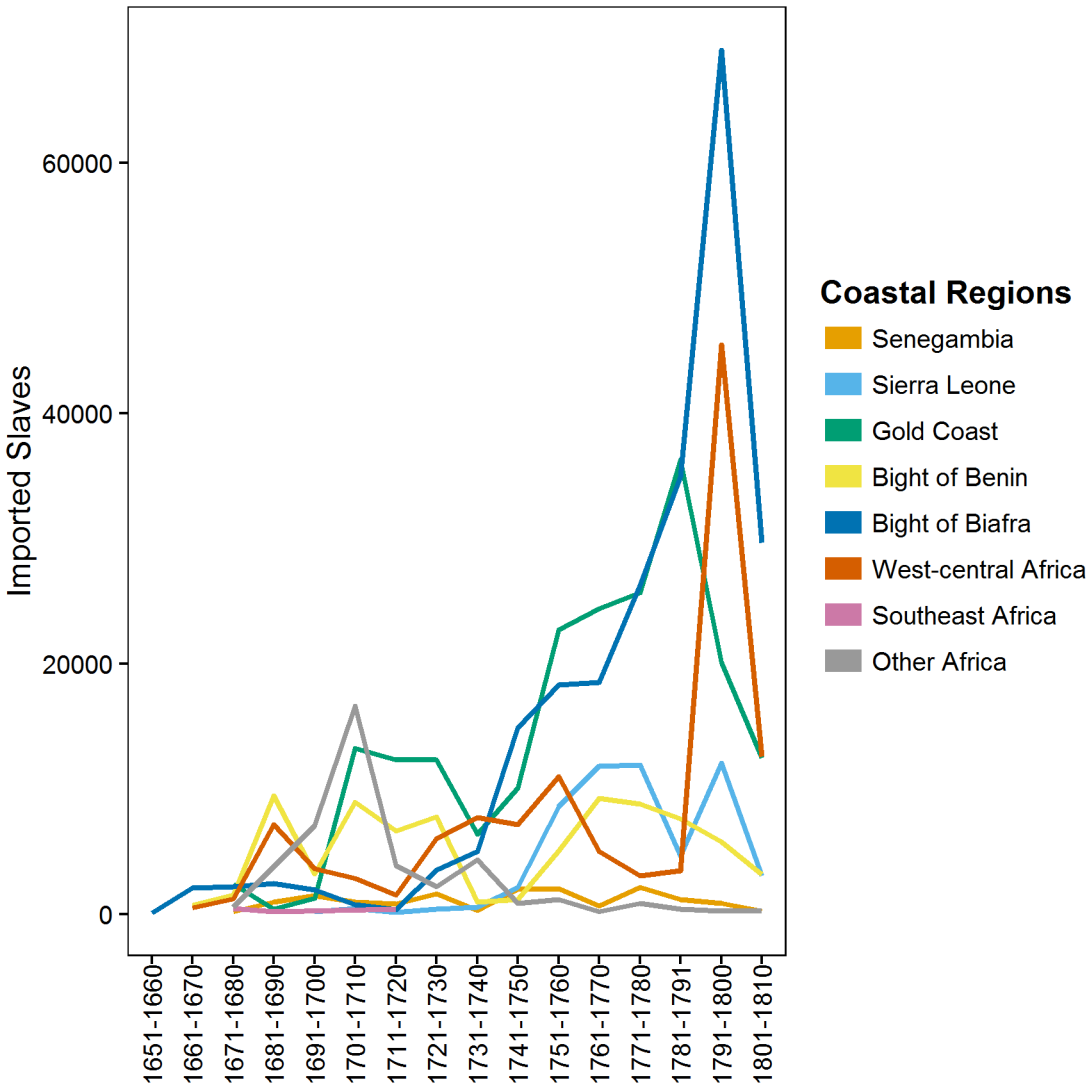


Figure 3.3 Line plot of imported number of slaves to Jamaica. Data from each African region considered in this analysis was clustered into 10-year intervals. Only those years of legal trade are considered. Contributory African coasts correspond to the map of Africa in Figure 3.2.

Earlier mtDNA studies of Jamaica suggest an almost entirely West African matrilineal origin with limited genetic contribution from either Eurasian or Asian/New World maternal lineages (McLean *et al.*, 2005; Torres *et al.*, 2007). This is consistent with the historical accounts of not only a high Sub-Saharan slave importation rate, but also the small population and forced endogamy of the European women on the island (Burnard, 1999; Burnard, 2006) as well as the great reduction in the population of indigenous groups shortly after contact with the Spanish (Whitehead, 1999).

The aim of this study is to investigate the origins of modern Jamaicans by applying an intra-continental admixture model to compare their mtDNA haplogroup distribution with presumptive source coasts of Africa. By comparing these intra-continental admixture coefficients to expected coefficients, insights into the demographic constraints on the slaves during the slave era in Jamaica were also sought. When considering the overwhelming proportion of slaves imported to Jamaica from the Bight of Biafra and West-central Africa immediately before the end of the slave trade, it is hypothesised that the mtDNA haplogroup profile would resemble these latter sources more closely than regions exploited earlier in the slave trade.

3.2 Methods

3.2.1 Sampled Jamaicans

To investigate the origins of modern Jamaicans and the selection pressures the slave population may have experienced, buccal cells were collected from 400 Jamaicans from around the island, including 107 athletes. These groups are described in Section 2.1 above. DNA from buccal cells was extracted and isolated before the mtDNA was amplified and sequenced, as described in Section 2.2.

3.2.2 Generation of Mitochondrial haplogroup data

Mitochondrial sequence data were then classified into 16 haplogroups to allow for a comparison with publicly available data from modern African groups to be made. The mtDNA haplogrouping process is described in Section 2.3 above.

3.2.3 Processing of Trans-Atlantic slave trade database

The number of slaves arriving in Jamaica and their points of origin were extracted from the comprehensive Trans-Atlantic slave trade database (www.slavevoyages.org; Eltis and Richardson 2010). This database is the most complete collection of slave trade data available, containing the records of 34,948 voyages between 1514 and 1866. It is estimated that over 95% of all British voyages are included in the database (Eltis, 2010). To facilitate the broader application of the database, the authors imputed observations for any missing values in the shipping records using the observed data. For example, if the departure and arrival destinations and the number of slaves departing are known, but the number of slaves arriving is not, data from similar vessels were used to infer an approximate value.

Only records indicating Jamaica as the imputed principal region of arrival were used in these analyses. The imputed principal region of slave purchase was used to define the departure location of each voyage. The imputed number of slaves represented the number of slaves arriving at Jamaican ports. Any voyages indicating "Other Africa" as the imputed principal region of importation were excluded from the intra-continental admixture analysis due to the ambiguity of departure locations.

Table 3.1 Imputed number of slave disembarked in Jamaica between 1655 and 1807 for voyages with known embarkation points. Only those years of legal trade are considered below. Percentages for each time-period are presented in brackets.

	Senegambia	Sierra Leone	Gold Coast	Bight of Benin	Bight of Biafra	West-central Africa	Southeast Africa	Unknown	Total
1651-1660	-	-	-	-	85 (100%)	-	-	-	85
1661-1670	-	-	-	704 (12.3%)	2101 (36.7%)	500 (8.73%)	-	2420 (42.27%)	5725
1671-1680	158 (1.47%)	-	2402 (22.36%)	1527 (14.21%)	2204 (20.52%)	1211 (11.27%)	462 (4.3%)	2779 (25.87%)	10743
1681-1690	960 (3.75%)	-	412 (1.61%)	9465 (36.93%)	2445 (9.54%)	7139 (27.86%)	171 (0.67%)	5037 (19.65%)	25629
1691-1700	1447 (6.13%)	205 (0.87%)	1270 (5.38%)	3181 (13.48%)	1926 (8.16%)	3593 (15.23%)	-	11977 (50.75%)	23599
1701-1710	928 (1.95%)	413 (0.87%)	13225 (27.81%)	8918 (18.76%)	730 (1.54%)	2843 (5.98%)	-	20491 (43.1%)	47548
1711-1720	813 (1.82%)	125 (0.28%)	12320 (27.51%)	6634 (14.81%)	360 (0.8%)	1497 (3.34%)	385 (0.86%)	22654 (50.58%)	44788
1721-1730	1603 (2.45%)	381 (0.58%)	12294 (18.81%)	7758 (11.87%)	3512 (5.37%)	6003 (9.18%)	-	33819 (51.73%)	65370
1731-1740	302 (0.46%)	559 (0.86%)	6389 (9.82%)	954 (1.47%)	5013 (7.7%)	7711 (11.85%)	-	44157 (67.85%)	65085
1741-1750	1993 (3.11%)	2142 (3.34%)	10064 (15.71%)	1155 (1.8%)	14869 (23.21%)	7134 (11.14%)	-	26711 (41.69%)	64068
1751-1760	2050 (2.64%)	8592 (11.07%)	22692 (29.23%)	5033 (6.48%)	18280 (23.55%)	10995 (14.16%)	-	9980 (12.86%)	77622
1761-1770	633 (0.79%)	11822 (14.7%)	24388 (30.32%)	9264 (11.52%)	18500 (23%)	4983 (6.19%)	-	10847 (13.49%)	80437
1771-1780	2114 (2.16%)	11880 (12.15%)	25662 (26.24%)	8802 (9%)	26305 (26.9%)	3052 (3.12%)	-	19968 (20.42%)	97783
1781-1791	1135 (1.2%)	4647 (4.93%)	36288 (38.48%)	7617 (8.08%)	34956 (37.07%)	3439 (3.65%)	-	6221 (6.6%)	94303
1791-1800	828 (0.53%)	12076 (7.66%)	20091 (12.75%)	5773 (3.66%)	69046 (43.82%)	45500 (28.87%)	-	4268 (2.71%)	157582
1801-1810	246 (0.37%)	3046 (4.56%)	12479 (18.68%)	3161 (4.73%)	29664 (44.41%)	12553 (18.79%)	-	5645 (8.45%)	66794
Total	15210 (1.64%)	55888 (6.03%)	199976 (21.57%)	79946 (8.62%)	229996 (24.81%)	118153 (12.74%)	1018 (0.11%)	226974 (24.48%)	927161

3.2.4 African populations and HVS-1 sequence data used for comparison

To permit the comparison of Jamaican mtDNAs with likely source populations, a dataset of 9,265 African HVS-1 sequences was collated from the literature to create representative mtDNA haplogroup distributions for each relevant African coast. African ethnic groups were assigned locations based on present day ranges with historic guidelines for the differing coastal boundaries superimposed. When the location of a sample collection was not specified in the accompanying literature, the ethnic group's present location according to the language atlas Ethnologue (Lewis, 2009) was used. The following eight historic regional definitions were used for dividing the African coastline:

- Senegambia comprised the area between the mouth of the Senegal River to the Nuñez River near Boké, Guinea.
- Sierra Leone comprised the area from the Nuñez River up to and including the Assini River in Côte d'Ivoire.
- Gold Coast ran east of the Assini River up to and including the Volta River.
- The Bight of Benin covered the Volta River in Ghana to the Nun River in the Niger delta of Nigeria.
- The Bight of Biafra ran east of the Nun River to Cape Lopez in modern Gabon.
- West-central Africa was defined as the rest of the western coast of the continent south of Cape Lopez until Tombua in southern Angola.
- Southeast Africa comprised anywhere east of the Cape of Good Hope up to the Rio Rovuma in northern Mozambique.

These coasts are defined in the Trans-Atlantic slave trade database (Eltis & Richardson, 2010) and are presented graphically in Figure 3.2. The ethnic groups contributing to the mtDNAs in each region are summarised in Section 7.2.

3.2.5 Genetic diversity indices of Jamaican and African mtDNAs

To investigate the internal genetic diversity of the studied source regions' mtDNA pools, the following diversity indices were calculated: number of different haplotypes (k), number of polymorphic sites (S) and mean number of pairwise differences between nucleotide sequences (π; Tajima, 1983). Demographic properties of the population samples were assessed using Tajima's D statistic, a

function comparing the number of polymorphic sites in a sample with the mean number of pairwise differences between haplotypes (Tajima, 1989). To test for any evidence of sudden population expansion, a mismatch analysis with 1,000 bootstrap replicates was performed to test for a unimodal distribution of pairwise differences. This analysis was done by calculating an index of raggedness around the mismatch distribution (RI) (Harpending, 1994) and sum of squared deviations (SSD) of the difference between the observed and the expected mismatch distributions (Schneider & Excoffier, 1999). All diversity indices were calculated using Arlequin 3.5 (Excoffier & Lischer, 2010).

3.2.6 Intra-continental admixture methodology

Intra-continental admixture coefficients based on haplogroup distributions were fitted to a model using a Markov-Chain Monte Carlo posterior sampling method assuming a multinomial distribution for the mtDNA haplogroup profiles. The model was fitted to quantify the magnitude of the impact of each African region on haplogroup frequencies in Jamaica. The number of mtDNAs in each assigned haplogroup present in the sample from Jamaica was defined as $n_i, (i = 1, \dots, 16)$, where 16 is the maximum number of classified haplogroups present in Jamaica. These clusters were assumed to be drawn from a multinomial distribution with parameters N and p_i, where N equalled the sample size of Jamaica: $N = \sum_{i=1}^{16} n_i$, and $p_i = \sum_{j=1}^{8} \alpha_j f_{ji}$ $(1 \leq i \leq 16)$, where p_i is the proportion of the i^{th} haplogroup in Jamaica, 8 is the number of source regions in Africa, $\mathrm{f_{ji}}$ is the frequency of the i^{th} haplogroup in the j^{th} source region, and the α_j are the intra-continental admixture coefficients. This model was analysed in a Bayesian framework, which meant that the distribution of the intra-continental admixture coefficients was explored through interpolation with the given data. The prior distribution of the intra-continental admixture coefficients was uniform on $\alpha_j \geq 0, \sum_j \alpha_j = 1$. The posterior distribution of α_j was explored with the Metropolis-Hastings algorithm and was summarised by the posterior mean of each α_j and its root-mean-square error about the mean. This model was first introduced by Salas *et al.* (2004a).

To investigate the discrete nature of each contributing group considering any effect population migrations or the trans-Saharan slave trade may have had on the clustering of populations, an exact test of population differentiation (a test

described in Section 2.4) was performed on haplogroup profile frequencies between each African coast in Arlequin 3.5.

All haplotypes not found in Sub-Saharan Africa were excluded from the Jamaican sample for the intra-continental admixture analyses to focus the analysis only on African groups and remain consistent with the historical embarkation data from the continent. Additional analyses were performed excluding more marginal groups with debatable influence on the slave trade, including the pygmy populations of equatorial Africa and populations from the Sahel, to account for the infrequent movement of slaves from further inland to the coast (Nunn, 2008). The Sahel here is defined geographically as 11.25°N-18.75°N, 16.875°W-35.625°E (Rowell, 2001).

3.2.7 Stable Population Simulation Model Description

A stable population simulation model was created to investigate the difference between the expected mtDNA haplogroup distribution given historically recorded demographic parameters and the observed mtDNA haplogroup distribution in the sampled Jamaicans. Stable population models have birth and mortality rates that remain constant through time while assuming the population age structure does not fluctuate. These models are used when adequate demographic data are not available (Coale & Demeny, 1967) and have been used previously to estimate population dynamics of slaves in the United States during the 19th century (Farley, 1965; Eblen, 1974). The numbers of Africans arriving in Jamaica annually from each of the eight previously described coasts of Africa were taken from the Trans-Atlantic slave trade database (Table 3.1). Only import data from 1600 to 1807 was considered for the model. The distribution of mtDNA haplogroups for each source region extracted from the literature detailed are presented in Section 7.3.

For each year and each coast an mtDNA haplogroup profile was drawn from a multinomial distribution with probabilities dictated by the coast's observed mtDNA haplogroup distribution. The number of draws was dictated by the number of recorded Africans arriving from that given coast during that year. At the end of each year, deaths were modelled by multiplying population size by a factor $d\ (<1)$, and births by a factor $b\ (>1)$. To account only for those births that reached reproductive age, the birth rate was then multiplied by the sum of the

probabilities of a slave reaching the age of 15 (q), a probability parameterised from age-specific death rates observed in slaves 0-14 years of age. The total simulated population size was calculated as:

Equation 3.1

$$\mathrm{N} = dbq \sum_{i\,=\,1600}^{1807} \sum_{j=1}^{8} Mult(imports_{i,j}, \pi_j)$$

where N is the population size of Jamaica, equivalent to the sum of the haplogroup distribution counts in 1807, $imports_{i,j}$ is the number of slaves imported from coast j in year i and π_j is the distribution of the 16 haplogroups for coast j. Haplogroup counts were rounded to the nearest whole number. The parameters d, b and q were informed by Higman's assessment of the slave population in Jamaica from 1808-1834, after the abolition of the slave trade, but before the abolition of slavery (Higman, 1976). Higman's data were available for the eight historic parishes of Jamaica, allowing for the calculation of means and standard deviations in the above population parameters. By sampling from these distributions, the model was stochastic (Table 3.2). The modelled population was sampled 390 times to match sampling the Jamaican individuals with Sub-Saharan maternal ancestry analysed in this chapter. Given its stochastic nature, the complete model was run 10,000 times to capture the variability in the outcomes.

Table 3.2 Stable model population parameter distributions. Parameters used in the stable population model defined in Equation 3.1 were summarised from Higman (1976). $N(\mu, \sigma^2)$ is a normal distribution with mean μ and variance σ^2

Parameter	Value	Description
d	$\frac{1000 - N(25.78, 1.381)}{1000}$	Mortality parameter
b	$\frac{1000 + N(22.94, 0.488)}{1000}$	Birth Parameter
q	$N(0.855, 0.042)$	Probability of surviving age 14

To control for multiple testing, differences between individual haplogroup frequencies in the observed and simulated data were deemed significant if the observed values were less than 0.3125th percentile or more than the 99.6875th

percentile of the simulated values. Mean haplogroup counts were then used as input in the intra-continental admixture algorithm described above in Section 3.2.6 to generate estimated intra-continental admixture coefficient for the simulated population. The same African source haplogroup frequencies used in the simulation were used in the intra-continental admixture calculations. For example, if the data excluding pygmy groups were used in creating the expected Jamaican data, the intra-continental admixture calculations were made using the source data excluding pygmies.

3.3 Results

3.3.1 Jamaican samples

The first hypervariable segment of mtDNA was successfully sequenced in 107 Jamaican athletes and 293 Jamaican controls. Sequence data were aligned and haplogrouped using the methodology described in Section 2.3. For this study, the athlete and control samples were considered together.

3.3.2 Jamaican haplogroups

Individual mitochondrial haplotypes were clustered into larger haplogroups and paragroups to facilitate meaningful comparisons between the Jamaican samples and the corresponding literature, presented in Section 1.7.1. This clustering process decreased the number of effective tests performed, thus reducing the likelihood of committing type I error. A consequence of this is a reduction precision, though the large number of unique haplotypes found in the Jamaican and African samples necessitated this level of clustering to eliminate tests comparing regions enriched for given haplotypes to regions where that haplotype may absent.

The mtDNA haplogroup distribution observed in the sample of Jamaicans is presented graphically in Figure 3.4. Most of the profiles observed in Jamaica were allocated to Sub-Saharan L haplogroups (97.5%), a result consistent with past studies showing very few non-African maternal lineages in Jamaica (McLean *et al.*, 2005). Two “other L” categories (“Other L0’1’2” and “Other L3”) were included to merge haplogroup sub-groupings occurring at low frequencies in the Jamaican

sample e.g. L2d and L3h. Combined, the non-Sub-Saharan "non-L/U6" paragroup, consisting of all haplogroups not commonly found in Sub-Saharan Africa, accounted for 2.5% of the genotyped Jamaicans. The North African haplogroup U6 was included separately given its arrival in the Sub-Saharan region as the result of intra-African migration prior to the Trans-Atlantic slave trade (Rando *et al.*, 1998).

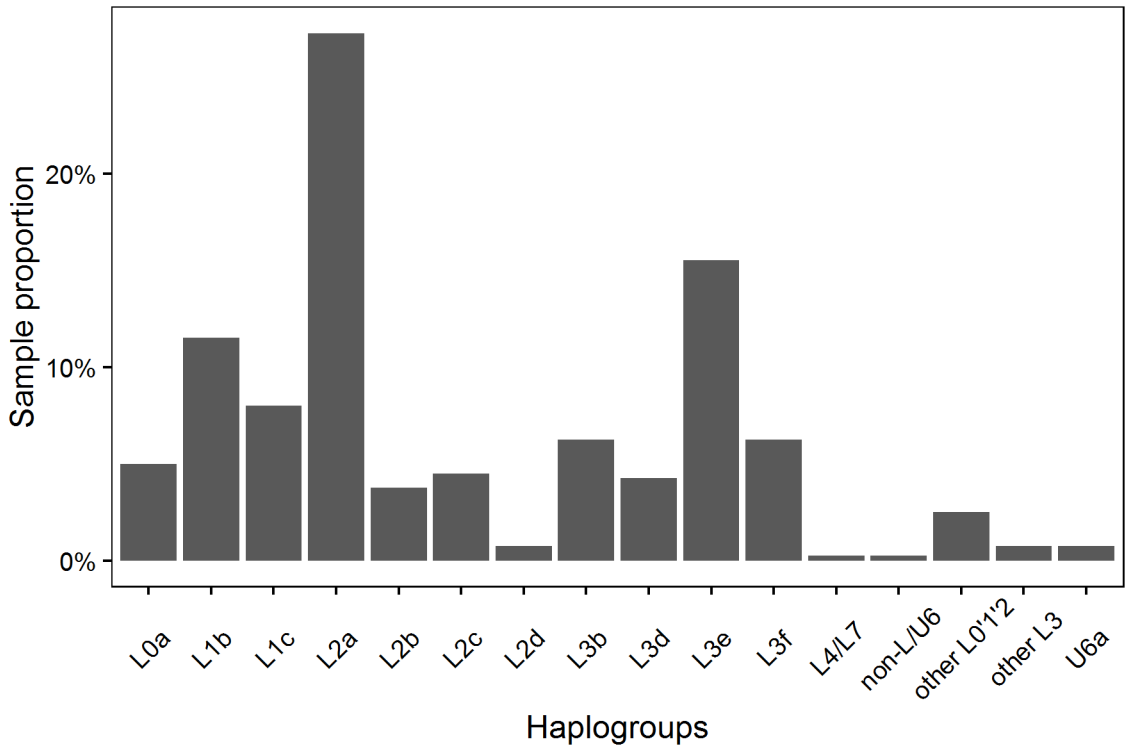


Figure 3.4 Proportion of mtDNA haplogroups or paragroups found in the 400 sampled Jamaicans.

While most Jamaicans had Sub-Saharan maternal lineages, ten Jamaicans were haplogrouped as non-Sub-Saharan. Their counts were as follows: Asian or New World haplogroups A2 (1), B2 (1) and D4 (2); Eurasian haplogroups H (1), J (2), U2 (1); and global macro-haplogroup M (2). Only one Jamaican profile could be unequivocally classified as belonging to typical Native American haplogroups A2. The sequences corresponding to haplogroups A2 and B2 were exact matches to sequences observed in El Salvador (Salas *et al.*, 2009), Costa Rica (Morera, 2002) and in the 'Hispanic' US sample of the United States' SWGDAM database (Monson *et al.*, 2002). There were several lineages most likely belonging to different sub-clades of the Pan-Asian haplogroup D4 and a member of haplogroup M30c1 of likely Indian origin (Maji *et al.*, 2009). Haplogroup U6 is typically from North Africa; however, the U6a profile observed in Jamaica (T16172C C16184T A16219G C16234T C16278T T16311C) has been reported in the Ikot Mbonde from Southeast Nigeria (Veeramah *et al.*, 2010) but not in North Africa. There are few haplotypes

of European ancestry in the Jamaican sample. The H2a2b1 profile (A16235G C16291T A16293G) was previously observed in Bolzano, Italy (Thomas *et al.*, 2008), in Galicia, Spain (Alvarez-Iglesias *et al.*, 2009) and in the United States (Behar *et al.*, 2007).

As shown in Section 7.4, most of the Jamaican haplotypes are present in populations sampled along the coastline of West Africa. Only eight different Jamaican haplotypes have frequencies above 2%. Of these, three of them (C16223T C16278T C16294T A16309G; T16126C C16187T T16189C C16223T C16264T C16270T C16278T A16293G T16311C and T16172C T16189C C16223T C16320T) were more frequent along the Gold Coast, Bight of Benin and Bight of Biafra than in any other African regions, while the remaining five (T16189C C16223T C16278T C16294T A16309G; T16126C C16187T T16189C C16223T C16264T C16270T C16278T T16311C; G16129A T16209C C16223T C16292T C16295T T16311C; T16124C C16223T and C16223T C16278T C16294T A16309G G16319A) are less geographically informative, being found in the entirety of Sub-Saharan Africa. The L2a1 profile C16223T C16278T C16294T A16309G G16319A was found nine times in Jamaica despite being absent in the samples from every African region. Typical East African haplogroups, such as L4/L7 and L5 profiles were virtually absent in Jamaica. There was a single individual belonging to L4b1, though this haplotype was found in very low frequency in the reference samples from the Bight of Biafra and the Gold Coast. This agrees with the historical documentation indicating that East Africa was not targeted by the European trading powers for the Trans-Atlantic slave trade.

3.3.3 Genetic diversity and demographic analyses

The results of the evaluation of genetic diversity and genetic composition of the parental populations and of the Jamaican mitochondrial gene pool are summarised in Table 3.3. All parental groups show expectedly high levels of diversity with regards to mean number of pairwise differences (π), as well as signs of expansion with regards to the stepwise expansion indices raggedness index (RI) and sum of squared deviations (SSD). All groups produced negative values for Tajima's D, suggesting that the observed variation is less than the expected variation for a given population size. This is consistent with either a selective sweep many

generations ago or population expansion and departure from mutation-drift equilibrium. Sampled Jamaicans fall within the range for the African diversity and demographic indices, providing no evidence of a founder effect.

Table 3.3 Genetic diversity and demographic statistics for the Jamaicans of African maternal descent and comparative population samples. Ten individuals with non-Sub-Saharan mtDNAs were excluded from these analyses.

Sample	N	k	S	π ± SD	D	RI	SSD
Jamaica	390	204	95	7.28 ± 3.42	-1.46*	0.0044	0.0009
Senegambia	892	361	118	6.93 ± 3.26	-1.60**	0.0037	0.0010
Sierra Leone	823	343	110	7.21 ± 3.38	-1.48*	0.0035	0.0016
Gold Coast	505	223	96	6.64 ± 3.14	-1.53*	0.0045	0.0007
Bight of Benin	421	196	97	6.75 ± 3.18	-1.58**	0.0048	0.0009
Bight of Biafra	3641	852	149	8.90 ± 4.10	-1.27*	0.0024	0.0014
West-central Africa	1403	370	122	9.05 ± 4.17	-1.15	0.0036	0.0004
South-east Africa	775	228	103	8.50 ± 3.94	-1.14	0.0038	0.0008
East Africa	805	402	139	8.76 ± 4.05	-1.55*	0.0043	0.0005

* p-value<0.05
** p-value<0.02

k = number of unique haplotypes
S = number of polymorphic sites
π ± SD = mean number of pairwise differences and standard deviation
D = Tajima's D
RI = raggedness index
SSD = sum of squared deviations

3.3.4 Intra-continental admixture analyses

An intra-continental admixture model using the observed mtDNA haplogroup distribution of sampled Jamaicans was fit using the observed mtDNA haplogroup distributions from the African coastal regions to estimate the proportion of maternal ancestry from each major slaving coastline. Four different combinations of presumptive and marginal African populations were used as a reference to investigate any influence the haplogroup frequencies in African populations unlikely to have contributed to the Trans-Atlantic slave trade may have had on the regional haplogroup profiles.

To accurately attribute the origins of each African coastal region, it was vital the mtDNA haplogroup distributions of each coast were unique across all African coastal regions. To test the independence of each coast, pairwise exact tests of population differentiation were performed on the haplogroup distributions from each coast in each combination of presumptive and marginal groups. The exact *P*-

values for every pairwise comparison were less than 1×10^{-4}, suggesting that each African coast had a distinct distribution of mtDNA haplogroups.

Estimated intra-continental admixture coefficients are summarised in Table 3.4, while haplogroup distributions for each reference dataset of African populations mentioned above are presented in Section 7.3. When marginal populations were not excluded from the reference dataset, the largest intra-continental admixture coefficient was associated with the Gold Coast (0.477 ± 0.12), while there were relatively small intra-continental admixture coefficients associated with the Bight of Biafra and West-central Africa (0.064 ± 0.05 and 0.089 ± 0.05, respectively). When excluding the Pygmy groups, the contribution from the Bight of Biafra and West-central Africa rose to their highest levels (0.095 ± 0.08 and 0.109 ± 0.06, respectively). The contributions of populations from the Bight of Benin were estimated to be much higher when populations from the Sahel were excluded from the reference datasets (0.123 ± 0.10 and 0.105 ± 0.09, compared with 0.261 ± 0.15 and 0.214 ± 0.14). The results from each of the four intra-continental admixture analyses suggested that East African groups had a negligible contribution to the population of modern Jamaica.

3.3.5 Stable population simulation results

A comparison between the distribution of simulated haplogroup frequencies from the stable population model and the observed Jamaican haplogroup frequencies are presented in Figure 3.5. The final simulated mean population size of Jamaica across all model runs (535,744, 95% CI [535,025, 535,786]) was larger than the observed population in 1808 when the slave trade was abolished (354,000; Higman, 1976). The observed pattern of simulated haplogroup frequencies does not fluctuate greatly when the differing African source datasets were used. Nevertheless, the simulation suggests that the observed Jamaican population consistently contained an excess of haplogroups L2a and paragroup “Other L0’1’2” and a dearth of haplogroup L1c. The significant difference between the simulated non-L/U6 paragroup frequency and the present frequency is an artefact of removing these haplotypes from the Jamaican sample before running the simulation. The remaining simulated haplogroup frequencies fall within the 95% confidence range.

Table 3.4 Intra-continental admixture coefficients and root mean square errors for parental populations calculated using observed and the means of the stochastically simulated haplogroup distributions from the stable population models. The source data used to create the simulated haplogroup data was identical to that used to estimate intra-continental admixture coefficients. Jamaica: *N* = 390.

Coastal Region	**All**			**excluding Pygmies**			**excluding Sahelian**			**excluding Pygmies & Sahelian**		
	N	observed	simulated	N	observed	simulated	N	observed	simulated	N	observed	simulated
Senegambia	892	0.049 ± 0.04	0.052 ± 0.04	892	0.048 ± 0.04	0.054 ± 0.04	39	0.075 ± 0.05	0.042 ± 0.03	39	0.072 ± 0.04	0.042 ± 0.03
Sierra Leone	823	0.096 ± 0.07	0.096 ± 0.07	823	0.092 ± 0.07	0.093 ± 0.07	659	0.092 ± 0.07	0.104 ± 0.08	659	0.092 ± 0.07	0.102 ± 0.08
Gold Coast	505	0.477 ± 0.12	0.220 ± 0.11	505	0.456 ± 0.11	0.222 ± 0.11	491	0.336 ± 0.15	0.208 ± 0.11	491	0.343 ± 0.14	0.206 ± 0.11
Bight of Benin	421	0.123 ± 0.10	0.127 ± 0.10	421	0.105 ± 0.09	0.154 ± 0.12	297	0.261 ± 0.15	0.128 ± 0.10	297	0.214 ± 0.14	0.163 ± 0.12
Bight of Biafra	3641	0.064 ± 0.05	0.260 ± 0.15	3097	0.095 ± 0.08	0.245 ± 0.14	3008	0.060 ± 0.05	0.289 ± 0.14	2464	0.091 ± 0.07	0.271 ± 0.15
West-central	1403	0.089 ± 0.05	0.210 ± 0.10	1314	0.109 ± 0.06	0.193 ± 0.08	1403	0.081 ± 0.05	0.198 ± 0.11	1314	0.097 ± 0.06	0.184 ± 0.08
Southeast Africa	775	0.092 ± 0.03	0.020 ± 0.02	775	0.085 ± 0.03	0.022 ± 0.02	775	0.088 ± 0.03	0.020 ± 0.02	775	0.083 ± 0.03	0.019 ± 0.02
East Africa	805	0.010 ± 0.01	0.014 ± 0.01	753	0.009 ± 0.009	0.017 ± 0.01	805	0.009 ± 0.01	0.011 ± 0.01	753	0.008 ± 0.01	0.013 ± 0.01

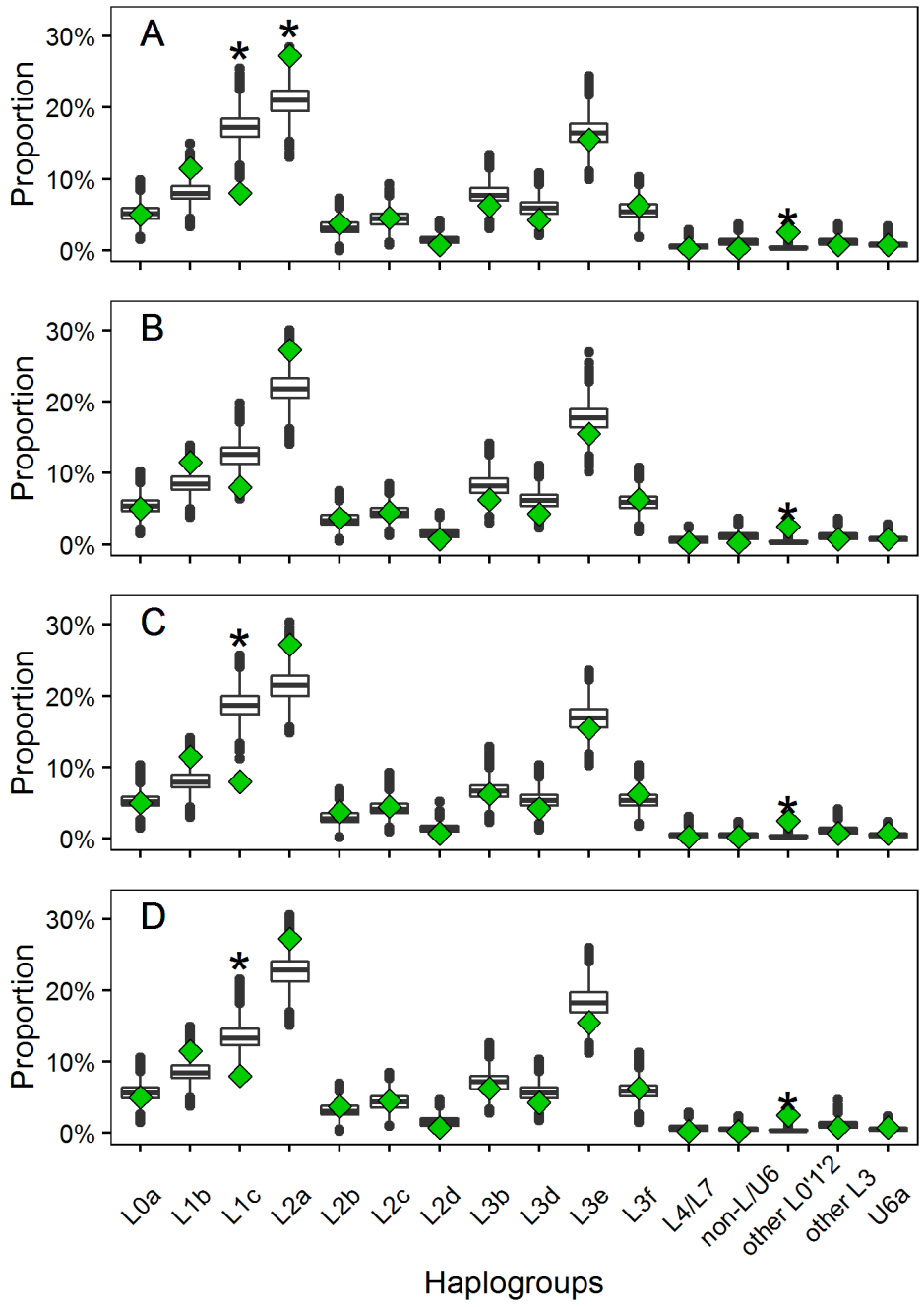


Figure 3.5 Boxplots of simulated haplogroup frequencies for the entire dataset (A), the dataset excluding pygmy groups (B), the dataset excluding groups from the Sahel (C), and excluding both marginal groups (D). The observed haplogroup frequencies are represented as green diamonds with significant differences represented with an asterisk.

The estimated intra-continental admixture coefficients for the simulated data are also presented in Table 3.4. The simulated haplogroup distribution led to a more evenly distributed set of intra-continental admixture coefficients across all source population datasets than did the observed haplogroup distribution. While the Gold Coast is still prominent (0.220 ± 0.11) when the complete African reference data are used, the intra-continental admixture coefficients of the Bight of Biafra (0.260

± 0.15) and West-central Africa (0.210 ± 0.10) were much larger than the coefficients calculated with the observed data (0.064 ± 0.05 and 0.089 ± 0.05, respectively). The contribution to the Bight of Biafra obtained its largest value when populations from the Sahel were excluded (0.289 ± 0.14, compared with 0.060 ± 0.05 from the observed dataset) while the West-central African coefficient obtained its largest value when marginal groups were included in the analysis (0.210 ± 0.10, compared with 0.089 ± 0.05 from the observed dataset). These results agree with the earlier hypothesis that the large influx of slaves from these regions near the end of the slave trade would make up most the Jamaican matrilineal pool. This allowed the simulated haplogroup frequencies to act as a suitable null model to assess any demographic pressures on the slave population by comparing those frequencies to the observed Jamaican haplogroup frequencies.

3.4 Discussion

The main findings of the investigations reported in this chapter are:

a) The Jamaican mtDNA haplogroup profile is significantly different from the distribution predicted under a stable population model. Apart from the scenario excluding Pygmy reference populations, there was a significantly lower proportion of haplogroup L1c in the Jamaicans than was predicted.
b) The Jamaican mtDNA more closely resembles the haplogroup profile of the modern-day Gold Coast despite large influxes of slaves from the Bight of Biafra and West-Central Africa near the end of the slave trade.

The African diaspora in Jamaica is the result of nearly 150 years of the Trans-Atlantic slave trade. By considering historical African departure points, estimates of the maternal contribution of each parental population were compared with historical arrival records. The results of the intra-continental admixture analyses suggest the mtDNA haplogroup distribution of Jamaica more closely resembles that of aggregated populations from the modern-day Gold Coast region despite an increasing influx of individuals from both the Bight of Biafra and West-central Africa during the final years of the slave trade. When considering what is known about the negative rate of natural population growth of slaves on Jamaica, these

results add an additional layer to the demographic history of Jamaica. Planters found it more economical to import new labour rather than invest in conditions that would encourage natural reproduction within their existing already in Jamaica (Sheridan, 1961). Coupling low fecundity with high mortality led to the expectation of a fluid demographic shift through time to a mitochondrial haplogroup distribution more closely resembling those groups arriving later during the slave trade. These results did not show this, suggesting a combination of greater mortality for later arriving slaves, assortative mating and other selection pressures were important to the creation of modern Jamaican matrilineal composition.

An assessment of the sampled Jamaican mitochondrial genomes indicated that the majority fell into the Sub-Saharan African macro-haplogroup L (Figure 3.4), in line with previous findings (McLean *et al.*, 2003; McLean *et al.*, 2005; Torres *et al.*, 2007; Benn-Torres *et al.*, 2008). The minority of observed haplotypes originating from outside Sub-Saharan Africa may highlight the rapid assimilation of a small, albeit global, workforce arriving in a post-emancipation Jamaica during the subsequent labour shortage (Richardson, 1989).

Genetic diversity was then assessed to identify any idiosyncrasies that may be present in the Jamaican sequence data when compared to the historically informed African datasets (Table 3.3). Results suggest that Jamaica and all parental groups have high levels of diversity and that any evolutionary forces or potential sampling biases in Jamaica or Africa have not compromised the intra-continental admixture analyses.

The estimated intra-continental admixture coefficients (Table 3.4) suggest that the sample of modern Jamaicans more closely resemble ethnic groups found in the Gold Coast than either the Bight of Biafra or West-central Africa. The exclusion of pygmy groups only negligibly increased the intra-continental admixture coefficient from the Bight of Biafra and West-central Africa. This increase was expected considering the pygmy groups' near-fixation of haplogroup L1c, a result of a population bottleneck in response to their marginalisation as a result of the Bantu Expansion (Batini *et al.*, 2011). Excluding populations from the Sahel reduced the intra-continental admixture coefficient for the Gold Coast while increasing that for the Bight of Benin. This was expected given the

disproportionate number of samples from the Sahel included from the Bight of Benin and consistent with the belief that most slaves originated from ethnic groups located near the coast (Nunn, 2008). Removing both pygmy and Sahelian groups did not change the intra-continental admixture coefficients markedly from the iteration removing Sahelian groups alone. The most striking result, however, was the low intra-continental admixture coefficients of both the Bight of Biafra and West-central Africa regardless of whether marginal groups were included in the estimations or not.

This is the first time a simulation model parameterised using recorded demographic statistics has been used to estimate predicted mtDNA frequencies. These predictions were then compared with observed frequencies to provide greater insight into the region-level origins of a population. This model was formulated according to slave importation data, the observed mtDNA haplogroup distributions from contributing coasts of Africa as well as historically informed birth and death rates. When all populations are included, two haplogroups and one paragroup were expected to have significantly different frequencies per the simulation model when compared to the observed values across the four source datasets used (Figure 3.5). The increased model estimates of both the Bight of Biafra and West-central Africa was likely due to the large observed frequencies of haplogroup L1c, a haplogroup found to be at high frequency in the mtDNA data from Bight of Biafra and West-central Africa (0.24 and 0.33, respectively). The sequences of Central African pygmies used in these analyses are predominantly haplogroup L1c (Quintana-Murci *et al.*, 2008), and their inclusion may have skewed the simulated frequencies away from a true representation of the expected haplogroup distribution.

While the intra-continental admixture analyses were conducted using haplogroup-level lineage data, it is also possible to calculate intra-continental admixture estimates using haplotype-level lineage data by searching for exact and fuzzy matches where either one or two markers in the motif differ. This methodology was introduced in Mendizabal *et al.* (2008) for Cubans and been used to infer the matrilineal origins of populations in El Salvador (Salas *et al.,* 2009), Mexico (Sandoval *et al.,* 2009), Nicaragua (Nuñez *et al.,* 2010) and Venezuela (Gómez-Carballa *et al.,* 2012*)*. This methodology offers much greater accuracy when estimating the origin of a given haplotype, although intrinsically it is much more

sensitive to the sampled haplotypes from the putative origins. In contrast, the haplogroup-level methodology used in this chapter theoretically captures all possible haplotypes within each haplogroup by comparing the frequencies of the haplogroups.

The haplotype-level method described above has been used to estimate the intra-continental origins of modern Jamaicans using the same cohort described in this chapter (Deason *et al.*, 2012), although the analyses were performed as part of a collaboration with Antonio Salas from the University of Santiago de Compostela. When only exact haplotype matches were considered, the result of these analyses showed that while the Gold Coast was still the most likely origin (0.20), the signal was markedly reduced when compared to the haplogroup-level analysis presented in this chapter (0.48). The Bight of Biafra and West-central Africa remain underrepresented from what was expected, though are considerably greater than the haplogroup method estimated (0.15, 0.11, respectively).

The intra-continental admixture estimates may suggest a preference among Jamaican planters to purchase slaves from certain regions over others. Historic evidence suggests the Jamaican planting class held the Akan of the Gold Coast in very high regards (Curtin, 1969), although some have suggested that planters were powerless in choosing specific ethnic groupings because of the heterogeneity of ethnic groups present on each arriving ship and the immediacy of labour needs (Burnard & Morgan, 2001).

As introduced in Section 1.5.3, the acclimatisation process for newly-arrived Africans was both mentally and physiologically stressful. Between a quarter and half of Africans died within their first three years on the island (Sheridan, 1985). The distance and time individuals spent travelling was negatively correlated with survival (Haines *et al.*, 2001) and as such, individuals embarking at African ports further from Jamaica would often arrive in a poorer state. It can be postulated therefore that despite more than half of all Africans shipped to Jamaica coming from the Bight of Biafra, they may have not survived the acclimatisation process as frequently as those individuals from further west along the coast. Longer journeys may explain the intra-continental admixture estimates from the Bight of Biafra falling below those hypothesised. Slaves arriving from Southeast Africa and

Madagascar were significantly disadvantaged in this respect, perhaps reflected in their negligible estimated contribution to the mtDNA pool of Jamaica (Table 3.4).

The development of modern Jamaican English may also provide insight into the demographic constitution of the island. The modern creolised English spoken on the island has been traced to Northern British and Irish overseers and bookmakers and the early African slaves they interacted with. During the initial era of slavery on the island (1655-1700), slave acculturation was a process characterised by direct contact between newly-arrived Africans and their European overseers. Though the Gold Coast contributed somewhat to the slave trade prior to 1700, the Akan-speaking groups from modern Ghana were thought to be the linguistic group with the greatest number of speakers (Patterson, 1966). These early slaves heavily influenced the development of the creole slave language and culture on the island (Cassidy & Le Page, 2002). Additionally, modern Jamaican English contains many loanwords of African origin, and many of those are etymologically from the Gold Coast region (Cassidy, 1966). A large part of the pidginisation was thought to have been completed within the first few decades, and as the proportion of Europeans began to shrink with the explosive increase in slave imports, newly-arrived Africans would be more reliant on established slaves for the acquisition of a common language (Mintz & Price, 1976). Africans arriving from the Gold Coast may have thus found the acclimatisation and acculturation process less stressful because of cultural and linguistic commonalities, although any speculation of greater fecundity and survivorship is difficult to defend solely on those grounds given the linguistic and ethnic variability that would have been present along the Gold Coast at the time (Law, 1991).

The historic evidence suggests that slave society on Jamaica operated in a very rigid social hierarchy. Jamaican-born slaves had much greater life expectancy, fecundity and upward social mobility than those born in Africa (Higman, 1976). The entire society was also highly endogamous, the one obvious exception being the high frequency at which European men took concubines and fathered children with their slaves (Burnard, 2006). This ultimately provided an opportunity for intergenerational mobility as slaves born of mixed parentage were often the recipients of more favourable positions (e.g. domestics and tradesmen) and were also more likely to be granted their freedom by their owners (Higman, 1976). Considering the paternal contribution by Europeans to modern Jamaicans has been

estimated at just over 40% (Torres *et al.*, 2007), African-European admixture may have played an important role in the genetic legacy of the slave population. It is conceivable then that the large influx of slaves from the Gold Coast during the formative years of slavery on Jamaica may have led to the establishment of a large population of mixed parentage slaves with ancestry from the Gold Coast. These slaves would then have been placed in more favourable positions, increasing their fertility when compared to newly arrived African slaves and ultimately perpetuating the ancestry of the Gold Coast.

Several assumptions were made in the formulation of the simulation model. Firstly, the model assumes no genetic contribution from non-African matrilines as the result of intra-continental admixture. The observed haplotypes in Jamaica indicated that roughly 2.5% of the population has mtDNAs not originating in Sub-Saharan Africa. However, to simplify the model, these intra-continental admixture events were not parameterised. Secondly, the age structure of arriving Africans and the Jamaican population was not considered. Mortality varied greatly by age, sex and birth location. To facilitate the modelling, however, a single death and adjusted birth rate were used. Failing to consider population demography may have led to an overestimation of the actual birth rate or underestimation in the actual death rate. This may explain why the observed slave population in 1808 is below the 95% confidence interval of the simulated population size in 1808. Thirdly, although the death and birth parameters were inferred from recorded data, the parameters were informed from the period after the slave trade ceased, but before slavery was abolished. Birth and death rates are likely to have varied greatly through time as technology improved and the island's economy diversified. As the influx of Africans had been stopped, it would have been in the planters' interest to impose amelioratory practices to encourage natural population growth and reduce mortality, although the continued negative natural growth of the slave population is evidence to suggest that few planters chose to do this (Higman, 1976; Morgan, 2006). Lastly, the observed distribution of mtDNA haplogroups in present day Jamaica is assumed not to have changed significantly since the end of the slave trade. Any movement of people to or from Jamaica occurring since the end of the slave era was assumed to have been statistical noise and inconsequential to the population's mtDNA haplogroup distribution, although the amount of actual change to the matrilineal gene pool since 1834 would be very difficult to measure.

More broadly, the use of imputed estimates for the slave trade network may have introduced a bias in the representation of certain regions. Complete voyage data were used to impute incomplete voyages and it is possible that voyages with incomplete records were fundamentally different from more successfully recorded ones. The mortality statistics from each coast had very large standard deviations, with the likely number of slaves arriving in the Americas falling within a broad range of potential values (Eltis, 2010). The imputed locations for embarkation and disembarkation were likely more accurate. Ships leaving Europe with goods for a particular market were expected at a particular port in Africa; these same ships were then expected to deliver slaves to a particular port in America (Eltis, 2010). It is impossible to know if there is a systematic bias in these data however, given the length of time since the slave trade took place.

In summary, despite the historical evidence that an overwhelming majority of slaves arrived in Jamaica from the Bight of Biafra and West-central Africa near the end of the slave trade, the mtDNA haplogroup profile of modern Jamaicans shows a greater affinity with groups found in the present-day Gold Coast region. Caution must be paid however to the scope of the analyses performed here. The Jamaican slave markets were the largest in the West Indies and anecdotal accounts exist of slaves being purchased in Jamaica for plantations in other parts of the New World (O'Malley, 2009), though it is difficult to accurately trace the ancestry of the resold slaves. Additionally, the mtDNA distributions of both Jamaica and the African source populations were assumed to have stayed relatively constant since emancipation. However, the environmental and selection pressures acting on these populations will certainly have changed through time, potentially influencing the observed genetic composition of both modern Jamaicans and modern Africans. Finally, the island was treated in these analyses as a closed system with regards to immigration and emigration after the abolition of slavery in 1834. The end of slavery in Jamaica brought about a change in economic climate, with many Jamaicans emigrating to other parts of the world, as well as foreign migrant labourers arriving from around the globe. Whether any of these effects have significantly affected the mtDNA distribution on the island is difficult to determine.

4

Do the mtDNA and NRY Haplogroup Profiles of African-American and Jamaican Elite Sprint Athletes Differ Significantly from Matched Controls?

4.1 Introduction

In the previous chapter, historic shipping records from the Trans-Atlantic slave trade and mitochondrial sequence data from Jamaica and Africa were contrasted to understand the likely origins of modern Jamaicans better. The results showed that maternal lineages currently associated with the Gold Coast were significantly more prevalent in Jamaica than would be expected based on the importation records and a simple model of population dynamics. The results also suggested that the African regions disproportionately represented in the importation records at the end of the Trans-Atlantic slave trade contributed significantly less to the modern Jamaican population than expected. These results provide evidence that there may have been either assortative mating among the slave population or reduced fitness of slaves arriving later.

Elite sprint running is a multifaceted phenotype associated with numerous environmental and genetic factors (Tucker & Collins, 2012). When considering that elite athlete status has an estimated heritability of 66% (de Moor *et al.*, 2007), the genetics of the elite sprint phenotype has the potential to contribute greatly to the understanding of the limits of human performance. Unfortunately, there has been limited success in identifying generalisable genetic markers associated with elite sprint athlete status (Ahmetov & Fedotovskaya, 2015).

Results from previous research investigating any associations between uniparental lineages and elite athlete status have been mixed. Statistically significant differences between athletes and controls have been found in maternal lineages

from samples from Japan (Mikami *et al.*, 2010), Spain (Castro *et al.*, 2007) and Kenya (Scott *et al.*, 2009). Statistically significant differences between athletes and controls have also been found in paternal lineages from Ethiopia (Moran *et al.* 2007). These results, however, are not generalisable because each population has experienced unique selection pressures and migration patterns leading to markedly different lineage frequencies around the world (Wallace, 2015).

While the mitochondrial genome contains the genes critical to encode the 13 polypeptides essential for oxidative phosphorylation (Mishra & Chan, 2014), there is no evidence that the non-recombining region of the Y chromosome contains genes associated with performance-related traits (Moran *et al.*, 2004). Elite athlete status, as well as physical attributes associated with elite sporting performance, including endurance and muscle strength, have been shown to be heritable (de Moor *et al.*, 2007; Costa *et al.*, 2012). Therefore, by comparing any differences in maternal and paternal lineages between athletes and controls, it is possible to test for homophily, the tendency for individuals to pair based on shared attributes. If homophily were acting in these populations, one would expect an enrichment over many generations for certain lineages containing the phenotypes associated with elite athlete status. If there is evidence of homophily, it may be possible to link functional variants to individual lineages to better inform future research into the genetics of elite sprint running.

The aim of this chapter is to test if elite sprint athletes from the African-American community of the United States and Jamaica form biologically distinct groups from their matched controls. Additionally, this chapter seeks to determine whether certain lineages are over- or under-represented in these athletes when compared with matched controls. Based on the significant differences between elite athletes and controls in other populations around the world, it was initially hypothesised that the distributions of the maternal and paternal lineages of African-American and Jamaican athletes would be significantly different from ethnically-matched controls. However, the small sample sizes in the mtDNA analysis in Jamaicans and the NRY analyses in both populations (Table 4.1) indicates that an unfeasibly large effect would be necessary to detect any significant differences. It was then hypothesised that the maternal lineages of African-American athletes and controls would be significantly different, while the remaining comparisons would not be, due to limited statistical power.

4.2 Methods

4.2.1 Study samples

Inclusion criteria for elite sprint athletes are found in Section 2. Only men have a Y chromosome, and as a result, the NRY analyses had fewer samples. The number of individuals analysed in this chapter are summarised in Table 4.1.

Table 4.1 African-American and Jamaican sample sizes in cases and controls by genetic marker.

	African-American		Jamaican	
	mtDNA	NRY	mtDNA	NRY
Athletes	105	50	107	44
Controls	1642*	246†	293	47

* Monson *et al.* (2002); Diegoli *et al.* (2009); Lao *et al.* (2010)

† Lao *et al.* (2010)

4.2.2 Mitochondrial DNA genotype and haplogroup data generation

One hundred and five African-American elite sprint athletes and 1,642 African-American control sequences from the literature (Monson *et al.*, 2002; Allard *et al.*, 2005; Diegoli *et al.*, 2009; Lao *et al.*, 2010) as well as 107 Jamaican athletes and 293 Jamaican controls were sequenced to evaluate the maternal lineages present in each cohort. The Jamaican samples are identical to those used in Chapter 3. Mitochondrial sequence data were generated, aligned and haplogrouped using methodology described in Section 2.3.

4.2.3 NRY genotyping methodology

The Y chromosome in 50 African-American athletes, 44 Jamaican athletes and 47 Jamaican controls were genotyped to evaluate the paternal lineages present in each sample. Publicly available Y chromosome data from 246 African-American controls were also used (Lao *et al.*, 2010). Previously identified SNPs maximising continental differentiation genotyped by Lao *et al.* (2010) were chosen to determine haplogroup assignment with the NRY. Twenty-four unique event polymorphisms (UEP) defining the major global branches of the Y chromosomal tree were genotyped in all male samples in two separate multiplex PCR reactions (Karafet *et al.*, 2008; de Filippo *et al.*, 2011). The UEPs (12f2, M106, M124, M145, M168, M170, M172, M174, M175, M20, M201, M207, M213, M214, M269, M45, M52, M69, M9, M91, M96, MEH2, SRY10831 and Tat) were then sequenced using two

multiplex SNaPshot® minisequencing assays of 12 SNPs each (Applied Biosystems, Foster City, CA, USA).

The SNaPshot® methodology allows for high throughput of many nonadjacent markers in a single reaction first by amplifying the markers, then annealing single-base extension primers before sequencing the product using capillary electrophoresis. An additional fluorescence detection step colours alleles according to the detected nucleotide. Where primer dimer formation made accurate base calls difficult or ambiguous, additional smaller individual-level multiplexes were performed by adjusting the concentrations of the extension primer. SNaPshot® genotyping protocol, primer sequences and concentrations were taken from the literature (Corach et al., 2010) and aligned to the NRY reference sequence (Reference Sequence NC_000024.9) using the online primer design tool Primer-BLAST (Ye et al., 2012). This was done as a precaution before primers were purchased to ensure the description of the primer sequences were free from obvious typographical errors.

Before beginning the SNaPshot® reaction, each individual primer pair was first validated in a uniplex PCR containing 1 ng of template DNA, 0.4 µM of both forward and reverse primers (Table 4.2, Applied Biosystems) and 0.6 units of Platinum Multiplex PCR Master Mix® (Applied Biosystems), a PCR master mix optimised for multiplexes. In the final multiplex PCR, 0.5 ng of template DNA was amplified in a 12.5 µl reaction volume containing 1× PCR buffer, 6.5 µM total $MgCl_2$, 200 µM of deoxynucleotides (dNTPs) and 2.5 units of AmpliTaq Gold® DNA polymerase. All initial PCRs were performed with an initial denaturation at 94ºC for 10 minutes followed by 35 cycles of 30 seconds at 94ºC, 30 seconds at 60ºC, 30 seconds at 72ºC and a final extension for 5 minutes at 72ºC. To eliminate excess primers and dNTPs, 2 µl ExoSAP-IT® (USB, Affymetrix, Cleveland, USA) was added to the PCR product and incubated at 37ºC for 30 minutes, followed by a final enzyme inactivation at 80ºC for 15 minutes.

Single-base extension (SBE) primers were then used to modify the molecular weight of the amplified DNA to differentiate each SNP in the multiplex PCR product during capillary electrophoresis. The weight of each SBE primer was controlled by designing the primer with a predefined number of redundant base motifs. Once the original PCR product had annealed with the set of SBE primers,

the different molecular weights made it possible to genotype up to twelve SNPs in a single reaction. Sequences for these SBE primers were also taken from Corach *et al.* (2010) and are presented in Table 4.2. Extension reactions were performed in a 5 µl reaction volume using 1 µl of the initial purified PCR product described above, 2.5 µl of SNaPshot® multiplex Ready Reaction Mix (Applied Biosystems) and 0.4 µM primer. Samples were set for an initial denaturation at 96ºC for 2 minutes, followed by 25 cycles of 10 seconds at 96ºC, 5 seconds at 50ºC and 30 seconds at 60ºC. To eliminate unincorporated ddNTPs, 1.25 µl SAP®-reagent (USB) was added and incubated at 37ºC for 1 hour. SAP was inactivated by incubation at 75ºC for 15 minutes. Two µl of the SAP-treated extension product was analysed with an ABI3100 Genetic Analyser using a 36-cm capillary array, polymer POP4 and Genescan 120 LIZ as internal size standard. Data were analysed using GENEMAPPER ID v3.2.1 software (Applied Biosystems). After background noise subtraction and colour separation, peaks were sorted into bins according to size by comparison to the internal size standard. After colours were assigned based on nucleotide to each SNP-specific peak in GENEMAPPER, the SNP calls were collated by individual. NRY haplogroups were then derived from these genotyped SNPs at the bifurcating internal nodes of a simplified phylogenetic tree (Figure 4.1) using the marker phylogeny described by Karafet *et al.* (2008).

Table 4.2 Forward, reverse and SBE primer sequences used in the SNaPshot® reaction to generate NRY data for each UEP. Len corresponds to the length of the SBE primer, while Dir corresponds to its direction.

UEP	PCR forward primer seq. (5-3')	PCR reverse primer seq. (5-3')	SBE primer sequence (5-3')	Len.	Dir.	*
12f2	CACTGACTGATCAAAATGCTTACAGAT	GGATCCCTTCCTTACACCTTATACA	GTGCCACGTCGTGAAAGTCTGACAAAACATGTAAGTCTTTAATCCATCTC	50	F	A/–†
M106	TGTACTTGGACAGGTGAAGCA	TCGCTTTTCCACCTACTCCT	GTCGTGAAAGTCTGACAATAGTTCCCTATGACAGTATC	38	F	A/G
M145	CCTCCCACTCCTTTTTGGAT	GCATACTTGCCTCCACGACT	GACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAATAGACACCAGAAAGAAAGGC	56	F	G/A
M172	TGAGCCCTCTCCATCAGAAG	GCCAGGTACAGAGAAAGTTTGG	CCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAACAAACCCATTTTGATGCTT	61	F	T/G
M174	TCTCCGTCACAGCAAAAATG	GACCCATCTTGCAAGGAAAA	CGTCGTGAAAGTCTGACAACCTTCTGGAGTGCCC	34	F	T/C
M175	GATTTAAACTCTCTGAATCAGGCACAT	TTCTACTGATACCTTTGTTTCTGTTCATTC	ACGTCGTGAAAGTCTGACAACACATGCCTTCTCACTTCTC	40	F	T/A
M207	GGGGCAAATGTAAGTCAAGC	TGTTCGCTGCTACGAATCTTT	GTCTGACAAAAGTCAAGCAAGAAATTTA	28	F	A/G
M69	TGGGTAGCCTGTTCAAATCC	TTCCCTTTGTCTTGCTGAAA	TGCCACGTCGTGAAAGTCTGACAAGGCTGTTTACACTCCTGAAA	44	F	T/C
M9	AGGACCCTGAAATACAGAACTG	AAATATTTCAACATTTCACAAAGGAA	ACTGCAAAGAAACGGCCTAAGATGGTTGAAT	31	F	C/G
M91	CAAAAATCCCCCTACATTGC	GCAGTGCCCTTCCAAATAAA	CCCCCCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAATTGCTATTCTGTTTTTTTT	65	F	T/A
M96	GCCAGCCAAGAATGAAGAGA	TGAGCTGTGATGTGTAACTTGG	GGAAAACAGGTCTCTCATAATA	22	R	G/C
Tat	GACTCTGAGTGTAGACTTGTGA	GAAGGTGCCGTAAAAGTGTGAA	CGTCGTGAAAGTCTGACAACTTCTCTCTTGCTGTGCTCTGAAATATTAAATTAAAACAAC	60	R	A/G
M124	TCAAAGTCACAGTATCTGAACTAGCA	TCATATTGAGATTTTGCTTTCCT	CCCCCCCCCCCCCCAGGTGCCACGTCGTGAAAGTCTGACAAGGGGAACAGGGAAGT	55	R	C/T
M168	TGTTTTGCAGAGACTTGGA	CTGCCCCTCTATCAGACCAT	GGTGCCACGTCGTGAAAGTCTGACAAGTTTTAATTCTTCAGCTAGC	46	R	G/A
M170	CAGCTCTTATTAAGTTATGTTTTCATATTCTGTG	GTCCTCATTTTACAGTGAGACACAAC	CAACCCACACTGAAAAAAA	19	R	T/G
M20	AGTTGGCCCTTTGTGTCTGT	CATGTTCAGTGCAAATGCAAC	CGTGAAAGTCTGACAACACATTTGTAGGTTCAACCAACTGTGGATTGAAAAT	52	F	A/G
M201	GATCTAATAATCCAGTATCAACTGAGG	CCAGCATCCTATCAGCTTCA	CAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAACTAAGTACCTATTACGAAAA	61	R	C/A
M213	CCATATAAAAACGCAGCATTCTGTT	TGGAGAGAACTTGAGAAAAAGTAGAGAA	TGACAATCAGAACTTAAAACATCTCGTTAC	30	R	A/G
M214	CCATGGTCCAATTGTACAGC	GAGGTCAAGGGTGTGGTGAG	CTAAACTAGGTGCCACGTCGTGAAAGTCTGACAAAGACACTGTCTGAAAACAAC	54	R	A/G
M269	AAGGGGAATGATCAGGGTTT	CCAAGGTGCTGGGATTACAC	TGCCACGTCGTGAAAGTCTGACAAGGAATGATCAGGGTTTGGTTAAT	47	F	T/C
M45	GAGAGAGGATATCAAAAATTGGCAGT	TGACAGTGGCACCAAAGGTC	AACAACTCAGAAGGAGCTTTTTGC	24	R	C/T
M52	CCTCAACTTCCCAGAGTGTTG	GACGAAGCAAACATTTCAAGAGAG	CGTCGTGAAAGTCTGACAAAATATCAAGAAACCTATCAAACATCC	45	R	T/G
MEH2	TTTGAGTAAGCCATCACCCC	TGCAAAAACTGCATTGATGA	CCCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAAATGTAATTTAAAGCATAGTG	63	F	GG/GT‡
SRY10831	TCATCCAGTCCTTAGCAACCATTA	CCACATAGGTGAACCTTGAAAATG	TCTGGCCTCTTGTATCTGACTTTTTCACACAGT	33	F	A/G

* Alleles displayed as Ancestral/Derived
† 12f2 variant appears is either present or absent
‡ Two nucleotide polymorphisms need to be genotype because a homologous region on the X chromosome is also amplified by the primers used

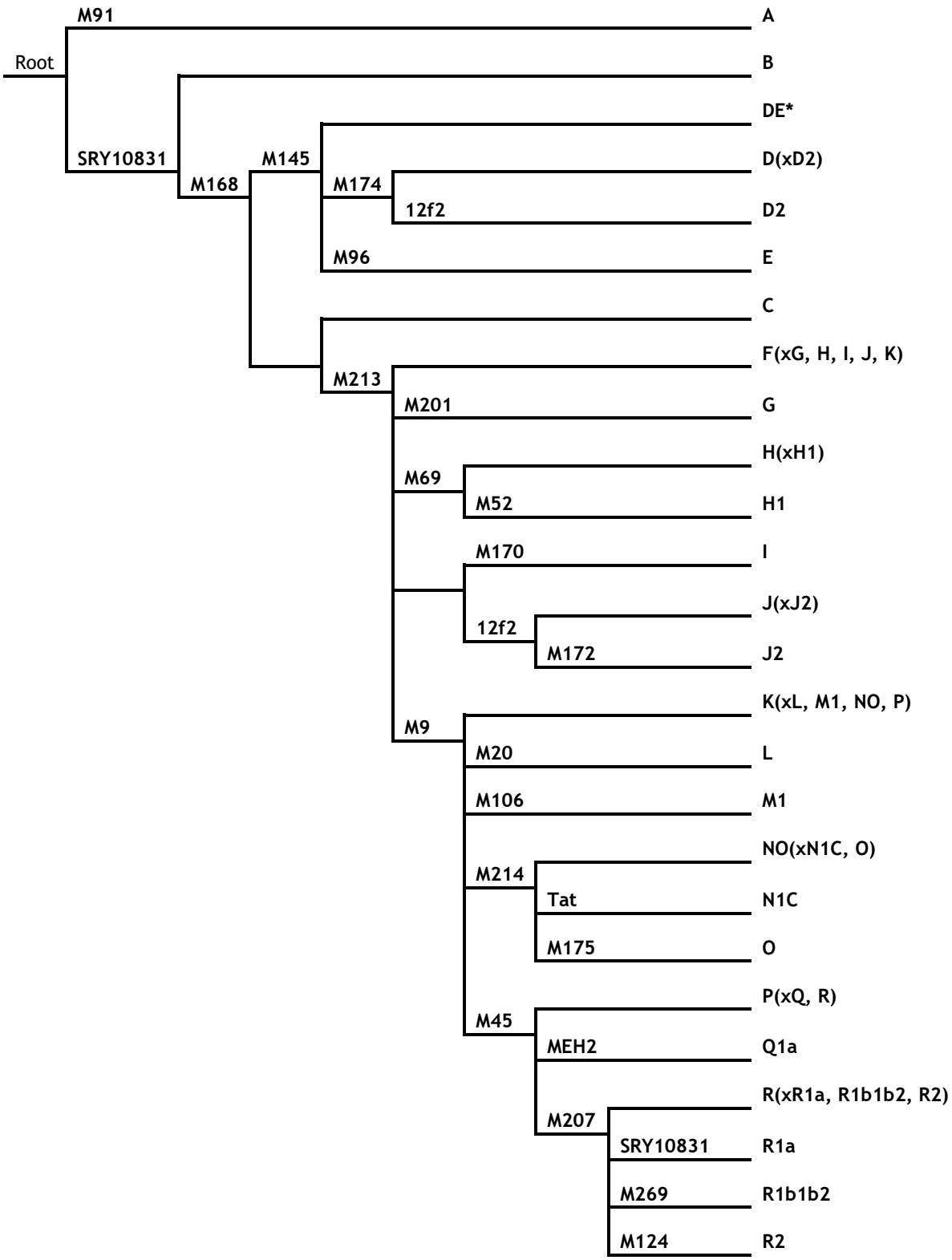


Figure 4.1 Simplified bifurcating phylogenetic tree of the NRY UEPs used in this study and their corresponding haplogroup.

4.2.4 Statistical analyses

Total cohort sequence diversity for the HVS-1 sequence data was determined by calculating the mean pairwise nucleotide differences between individuals within each population (π) assuming no recombination between sites and neutral evolution (Tajima, 1993). These values were calculated to assess the genetic variability in the mtDNAs in this study. Exact tests of sample differentiation based on total haplogroup frequencies (as described in Section 2.4) were performed to evaluate whether athletes form a distinctive group from controls separately in African-Americans and Jamaicans. Rejection of the null hypothesis then indicates that some haplogroup or groups are significantly more common in athletes than in controls. These calculations were performed in Arlequin 3.5 (Excoffier & Lischer, 2010).

To determine the strength of any association between athlete status and either mtDNA or NRY haplogroups, a series of logistic regressions were fit to these haplogroup count data. Each test took the form of a 2 × 2 contingency table, which meant the number of athletes and controls observed to have a haplogroup of interest were compared to the sum of the remaining haplogroups in athletes and controls. As the number of performed tests increases, the chance of type I error also increases. To help account for this, odds ratios (ORs) from the fitted logistic models were calculated with Bonferroni-adjusted confidence intervals. Assuming a critical α of 0.05, the new conservative significance threshold for the mtDNA regressions with 16 independent tests was adjusted to $\frac{0.05}{16} = 3.125 \times 10^{-3}$ while the NRY regressions with 3 independent tests was adjusted to $\frac{0.05}{3} = 1.667 \times 10^{-3}$.

Statistical power for the comparison of haplogroup distributions was calculated *a posteriori* with the online tool mitPower (Pardo-Seco *et al.*, 2013). Power was estimated using 2 × k contingency tables where k equals the number of non-empty haplogroups present in both athletes and controls for each population. Significance was accepted at $P < 0.05$ using a Fisher's exact test and was implemented over 1,000 Monte-Carlo simulations to reduce the potential computational time that may result from the increased dimensionality. Power values above 80% were deemed adequate, meaning that there would be an 80% chance of calculating a P-value of less than 0.05 if the estimated difference between the populations actually exists. Values below 80% signified that there was

not sufficient power to detect any difference in the current sample even if it were to exist in the population.

4.3 Results: mitochondrial DNA

4.3.1 Summary statistics: mtDNA

Total mtDNA sequence diversity was evaluated for the entire African-American and Jamaican cohorts, in addition to the athlete and control groups in both samples (Table 4.3). Results suggest similar amounts of diversity between athletes and controls in both groups (8.09 and 8.12 in African-Americans and 8.66 and 8.05 in Jamaicans).

Table 4.3 Total nucleotide sequence diversity (π) and standard deviation of composite African-American and Jamaican cohorts, as well as by athlete/control status.

African-American (combined)	8.38 ± 3.88
Athletes	8.09 ± 3.78
Controls	8.12 ± 3.80
Jamaican (combined)	9.81 ± 4.50
Athletes	8.66 ± 4.03
Controls	8.05 ± 3.75

Individual haplotypes were clustered into larger phylogenetically-informed haplogroups to facilitate a direct comparison between lineage distributions of athletes and controls (Figure 4.2). In analyses like those performed in the previous chapter, two "other L" paragroups ("Other L0′1′2" and "Other L3") were included for haplogroup subgroupings occurring at low frequencies in both African-Americans and Jamaicans, e.g. L2d and L3h. A "non-L/U6" paragroup was included for all haplogroups not commonly found in sub-Saharan Africa, including those with European or Asian origins. Combined, this non-sub-Saharan paragroup was observed in 17.1% of the African-American athletes compared with 7.7% of the African-American controls, and in 1.9% of the Jamaican athletes compared with 1.7% of the Jamaican controls. Apart from the African-American athletes, these values are in line with the frequencies expected.

4.3.2 Statistical power of mtDNA analyses

The estimates of statistical power suggest that sample sizes were large enough with sufficient variation to reject the null hypothesis that there are no differences between the haplogroup distributions of African-American athletes to controls (99.7%), African-American athletes to Jamaican athletes (99.6%) and African-American controls and Jamaican controls (99.8%). The comparison between Jamaican athletes and Jamaican controls, however, is only 58.4%.

4.3.3 Population differentiation for mtDNA haplogroups

To test for any overall differences in mtDNA haplogroup distribution between athletes and controls, exact tests of population differentiation were carried out on the haplogroup distributions. When African-American athletes and African-American controls were compared, the two distributions were found to be significantly different ($P < 0.001$). This difference in distributions is illustrated in Figure 4.2A. No significant difference was detected between Jamaican athletes and controls ($P = 0.57 \pm 0.01$, Figure 4.2B).

4.3.4 Haplogroup-specific logistic regression results for mtDNA

The results of the haplotype-specific fitted logistic regressions are presented for African-Americans in Table 4.4 and Jamaicans in Table 4.5. A forest plot of odds ratios and Bonferroni-adjusted 99.69% confidence intervals from both sample populations is shown in Figure 4.3. The “non-L/U6” paragroup was found to be significantly overrepresented in the athletes compared with controls (OR = 2.47; 99.69% CI [1.02-5.31]). No individual haplogroup showed any significant overrepresentation or underrepresentation in the Jamaican dataset. Insignificant and unbounded ORs approaching infinity were observed in both datasets when a particular haplogroup was not observed in either athletes or controls. This occurred for “other L3”, U6 and L4/L7 in African-Americans and U6 in Jamaicans. Haplogroup L5 was entirely absent from African-American athletes and controls while haplogroup L4/L7 was entirely absent from Jamaican athletes and controls.

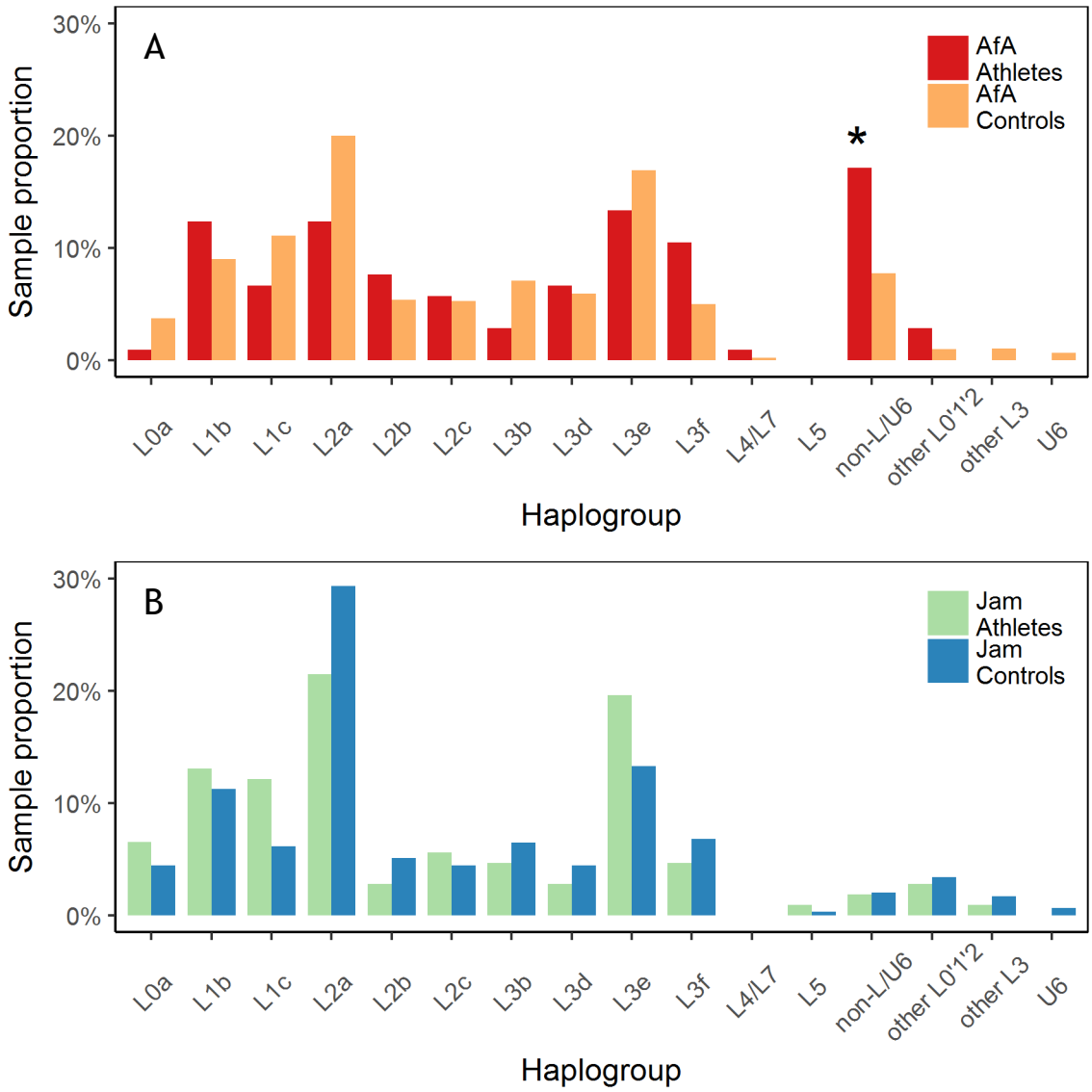


Figure 4.2 Mitochondrial haplogroup distributions of African-American (AfA, A) and Jamaican (Jam, B) athletes and their respective controls. The asterisk denotes significant odds ratio after controlling for multiple testing with a Bonferroni correction. There was a significant difference between the haplogroup distributions of African-American athletes and their controls, although there was no significant difference between the haplogroup distributions of Jamaican athletes and their controls.

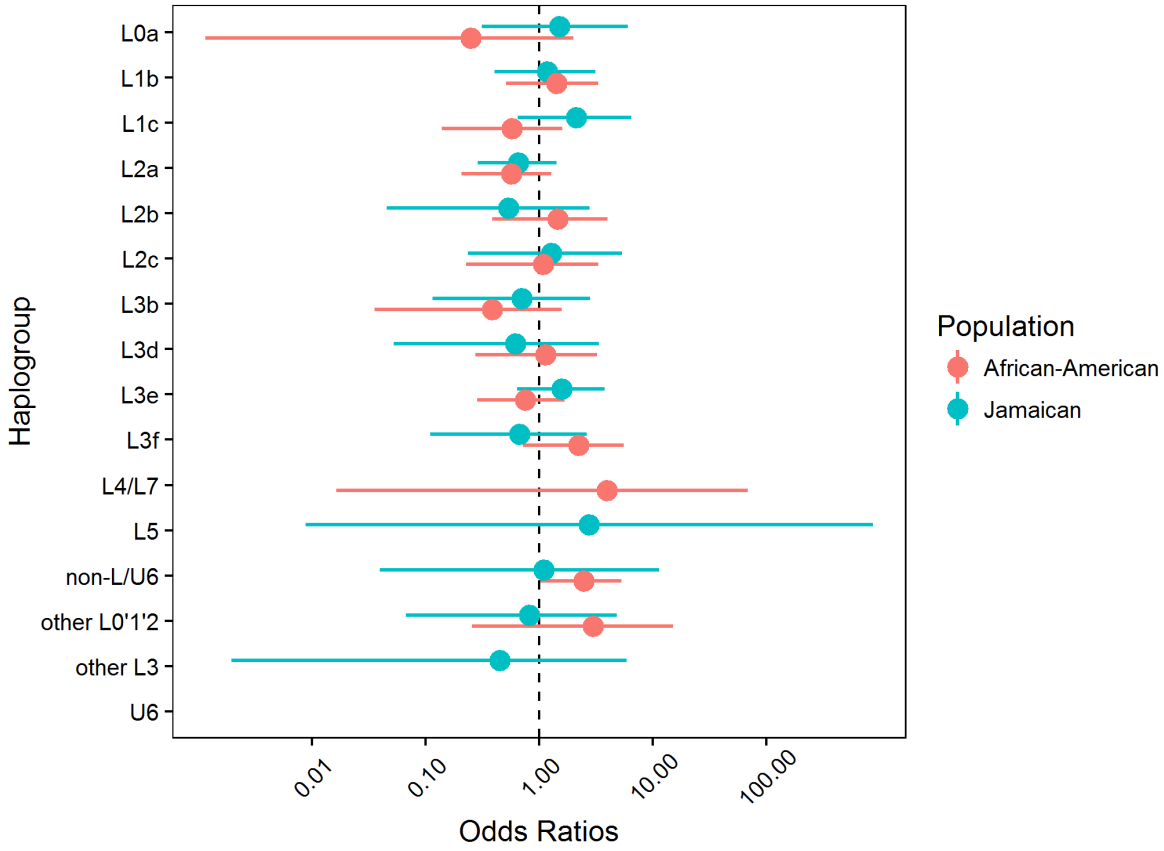


Figure 4.3 Forest plot depicting the mtDNA haplogroup odds ratios for the set of 2 × 2 logistic regression models in African-Americans and Jamaicans. Odds ratios and Bonferroni-adjusted 99.69% confidence intervals are displayed on a log scale. The dashed line indicates an odds ratio of 1. Empty haplogroups indicate that not enough data was present for the model to assign an odds ratio.

Table 4.4 African-Americans: mitochondrial haplogroup counts, frequencies and odds ratios with 99.69% confidence intervals of athletes and controls. Significant overrepresentation is denoted with an asterisk.

Group	Athletes	Controls	OR
L0a	1 (0.95%)	61 (3.71%)	0.25 [0-1.99]
L1b	13 (12.38%)	148 (9.01%)	1.43 [0.52-3.3]
L1c	7 (6.67%)	182 (11.08%)	0.57 [0.14-1.6]
L2a	13 (12.38%)	328 (19.98%)	0.57 [0.21-1.28]
L2b	8 (7.62%)	88 (5.36%)	1.46 [0.39-4]
L2c	6 (5.71%)	87 (5.30%)	1.08 [0.23-3.3]
L3b	3 (2.86%)	116 (7.06%)	0.39 [0.04-1.58]
L3d	7 (6.67%)	97 (5.91%)	1.14 [0.27-3.26]
L3e	14 (13.33%)	278 (16.93%)	0.75 [0.29-1.68]
L3f	11 (10.48%)	82 (4.99%)	2.23 [0.72-5.57]
L4/L7	1 (0.95%)	4 (0.24%)	3.94 [0.02-68.59]
L5	0 (0%)	0 (0%)	NA
non-L/U6	18 (17.14%)	127 (7.73%)	2.47 [1.02-5.31] *
other L0'1'2	3 (2.86%)	16 (0.97%)	2.99 [0.26-15.16]
other L3	0 (0%)	17 (1.04%)	NA
U6	0 (0%)	11 (0.67%)	NA

Table 4.5 Jamaicans: mitochondrial haplogroup counts, frequencies and odds ratios with 99.69% confidence intervals for athletes and controls.

Group	Athletes	Controls	OR with 99.69% CI
L0a	7 (6.54%)	13 (4.44%)	1.51 [0.31-6.03]
L1b	14 (13.08%)	33 (11.26%)	1.19 [0.40-3.13]
L1c	13 (12.15%)	18 (6.14%)	2.11 [0.65-6.52]
L2a	23 (21.5%)	86 (29.35%)	0.66 [0.29-1.42]
L2b	3 (2.8%)	15 (5.12%)	0.53 [0.05-2.78]
L2c	6 (5.61%)	13 (4.44%)	1.28 [0.24-5.36]
L3b	5 (4.67%)	19 (6.48%)	0.71 [0.12-2.82]
L3d	3 (2.8%)	13 (4.44%)	0.62 [0.05-3.37]
L3e	21 (19.63%)	39 (13.31%)	1.59 [0.64-3.78]
L3f	5 (4.67%)	20 (6.83%)	0.67 [0.11-2.64]
L4/L7	0 (0%)	0 (0%)	NA
L5	1 (0.93%)	1 (0.34%)	2.75 [0.01-868.47]
non-L/U6	2 (1.87%)	5 (1.71%)	1.10 [0.04-11.43]
other L0'1'2	3 (2.8%)	10 (3.41%)	0.82 [0.07-4.85]
other L3	1 (0.93%)	6 (2.05%)	0.45 [0.00-5.92]
U6	0 (0%)	2 (0.01%)	NA

4.3.5 Comparison of mtDNA haplogroups across African-American and Jamaican cohorts

An exact test was performed between the mtDNA haplogroup distributions of African-American and Jamaican athletes as well as African-American and Jamaican controls to test for the presence of a phenotype-specific mtDNA haplogroup distribution. The difference in haplogroup distributions between Jamaican and African-American athletes (Figure 4.4A) proved to be significant ($P < 0.001$). African-American athletes have a much larger proportion of the non-Sub-Saharan lineages compared to Jamaican athletes. The comparison between African-American controls and Jamaican controls was also significant ($P < 0.001$, Figure 4.4B).

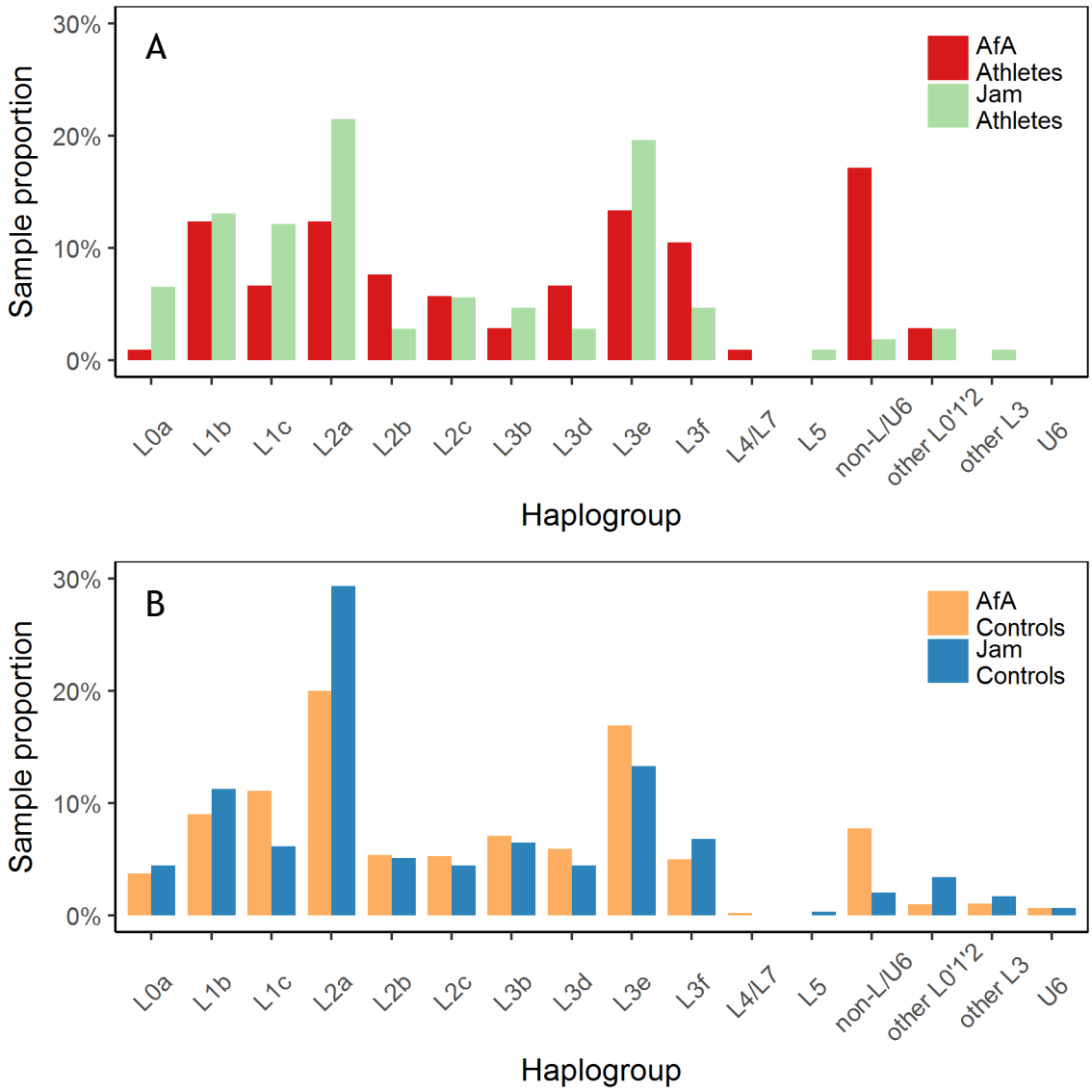


Figure 4.4 Inter-population comparison of mtDNA haplogroup distributions in African-American (AfA) athletes and Jamaican (Jam) athletes (A) as well as African-American controls and Jamaican controls (B). Both paired comparisons of haplogroup distributions were significantly different.

4.4 Results: Non-recombining Y chromosome

4.4.1 Summary statistics: NRY

Most African-American and Jamaican individuals were observed to have either African or European NRY haplogroups. Because of the assay used, the sensitivity within the largest African haplogroup E was sacrificed to detect a broader range of lineages. The male lineages of African-American athletes and controls also consisted predominantly of mostly African haplogroup E (*N* = 38; 76% and *N* = 163; 66%, respectively) and European haplogroup R1b1b2 (*N* = 10; 20% and *N* = 56; 23%, respectively). The remaining lineages observed in the African-American athletes

were Pan-African haplogroup A (*N* = 1) and South Asian haplogroup F(xG, H, I, J, K; *N* = 1). The patrilines of both Jamaican athletes and controls were predominantly haplogroup E (*N* = 29; 66% and *N* = 37; 79%, respectively), and the most common European haplogroup R1b1b2 (*N* = 7; 16% and *N* = 6; 13%, respectively). Both Jamaican athletes and controls had a single individual with the European lineage R1a. There was a greater number of unique lineages detected in Jamaican athletes than controls, including Central African haplogroup B (*N* = 1), Northern European haplogroup I (*N* = 3) and Mediterranean haplogroup J2 (*N* = 3), that were not observed in controls. Among the Jamaican control sample, haplogroup F(xG, H, I, J, K; *N* = *2*), and East Asian haplogroup Q1a (*N* = 1) were observed.

To facilitate meaningful comparisons of NRY haplogroup frequencies between athletes and controls, haplogroups were clustered into haplogroup E, haplogroup R and A(xE, R), representing all other haplogroups not belonging to either E or R. This includes a broad spectrum of global haplogroups, including more uncommon African and European haplogroups. Haplogroup distribution comparisons for African-American and Jamaican athletes are presented in Figure 4.5.

4.4.2 Statistical power of NRY analysis

Estimations of statistical power were also calculated for NRY haplogroup comparisons. Estimates were much lower when compared to those for mtDNA. Assuming a power threshold of 80%, the comparison between African-American athletes and controls (19.6%), African-American controls and Jamaican controls (29.8%), as well as for Jamaican athletes and Jamaican controls (29.1%) and African-American athletes and Jamaican athletes (39.5%) were all below the power necessary to detect a significant difference between haplogroup distributions from the current samples.

4.4.3 Population differentiation for NRY haplogroups

To test for any overall differences between the NRY haplogroup distributions observed in athletes and controls, exact tests of population differentiation were performed on the haplogroup distributions. There was no evidence that the

African-American athlete NRY haplogroupings were statistically different from their controls (P = 0.334 ± 0.01), nor was there evidence that Jamaican athlete haplogroups were significantly different from their controls (P = 0.268 ± 0.01).

4.4.4 Haplogroup-specific logistic regression results for NRY haplogroups

The logistic model OR results and 98.33% confidence intervals for E, R, and A(xE, R) in African-Americans and Jamaicans are presented in Table 4.6 and Table 4.7, respectively, to illustrate overrepresentation of any particular haplogroup when athletes are compared to controls. A forest plot of NRY odds ratios is presented in Figure 4.6. No grouping showed significant overrepresentation for athlete status in either the African-American or the Jamaican cohort.

Table 4.6 African-Americans: NRY haplogroup counts, frequencies and odds ratios with 98.33% confidence intervals for athletes and controls.

Group	Athletes	Controls	OR; 98.33% CI
A(xE, R)	2 (4.0%)	24 (9.68%)	2.59 [0.58-28.48]
E	38 (76.0%)	163 (65.73%)	1.61 [0.71-4.02]
R	10 (20.0%)	59 (23.79%)	0.79 [0.29-1.89]

Table 4.7 Jamaicans: NRY haplogroup counts, frequencies and odds ratios with 98.33% confidence intervals for athletes and controls.

Group	Athletes	Controls	OR; 98.33% CI
A(xE, R)	7 (16.9%)	3 (6.38%)	0.36 [0.05-1.87]
E	29 (66.0%)	37 (78.72%)	0.52 [0.16-1.62]
R	8 (18.1%)	7 (14.89%)	1.27 [0.32-5.16]

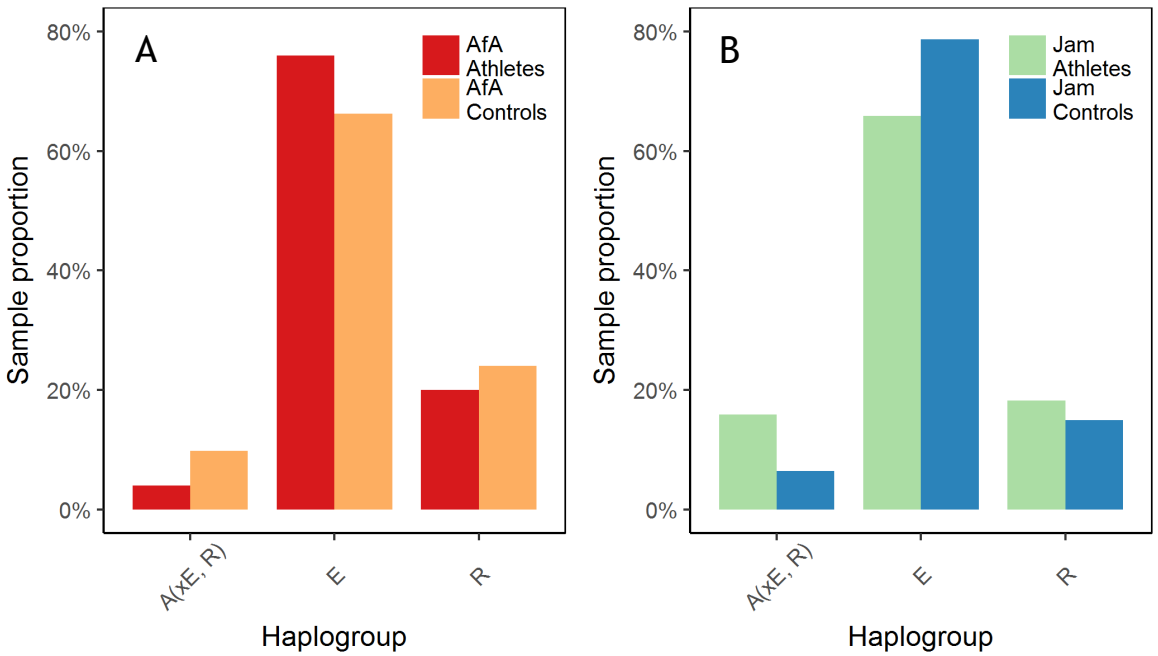


Figure 4.5 NRY haplogroup distributions of African-American (AfA, top) and Jamaican (Jam, bottom) athletes and their respective controls. No haplogroup in either population showed significant overrepresentation assuming a Bonferroni-corrected α = 0.0167. There were no significant differences between the haplogroup distributions of African-American athletes, Jamaican athletes and their respective controls.

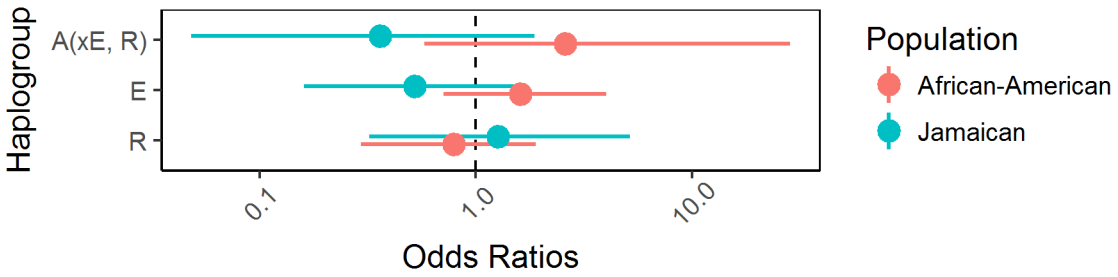


Figure 4.6 Forest plot depicting the NRY group odds ratios for the set of 2 × 2 logistic regression models in African-Americans and Jamaicans. Odds ratios and 98.33% confidence intervals are displayed on a log scale to improve clarity. The dashed line indicates an odds ratio of 1.

4.4.5 Comparison of NRY haplogroups across African-Americans and Jamaicans

An exact test was performed between the NRY haplogroup distributions of African-American and Jamaican athletes as well as African-American and Jamaican controls to test for the presence of a phenotype-specific NRY haplogroup distribution. There was evidence that the haplogroup distributions of African-American and Jamaican athletes (P = 0.039 ± 0.001) and African-American controls and Jamaican controls (0.002 ± 0.001) were different. Unfortunately, the limited statistical power made it inadvisable to draw firm conclusions about whether these differences were statistically significant. A comparison of these distributions is presented in Figure 4.7.

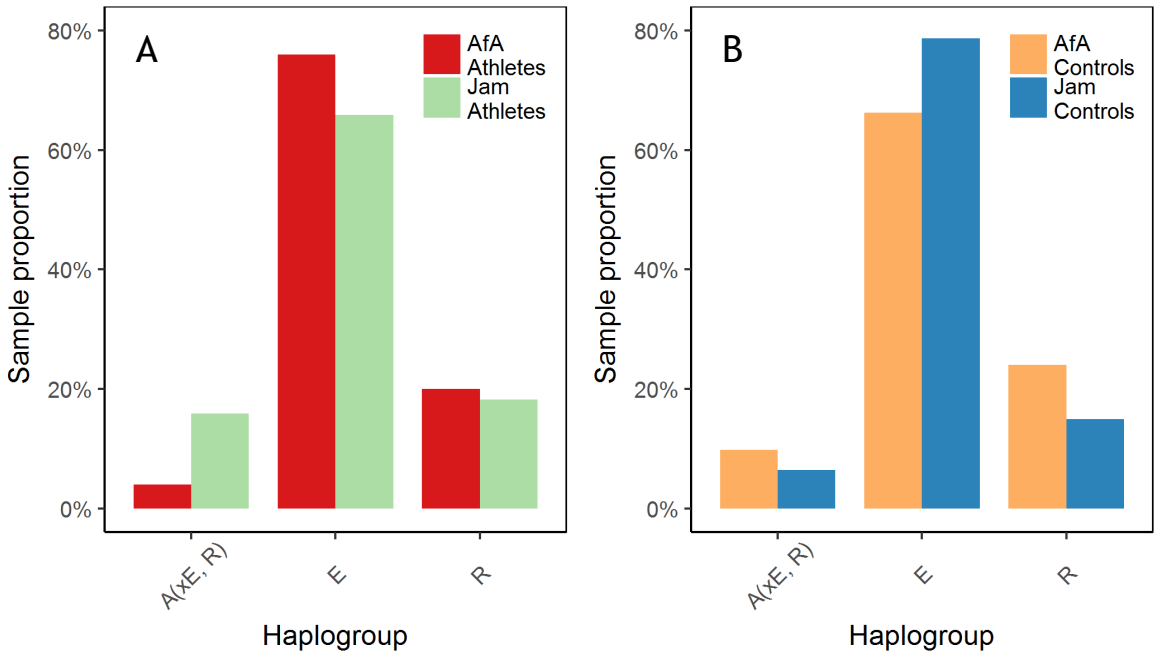


Figure 4.7 Inter-population comparison of NRY haplogroup distributions in African-American (AfA) athletes and Jamaican (Jam) athletes, A, as well as African-American controls and Jamaican controls, B. Both paired comparisons of haplogroup distributions were significantly different.

4.5 Discussion

The main findings of the investigations reported in this chapter are:

a) Mitochondrial haplogroup distributions were significantly different between African-American athletes and controls. The "non-L/U6" paragroup was significantly overrepresented in African-American athletes.

b) Statistical power was too poor to draw any conclusions differences between mitochondrial haplogroup distributions of Jamaican athletes and controls.

c) Mitochondrial haplogroup distributions were significantly different between athletes from both sample populations and between controls from both sample populations.

d) Statistical power was too poor to draw any conclusions about differences between NRY haplogroup distributions of African-American athletes and controls or Jamaican athletes and controls.

e) Statistical power was too poor to draw any conclusions about the differences between NRY haplogroup distributions of African-American and Jamaican athletes and between African-American controls and Jamaican controls.

It was hypothesised that there would be significant differences between the lineage distributions of African-American and Jamaican elite sprint athletes and their respective controls. While there was not sufficient statistical power to draw any conclusions between the maternal lineages in Jamaican athletes and controls and paternal lineages in African-American and Jamaican athletes and controls, this hypothesis was upheld for the maternal lineages in the sample of African-American athletes and controls. Many facets of sprint physiology have been studied to determine these populations' dominance in sprint athletics at the international level (Pitsiladis *et al.*, 2011; Majumdar & Robergs, 2011a; Majumdar & Robergs, 2011b). Previous research has identified candidate genes thought to influence sprint performance in humans (Montgomery *et al.*, 1998; Yang *et al.*, 2003). However, a previous candidate gene study failed to find any significant differences between elite African-American and Jamaican sprint athletes and controls (Scott *et al.*, 2010). Considering previous findings, the differences in the uniparental lineages found in the sampled African-Americans may provide further insight into how ancestry influences elite athlete status.

Mitochondrial genomes from African-American and Jamaican sprint athletes and their respective controls were genotyped to test for any differences in matrilineal composition. There was sufficient total nucleotide sequence diversity to rule out potentially confounding genetic drift in the maternal lineages of both athletes and controls. These values, shown in Table 4.3, were greater to those previously reported from groups of West African descent (π = 7.2 - 8.2) in the Western Hemisphere (Salas *et al.*, 2004). The comparatively large values for π in the African-Americans and Jamaicans may also reflect the substantially larger sample sizes used in this study when compared to Salas *et al.* (2004a), as calculations of π has been shown to be sensitive to the number of unique haplotypes present in a sample (Helgason *et al.*, 2003).

There was significant overrepresentation of the "non-L/U6" mitochondrial paragroup among African-American athletes when compared to controls (Figure 4.2). These results may also suggest a functional difference between non-Sub-Saharan—predominantly European—mitochondrial lineages and Sub-Saharan mitochondrial lineages. Sprint running has been shown to rely at least partially on oxidative phosphorylation, a process of ATP regeneration heavily reliant on proteins encoded by mtDNA (Section 1.2.2). Recent work by Kenney *et al.* (2014)

comparing the bioenergetic capacities of mitochondria isolated from African haplogroup L and European haplogroup H cell lines found that the African type had decreased ATP turnover rates, consistent with more efficient energy production. Considering those findings, it seems unlikely that non-Sub-Saharan mitochondrial lineages confer an advantage to sprint performance.

Alternatively, no individual haplogroups were associated with athlete status in the African-American sprint athletes, providing evidence against potential assortative mating for the suite of phenotypes associated with elite sprint athletics. Genetic admixture in African-Americans is widely acknowledged to be the result predominantly of gene flow from European men to African women (Parra *et al.*, 1998), and while the controls fit this pattern, the athletes did not. The overrepresentation of the non-Sub-Saharan types in athletes may indicate that African-Americans with atypical genetic admixture have a greater likelihood of becoming elite sprinters when compared to African-Americans with Sub-Saharan mtDNA. European-Americans have enjoyed a privileged socioeconomic position relative to African-Americans during the period of formation of these populations (Rank, 2009). Access to resources has been shown to play a crucial role in the likelihood of a potentially-elite athlete engaging in high-quality training (Fu, 2008). African-American athletes with a more recent European-American ancestor could also be of a higher socioeconomic status and therefore result in a more optimal environment in which to develop athletically. However, the extrapolation of these results is purely speculative, and to detect any association between athlete status and socioeconomic status would require measurement socioeconomic status directly or use of a more informative proxy, such as genetic ancestry (Kirkegaard *et al.*, 2016). The relationship between a more recent admixture event and athlete status is tested in Chapter 5.

Contrary to the earlier hypothesis, the mtDNA haplogroups of the sampled Jamaican athletes did not significantly differ from controls (Figure 4.2). Jamaica has a comparatively homogeneous ethnic identity (Fearon, 2003) and the prestige of the “Champs” athletics meet (Section 1.1.2) may encourage children from all backgrounds to pursue the sport, ultimately leading to a more homogenous distribution of lineages in athletes. However, it is much more likely that these tests suffered from a lack of statistical power, making it difficult to refute the

null hypothesis that Jamaican athletes and controls have the same haplogroup distribution when the likely size of the signal is considered.

The paternal lineages of both African-Americans and Jamaicans were largely a mixture of European and African haplotypes. Because men are effectively able to reproduce without an interbirth interval, the Y-chromosome has a different set of life history constraints to the mitochondrial genome and has much lower genetic diversity (Jobling & Tyler-Smith, 2003). The high frequency of R1b1b2 in both samples likely stems from this haplogroup's high prevalence in Great Britain and Western Europe (Balaresque *et al.*, 2010), not unexpected given the known source of European migrants to both the United States and Jamaica.

There were no significant differences between the NRY haplogroup profiles of African-American or Jamaican athletes and their respective controls (Table 4.6 and Table 4.7). The paternal lineages found in Western Europe and West Africa show signs of reduced variability from the population bottleneck which resulted from cultural expansion in Europe and Africa, respectively (Cruciani *et al.*, 2002; Balaresque *et al.*, 2010). This reduced variability potentially made any significant differences between athletes and controls more difficult to detect in these populations when compared to mtDNA. Statistical power was poor when comparing athletes and controls. Future work should concentrate on genotyping a greater number of polymorphic loci with a specially designed assay to improve haplogroup resolution in targeted lineages. When combined with a greatly increased sample size, this process would not only improve statistical power, but also allow for improved resolution of the lineages present in the samples to be inferred.

There was a significant difference between the mtDNA haplogroup distribution of African-American and Jamaican athletes. These results suggest that the elite sprint phenomenon may not manifest itself in these populations as a distribution of maternal lineages and that there may be different biological mechanisms underlying elite sprint running in the two populations. There were also significant differences between African-American and Jamaican controls. These results were expected when considering their dissimilar relationships with the Trans-Atlantic slave trade and subsequent differences in selection pressures acting on the two populations. Unfortunately, due to lack of statistical power, nothing definitive can be learned from the paternal lineage comparisons between African-American

athletes and Jamaican athletes and between African-American controls and Jamaican controls.

In conclusion, there was a significant difference between the mtDNA haplogroup distributions of African-American sprint athletes and controls driven by the overrepresentation of lineages originating from outside Sub-Saharan Africa. This result suggest that socioeconomic or other environmental variables may be important in determining elite sprint athlete status in self-declared African-Americans, although inferring genome-wide ancestry and the estimated number of generations since admixture began will undoubtedly be more informative to test this hypothesis. Using genome-side SNP array data, these tests are performed in the subsequent chapter. There were no discernible differences in the uniparental markers between Jamaican sprint athletes and controls. These results indicate that Jamaican athletes originate from the same population as Jamaican controls. There were significant differences between African-American and Jamaican sprint athletes, suggesting that there was not a particular ancestral make-up associated with the elite sprint phenomenon. There were significant differences between the maternal lineages distributions of African-American and Jamaican controls. These results reaffirm the different demographic histories and ancestral make-up of the two groups. Unfortunately, there was not enough statistical power to draw conclusions about the paternal lineage distributions between control groups. Since atypical directional admixture appears to be an important predictor of elite athlete status, the next chapter will examine fine-scale autosomal admixture to evaluate any influence locus-specific ancestry has on elite sprint athlete status.

5

Is there any evidence for an association between locus-specific genetic ancestry and elite sprint status in populations of African-Americans and Jamaicans?

5.1 Introduction

The previous chapter examined mitochondrial and Y chromosome haplogroup profiles of African-American and Jamaican athletes and controls to determine if there was any evidence that the population distributions of athletes and controls differed significantly. The results suggested that while Jamaican athletes and controls likely originated from a homogeneous population, the matrilineal pool of African-American athletes was significantly different from African-American controls. The overrepresentation of non-Sub-Saharan lineages in the African-American athletes was contrary to what was expected, and the results suggested that potential genetic differences between African-American elite sprint athletes and controls might be linked to admixture between West Africans and Europeans.

Genetic ancestry is inferred by comparing the allele frequencies in known geographically isolated source populations with an admixed population. The choice of these source populations is critical, since chosen populations act as an approximation for the true ancestral populations (Shriner, 2013). African-Americans and Jamaicans are predominantly admixed between West Africans and Europeans. However, their actual ancestral populations are unlikely to be represented fully by publicly available genotype data given the broad geographical areas and genetic diversity of each source population. Fortunately, since the number of generations since admixture began is sufficiently small (Shriner, 2013), source populations in Africa and Europe are unlikely to have experienced much genetic drift since admixture began (Keinan *et al.*, 2009). Additionally, the

observed genetic distance between West Africans and Europeans, defined as the amount of genetic divergence between two groups, is greater than the genetic distance within either group (1000 Genomes Project Consortium, 2010), allowing publicly available West African and European genetic data to serve as representatives for the actual unknown source populations.

Genetic ancestry is invaluable for inferring population history of admixed populations and can be divided into both local and genome-wide ancestry. Local ancestry refers to the assignment of genetic ancestry from a source population to individual loci within each chromosomal segment, while genome-wide ancestry refers to the proportion of genetic ancestry a source population contributes across the full length of the genome. Once the ancestral origins of each chromosomal segment are inferred to have zero, one or two copies derived from a given source population, it is also possible to estimate the number of generations since admixture began for each individual by identifying the number of times the ancestry switches across the chromosomes (Price *et al.*, 2009).

The number of switches increases with each generation from the initial admixture event because of meiotic recombination. With each subsequent generation, these chromosomal segments continue to split at a predictable rate. The amount of ancestral fragmentation along the chromosome can be used to infer the number of generations since the initial admixture event. It is then possible to test for an association between these inferred generations times and a phenotype of interest. For example, European-Americans are more likely to be in better socioeconomic standing than African-Americans in the United States (DeNavas-Walt *et al.*, 2012), and a negative relationship has been shown between socioeconomic status and higher genome-wide African ancestry in African-Americans (Kirkegaard *et al.*, 2016).

These same estimates of local and genome-wide ancestry can be used to detect genetic markers associated with a given phenotype. When there are observable differences between the prevalence of a phenotype in the isolated source populations of an admixed group, it is possible to map the local genetic ancestry in these admixed populations to loci associated with a given phenotype. Because local ancestry is highly autocorrelated along each chromosome, fewer effective tests for association are made, increasing the statistical power (McKeigue, 1998;

Hoggart *et al.*, 2004; Shriner *et al.*, 2011). Since there is an observable difference in the number of elite sprint athletes from Africa and Europe, admixture mapping may be insightful in detecting the causative variants for elite sprint athlete status in African-Americans and Jamaicans. Admixture mapping has been used successfully in the past to map phenotypes to genomic loci in African-Americans, including asthma (Mathias *et al.*, 2010), renal disease (Kao *et al.*, 2008) and prostate cancer (Freedman *et al.*, 2006). Admixture mapping is also likely to be effective in Jamaicans given their admixed population history, although there have not been any published studies to date.

The aim of this chapter was firstly to determine if the genome-wide ancestry estimates are significantly different between athletes and controls. Based on the results from the previous chapter, it was hypothesised that African-American athletes would have a significantly lower proportion of genome-wide African ancestry than controls, while Jamaican athletes would have similar levels of genome-wide African ancestry when compared to controls. Additionally, it was also hypothesised that because of the more favourable socioeconomic status associated with lower levels of genome-wide African ancestry, the number of generations since admixture began in African-American athletes would be significantly lower than African-American controls, and that there would be no significant difference in time since admixture began for Jamaican athletes and controls. Finally, it was hypothesised that regions of the genome associated with elite sprint running would be enriched for European ancestry in African-American athletes when compared to African-American controls while there would be no significant differences between Jamaican athletes and controls. This was the first known attempt to use admixture mapping to identify genomic loci associated with elite athlete status.

5.2 Methods

5.2.1 Sample size and genotyping methodology

One hundred and eighteen African-American athletes and 47 controls were genotyped using the HumanOmni1-Quad BeadChip (Illumina, CA, USA). Additionally, 93 Jamaican athletes and 101 Jamaican controls were genotyped: eight athletes and eight controls were genotyped at 1,140,419 markers using the

HumanOmni1-Quad BeadChip (Illumina, CA, USA) and 185 individuals at 731,442 markers using the HumanOmniExpress BeadChip (Illumina, CA, USA). These African-American and Jamaican data were generated by Guan Wang and Noriyuki Fuku at the Tokyo Metropolitan Institute of Gerontology. An additional 350 African-American controls genotyped using the HumanOmni1-Quad Beadchip from an unrelated study investigating ischemic stroke in young adults were used to bolster the African-American control numbers (Cheng *et al.*, 2011).

5.2.2 Quality control

A published protocol describing the quality control process for genotype data from the Illumina Infinium HumanExome BeadChip (Guo *et al.*, 2014) was adapted for the current study to process the African-American and Jamaican genotype data using the GENOMESTUDIO genotyping module (Illumina, CA, USA). The recorded sex of each sample was confirmed by the software using the rate of heterozygosity found on the X chromosome. Once these data were read into the software, the proprietary GenTrain algorithm clustered each SNP and assigned a score to the fit. The GenTrain score is calculated using a penalised neural-network clustering algorithm developed to mimic genotype calls made visually by an expert investigator (Illumina Inc., 2005). Samples with a call success rate below 95% were excluded from the dataset. Because genotypes from excluded samples may have negatively affected the GenTrain clustering, SNPs with a GenTrain score below 0.7 were clustered again. After the autosomes were extracted, SNPs with a call frequency below 95% were excluded. Finally, the data were extracted in PED format for use with the PLINK genome analytical software (Purcell *et al.*, 2007). The cleaned African-American and Jamaican genotype data were then combined and any SNPs not present in both cohorts were excluded from the dataset.

5.2.3 Reference genotype data

West African and Western European genotype data from the third phase of the publicly available 1000 Genomes Project (1000 Genomes Project Consortium, 2012) were used as reference populations for detection of genetic ancestry in the cohorts of African-Americans and Jamaicans. Due to the size of the 1000 Genomes dataset, each chromosome was processed individually to circumvent limitations

with computer memory. These downloaded data were first converted from the compressed VCF file format into the PED format most commonly used with PLINK. Next, only those individuals belonging to either West African (Esan of Nigeria, *N* = 99; Gambian in Western Divisions of Gambia, *N* = 113; Mende of Sierra Leone, *N* = 85; Yoruba of Nigeria, *N* = 108) or Western European (Utah residents with Northern and Western European ancestry, *N* = 99; British in England and Scotland, *N* = 91; Iberian populations in Spain; *N* = 107) were extracted. Other African or European populations available from 1000 Genomes were not included in the combined reference populations due to their low likelihood of contributing to genetic admixture in either African-Americans or Jamaicans (e.g. Finnish from Finland and Luhya from Kenya). Lastly, with the sample size greatly reduced, the individual chromosomes were combined into a working reference dataset.

Next, the subset of the 1000 Genomes reference data described above were combined with the African-American and Jamaican data to create the working dataset used for the genetic ancestry inferences. Data were combined according to methodology described in the PLINK documentation to systematically address any DNA strand direction misalignments that may arise from using different genotyping technologies. First, an attempt was made to merge the combined African-American and Jamaican and the reference populations from 1000 Genomes. Any SNPs with a potential strand mismatch were automatically stored in an accompanying file before the merging procedure terminated. Next, the stored SNPs were flipped in the African-American and Jamaican dataset before an attempt was made again to merge these data with the reference data. If any SNPs were still thought to be misaligned, it is possible that these SNPs are triallelic. Plink is unable to handle triallelic SNPs, and as a result, these SNPs were then removed from both the admixed and reference datasets. Once these SNPs were removed, these data merged without issue, leaving 641,575 SNPs in the working dataset. This value greatly exceeded the suggested number of 50,000 markers required to detect all changes in local ancestry in a sample population of African-Americans (Shriner, 2013). These data were then separated again into African-American, Jamaican, African and European samples and saved in the BINBAM format for further analysis.

5.2.4 Local ancestry estimation

There are several different programs that can be used to estimate local ancestry. However, given the structure and composition of the available data, the program ELAI (Efficient Local Ancestry Inference) was selected (Guan, 2014). The available African-American and Jamaican data were unphased, meaning that individual haplotypes were not known. While it was possible to determine the haplotype ordering of unphased data using a similar population sample (Howie *et al.*, 2009), this process had the potential to introduce spurious results in downstream analyses. ELAI can process both phased and unphased data as input, the difference largely being computation time as haplotypes must be inferred from input data as part of the algorithm (Guan, 2014).

ELAI uses a two-layer Hidden Markov model to estimate local ancestry. This is roughly equivalent to integrating existing algorithms by combining lower layer haplotype clustering found in programs such as fastPHASE (Scheet & Stephens, 2006) and BEAGLE (Browning & Browning, 2007) with upper layer clustering algorithms such as STRUCTURE (Pritchard *et al.*, 2000) and ADMIXTURE (Alexander *et al.*, 2009). ELAI is designed to learn the haplotype structure from the data. Calculating admixture in this way has the benefit of detecting and condensing haplotype blocks of a particular ancestry. Since African-Americans and Jamaicans are predominantly of African and European ancestry, two upper layer clusters were assumed for these estimations. Following the recommendations for African-Americans published alongside ELAI (Guan, 2014), ten lower layer clusters were assumed to make the haplotype-level inferences.

Ten generations of admixture were assumed in these calculations for African-Americans as recommended by the author. This allowed ELAI to model the gradual amounts of interbreeding through time. There is considerable variation in the estimated number of generations since admixture began in the literature for African-Americans, ranging from six (Tian *et al.*, 2006) to 20 generations (Smith *et al.*, 2004). ELAI has been shown in simulation studies to infer robust estimates for ancestry even when the number of generations varies by up to a multiple of two (Guan, 2014). As such, it was also assumed that admixture began ten generations ago for the Jamaicans despite no available reference in the literature.

ELAI uses an expectation maximisation (EM) algorithm to estimate the number of alleles for a chosen reference population at each locus, a value known as dosage. An example of these dosages and the density of markers used to calculate dosage is shown below in Figure 5.1. Because the high dimensionality of the parameter set used by ELAI may result in multimodal distributions, the EM tends to converge on local rather than global modes. This may happen when the initial seed value is too close to a local maximum of the underlying likelihood function. To minimise any influence these local convergences may have had on the model outcome, the mean of ten runs of ELAI were calculated to achieve robust estimates for allelic dosage.

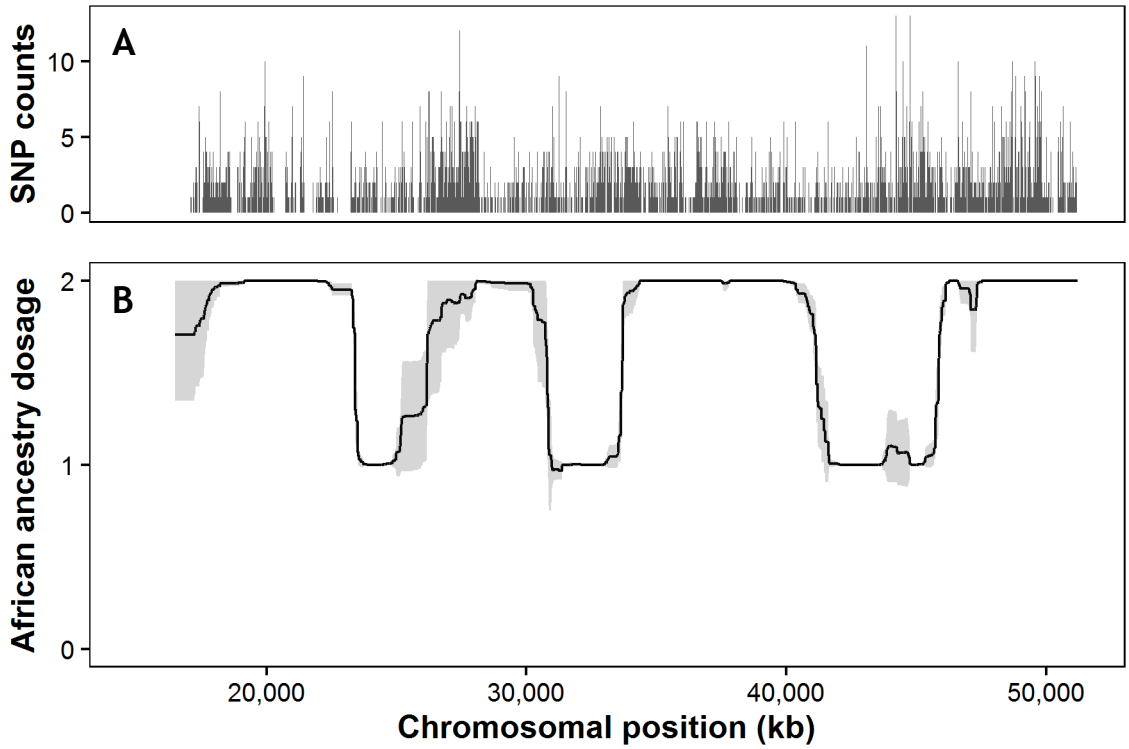


Figure 5.1 Example mean local African ancestry dosage calculations after 10 runs of ELAI for a single individual along chromosome 22. A) The histogram shows the relative density of SNPs along the chromosome. B) The line plot shows the estimated number of African alleles present at each point in the chromosome with the grey shading indicating standard deviation across the 10 runs. Six switches in ancestry are shown in this example.

5.2.5 Genome-wide ancestry calculation and comparison

Genome-wide estimates of African ancestry were calculated by taking the mean of the local ancestry dosages across the length of the genome and dividing by two, the number of alleles at each locus. To test if the distribution of genome-wide African ancestry coefficients were significantly different between athletes and controls and to accommodate the long left tail typically seen in the distributions of African ancestry in African-Americans and Jamaicans (Section 1.7.3), the non-parametric Mann-Whitney-U test was used.

5.2.6 Generations since admixture

To estimate the number of generations since the initial admixture event, local African ancestry dosage data were used to calculate the number of ancestry switches per individual across the entire genome. A switch in ancestry was calculated each time the estimated ancestry dosage fixed on either zero, one or two African alleles (see Figure 5.1 above). The following equation was then applied:

$$G = \frac{N}{4\mu(1-\mu) \times 36.25}$$

where G was the estimated number of generations since the initial admixture event, N was the number of switches, μ was the estimated genome-wide African ancestry, and 36.25 was the mean length of the autosomes for African-Americans in Morgans, a measure of genetic linkage between chromosomal positions (Hinch *et al.*, 2011). A Mann-Whitney-U test was used to compare the estimated number of generations between athletes and controls in both cohorts. This test was used to determine whether there was evidence that athletes and controls had differing estimates for their distributions of number of generations since admixture.

5.2.7 Bayesian methodology for admixture mapping

Admixture mapping was processed in a Bayesian framework to test for an association between local ancestry dosage and elite athlete status. This methodology was published as part of the six-part joint ancestry and association algorithm BMIX introduced by Shriner *et al.* (2011). The methodology was adapted

specifically for this study to accommodate the binary case-control response variable instead of the continuous response cases-only framework presented originally in Shriner *et al.* (2011). This method directly measures the probability of a hypothesis given the data and a set of assumptions by using posterior probabilities, as opposed to *P*-values which would measure the probability of a set of data given a hypothesis. Stated another way, the posterior probability assigns a probability to the test hypothesis being true, while the P-value assigns the probability of observing hypothetical data as or more extreme than the actual data assuming the null hypothesis is true. The naïve objective threshold for posterior probability significance was 0.5, at which point the hypothesis favoured by the posterior odds would switch to a test hypothesis of an association with athlete status at the given locus.

5.2.7.1 Applied correction for multiple testing of admixture mapping

The number of effective tests was needed to calculate a population-specific significance threshold for the genome-wide associations being tested. The number of tests was estimated using an autoregressive model, considering the correlation of dosage estimates from one locus to the next along the genome. This method allows for the applied correction of multiple testing, or testing burden, for the association analysis to be estimated. This correction was done by evaluating the frequency at which dosage values change along the genome, providing an indicator of the total number of distinct ancestral components. These calculations were performed using the R package CODA (Plummer *et al.*, 2006). The number of tests was calculated for each individual, and the mean value across all individuals was used in calculating the genome-wide significance threshold. For African-Americans, the mean number of effective tests was 330.887 leading to an $\alpha = \frac{0.05}{330.887} = 1.51\times10^{-4}$ level of significance, while the mean number of effective tests in the Jamaican cohort was 411.381 with an $\alpha = \frac{0.05}{411.381} = 1.22\times10^{-4}$ level of significance.

The non-centrality parameter λ was used to calculate the central χ_1^2 density to test the hypothesis that ancestry at a given locus was significantly associated with athlete status. Assuming that statistical power is $1-\beta$ = 0.8, the two-sided non-centrality parameter λ was calculated as

$$1-\beta = \phi(\sqrt{\lambda} - \phi^{-1}\left(1-\frac{\alpha}{2}\right)) + \phi(-\sqrt{\lambda} - \phi^{-1}\left(1-\frac{\alpha}{2}\right))$$

where β was the type II error rate, ϕ was the standard normal cumulative distribution function, ϕ^{-1} was the standard normal quantile function and values of α were as stated above. The values of λ were 21.445 and 21.946 for African-American and Jamaicans, respectively.

5.2.7.2 Admixture mapping

Binary logistic regressions were fitted at each locus to assess the association between African ancestry dosage and athlete status. Genome-wide African ancestry estimates were added as a covariate to control for underlying population structure such that:

$$logit(athlete\ status) \sim local\ ancestry + global\ ancestry$$

Once the model had been fitted, the corresponding *P*-value for the local ancestry covariate from each locus-specific model was then converted into posterior probabilities to test the null hypothesis that a given locus was not associated with athlete status.

P-value conversion was accomplished first by transforming the quantile associated with the locus-specific *P*-value into a χ_1^2 statistic. This statistic, D, was then converted into a posterior probability using Bayes' theorem:

$$P(H_1|\mathrm{D}) = \frac{P(\mathrm{D}|H_1)P(H_1)}{P(\mathrm{D}|H_1)P(H_1) + P(\mathrm{D}|H_0)P(H_0)}$$

where $P(H_0)$ is the prior probability of the null hypothesis defined as 1 over the effective number of tests ($\frac{1}{330.887} = 3.02 \times 10^{-3}$ and $\frac{1}{411.381} = 2.43 \times 10^{-3}$ for African-Americans and Jamaicans, respectively), $P(H_1)$ is the prior probability of rejecting the null hypothesis $1 - P(H_0)$, $P(\mathrm{D}|H_0)$ is probability of D given the null hypothesis, calculated as the χ_1^2 density function for D centred at 0, $P(\mathrm{D}|H_1)$ is

the probability of D given the test hypothesis, calculated as the χ_1^2 density function for D centred at λ, the non-centrality parameter defined in Section 5.2.7.1 above. As each locus-specific association model was fitted independently from one another, calculations were performed in parallel using the doParallel (Revolution Analytics & Weston, 2014) package in R to decrease total running time.

5.2.8 Association assessment

Two separate summary figures were generated to illustrate any association between local ancestry dosage and athlete status. Quantile-Quantile (Q-Q) plots were generated to assess the resulting significance of any association under the assumption that locus-specific *P*-values were random and uniformly distributed under the null hypothesis of no association. While deviances from the line of perfect correspondence are expected at the extreme right tail of the distribution when highly significant associations are found, a substantial proportion of points above the expected value may indicate that either the underlying data or the model are problematic, while a large proportion of points below the expected value likely indicates a lack of statistical power. The genomic inflation factor λ was calculated by dividing the median of the association *P*-values chi-squared test statistic by the expected median from a chi-squared distribution with one degree of freedom. Genomic inflation factors above one indicate that there may be population stratification that has not been accounted for in the association tests.

Manhattan plots were also generated to illustrate any significant associations resulting from the admixture mapping. Posterior probabilities generated above were plotted against their genomic position with significant associations visible as peaks rising above statistical noise. Peaks were defined here as the cluster of SNPs rising above the 0.5 posterior probability significance threshold. Solitary points rising above the threshold were likely false positives, although given the nature of the SNP density and haplotype structure, these points could also represent a small ancestral genomic block with only one typed SNP.

5.3 Results

To test for any association with local genetic ancestry and athlete status, the mean and standard deviation of the local African ancestry dosages were summarised over the 10 ELAI model runs. By taking the mean of the local ancestry dosages along the entire genome and dividing the estimated number of alleles by two, it was possible to infer genome-wide African ancestry. The estimated genome-wide mean African ancestry for African-American athletes was 71.0% (SD = 27%, range [0% - 100%]) while the estimated mean genome-wide African ancestry for African-American controls was 81.8% (SD = 9.53%, range [45% - 100%]). Two of these athletes and 12 of these controls were inferred to have 100% African ancestry. The estimated mean genome-wide African ancestry for the Jamaican athletes was 87.0% (SD = 11.0%, range [49% - 100%]) while the estimated mean genome-wide African ancestry for Jamaican controls was 86.0% (SD = 13.0%, range [47% - 100%]). Ten of these athletes and nine of these controls were inferred to have 100% African ancestry.

Upon inspection of the genome-wide ancestry estimations, there were 11 African-American athletes with less than 5% genome-wide African ancestry. The next closest athlete had roughly 45% African ancestry. While African-Americans with very little African ancestry have been observed in the literature (Sinha *et al.*, 2006; Bryc *et al.*, 2010), there were concerns that these individuals would confound any associations based on ancestry as there were not a comparable number of controls with very little African ancestry. After removing those 11 athletes with estimated genome-wide African ancestry below 5%, the estimated mean genome-wide African ancestry for the remaining 97 African-American athletes rose to 78% (SD = 16%, range [46% - 100%]), much closer to the mean of 71% observed in the control sample.

The results of the reduced set of African-American athletes and controls and Jamaican athletes and controls indicated that there was no significant difference between the estimates of genome-wide African ancestry (P = 0.480 and P = 0.692, respectively). The genome-wide ancestry estimations are presented as ordinal line plots for African-Americans (Figure 5.2A) and Jamaicans (Figure 5.2B). A truncated version of these results including the athletes excluded for very low amounts of African ancestry is included in Section 7.5.

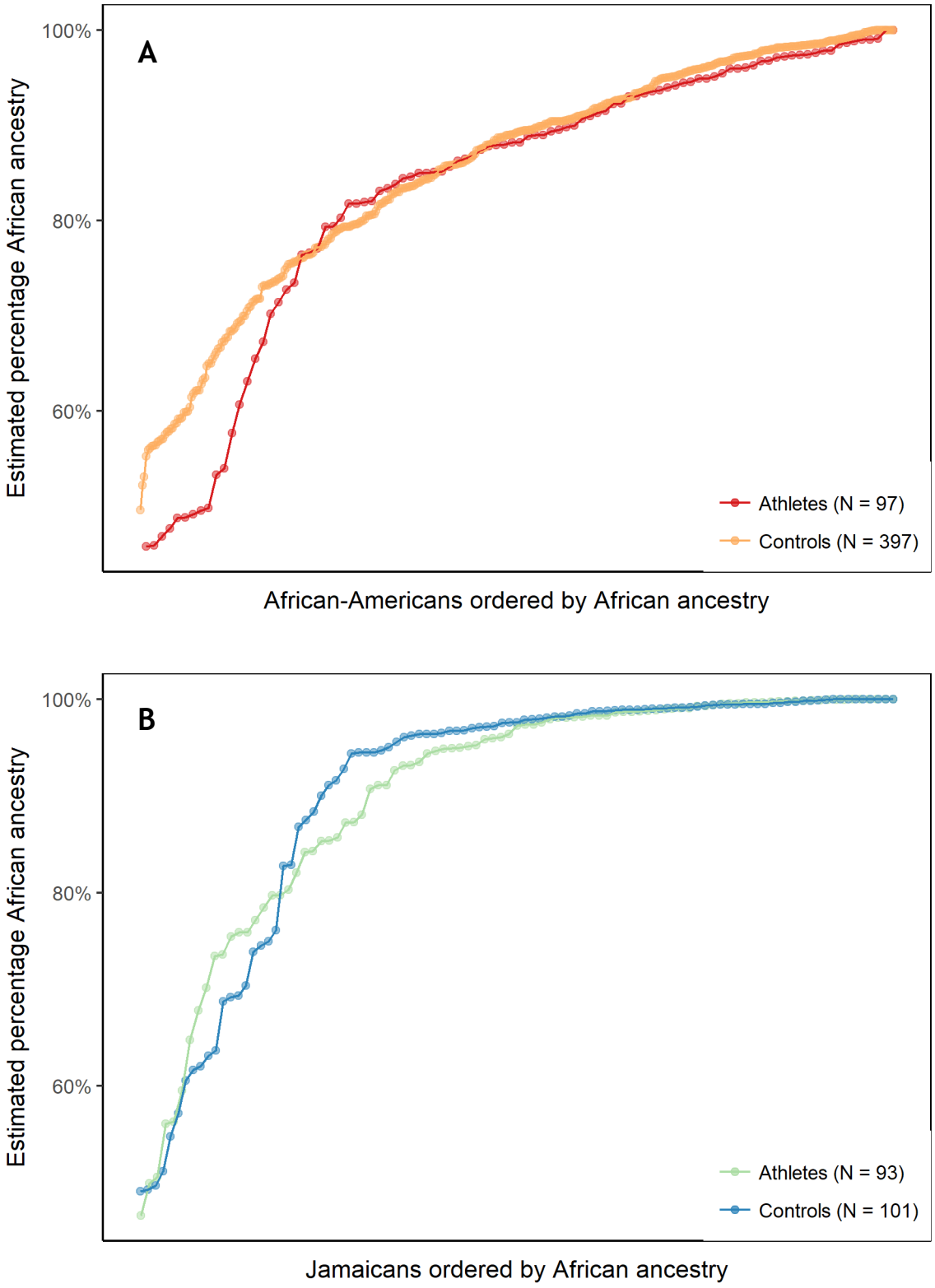


Figure 5.2 Ordered line plot of genome-wide African ancestry for A) African-Americans with the 11 outliers excluded and B) Jamaicans. Each point represents an individual.

5.3.1 Estimated generations since admixture

The estimated number of generations since admixture as a function of inferred percentage African ancestry, assuming a simplistic model of a single admixture event for each individual, is presented for African-Americans and Jamaicans in Figure 5.3A and Figure 5.3B, respectively. These estimates were not rounded to the nearest full generation. As an artefact of the formula used to calculate the estimated number of generations, individuals with fewer detected switches in ancestry than would be expected given their genome-wide African ancestry could potentially have an estimated number of generations since admixture below one (e.g. an individual with 50% genome-wide African ancestry and 36 or fewer detected switches).

The median (and range) of generations since admixture for African-American athletes was 13.81 [0.50-19.55] generations while the median of generation since admixture for African-American controls was 13.95 [0.30-21.85] generations. Individuals without any detectable switches in ancestry were excluded from these summary statistics. The results of the Mann-Whitney-U test suggest that there was no significant difference between the number of switches in African-American athletes and controls (P = 0.182). The median of estimated generations since admixture for Jamaican athletes was 13.89 [1.63-20.15] generations while the median of estimated generations since admixture for Jamaican controls was 13.54 [0.52-46.95] generations. The results of the Mann-Whitney-U test suggest that there was no significant difference between the number of switches in Jamaican athletes and controls (P = 0.923). The overall trend in Figure 5.3A and Figure 5.3B is symmetrical, with the estimated number of generations since the initial admixture event rising as and estimated individual's African ancestry approaches the limits of either 0% or 100%. The variance in the estimated number of generations also rises with the increase in African ancestry in both populations as well, suggesting the algorithm developed to detect ancestry switches may lose sensitivity as ancestry increases.

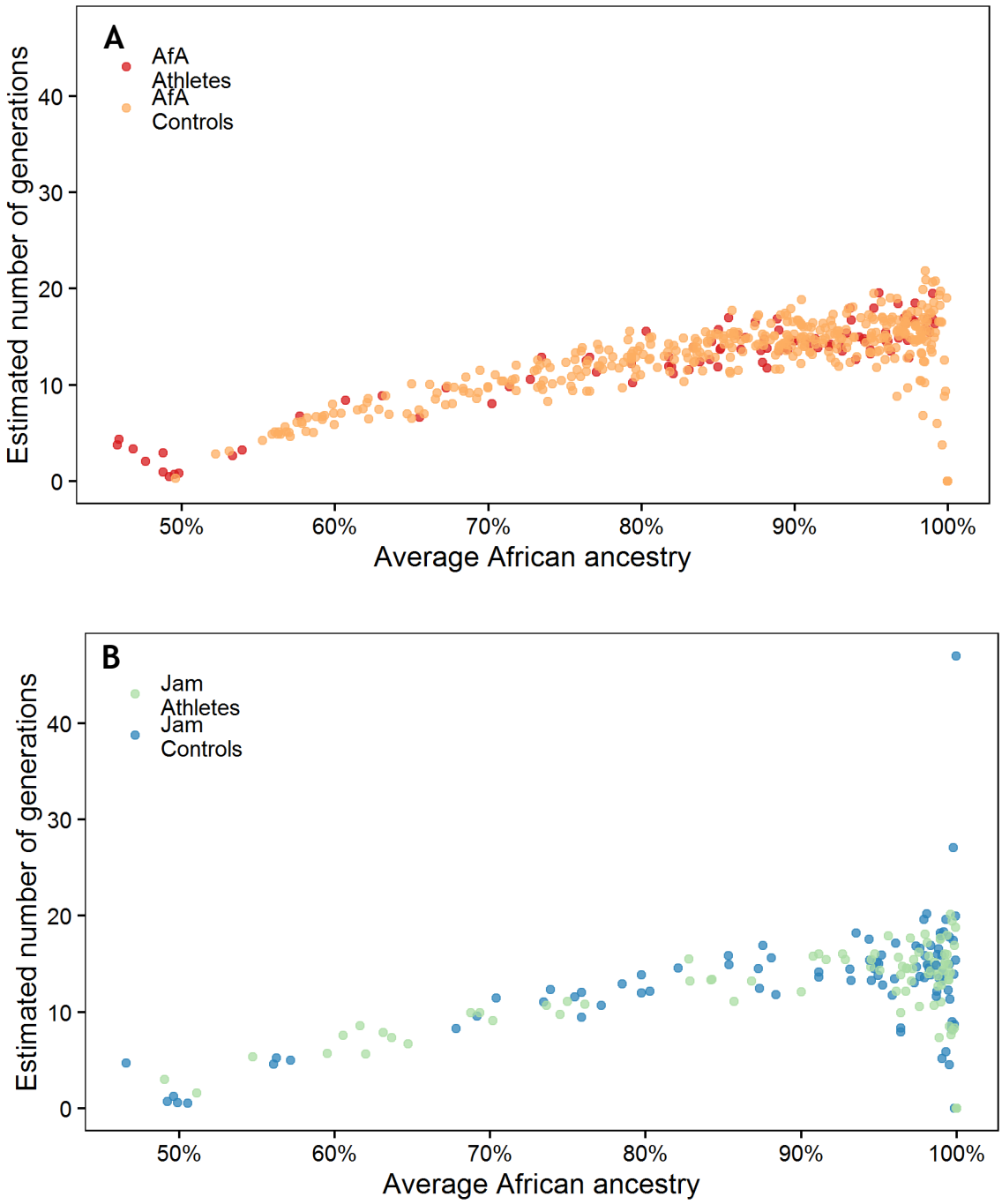
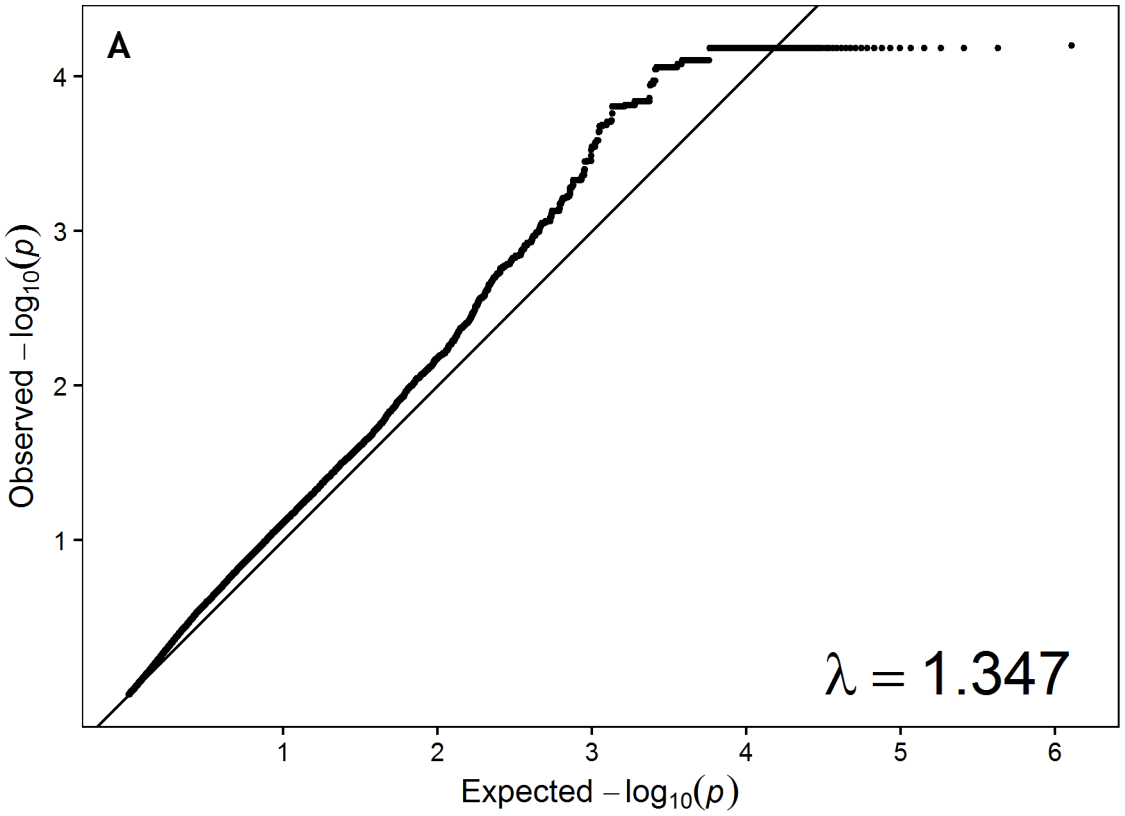


Figure 5.3 Relationship between estimated average African ancestry and the estimated number of generations since admixture for A) African-American and B) Jamaican athletes and controls. Each point represents an individual. Generations have not been rounded to the nearest full generation. Individuals with an inferred African ancestry of 100% are not included in these figures because they did not have any detectable European ancestry to model.

5.3.2 Admixture Mapping

Q-Q plots were generated to evaluate whether the *P*-values fitted a random uniform distribution. The African-American distribution (Figure 5.4A) rises from the expected line quickly, indicating a greater number of strong associations than would be expected by chance. However, the distribution has a strong right tail below the line of perfect correspondence, indicating fewer significant associations at this end of the distribution than would be expected. The inflation factor of 1.347 suggested the calculated *P*-values fell outside the expected random uniform distribution. These results were in stark contrast to the Jamaicans (Figure 5.4B), who appear to have a distribution that is below one that would be expected. The inflation factor of 0.981 indicated that the calculated *P*-values fell outside the expected random uniform distribution.



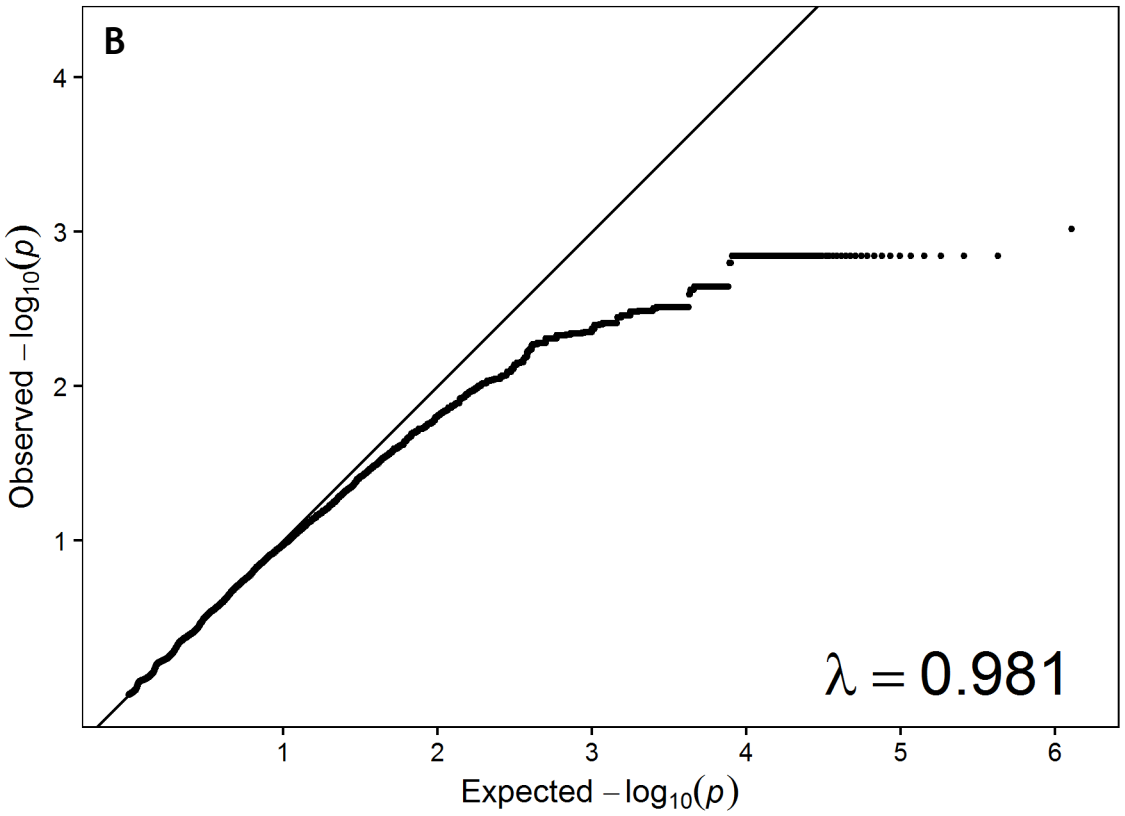


Figure 5.4 Q-Q plot for admixture mapping for genome-wide ancestry association testing of local African ancestry in A) African-American and B) Jamaican athletes and controls. Each point represents the expected *P*-value for an association with athlete status at a given locus when compared to the observed *P*-value from the association with athlete status at each locus. The genomic inflation factor λ is presented in the bottom right corner. The value for African-Americans (1.347) indicates that there may be additional population stratification that has not been controlled for, while the value for Jamaicans (0.981) indicates that there was not sufficient statistical power to detect any associations.

Manhattan plots summarising the association mapping are presented for African-Americans and Jamaicans in Figure 5.5A and Figure 5.5B, respectively. These plots are used to visualise the loci associated with athlete status, with peaks rising above the dashed line at 0.5 highlighting regions of the genome more likely to be associated with athlete status than not associate with athlete status. The admixture mapping in African-Americans identified a single significant 4.2 megabase peak at 4q13.1, with an index SNP at the intergenic SNP rs13118127. The mean local African ancestry dosage for this locus was 1.58 alleles for athletes and 1.76 alleles for controls. According to 1000 Genomes, the minor allele is present in 35.4% of the European populations analysed, but only 3.5% of the West African populations analysed (1000 Genomes Project Consortium, 2012). Despite the intergenic location of the index SNP, the significance peak contains the

Adhesion G protein-coupled receptor L3 (*ADGRL3*) gene (formerly named *LPHN3*) which encodes a member of the latrophilin subfamily of G-protein coupled membrane-receptors (Martinez *et al.*, 2011). The index SNP is also located 114 kilobases upstream from the Trans-2,3-Enoyl-CoA Reductase-Like (*TECRL*) gene which encodes for TECRL that belongs to the steroid 5-alpha reductase protein family. Admixture mapping found no significant differences between the ancestry dosages of Jamaican athletes and controls at any locus.

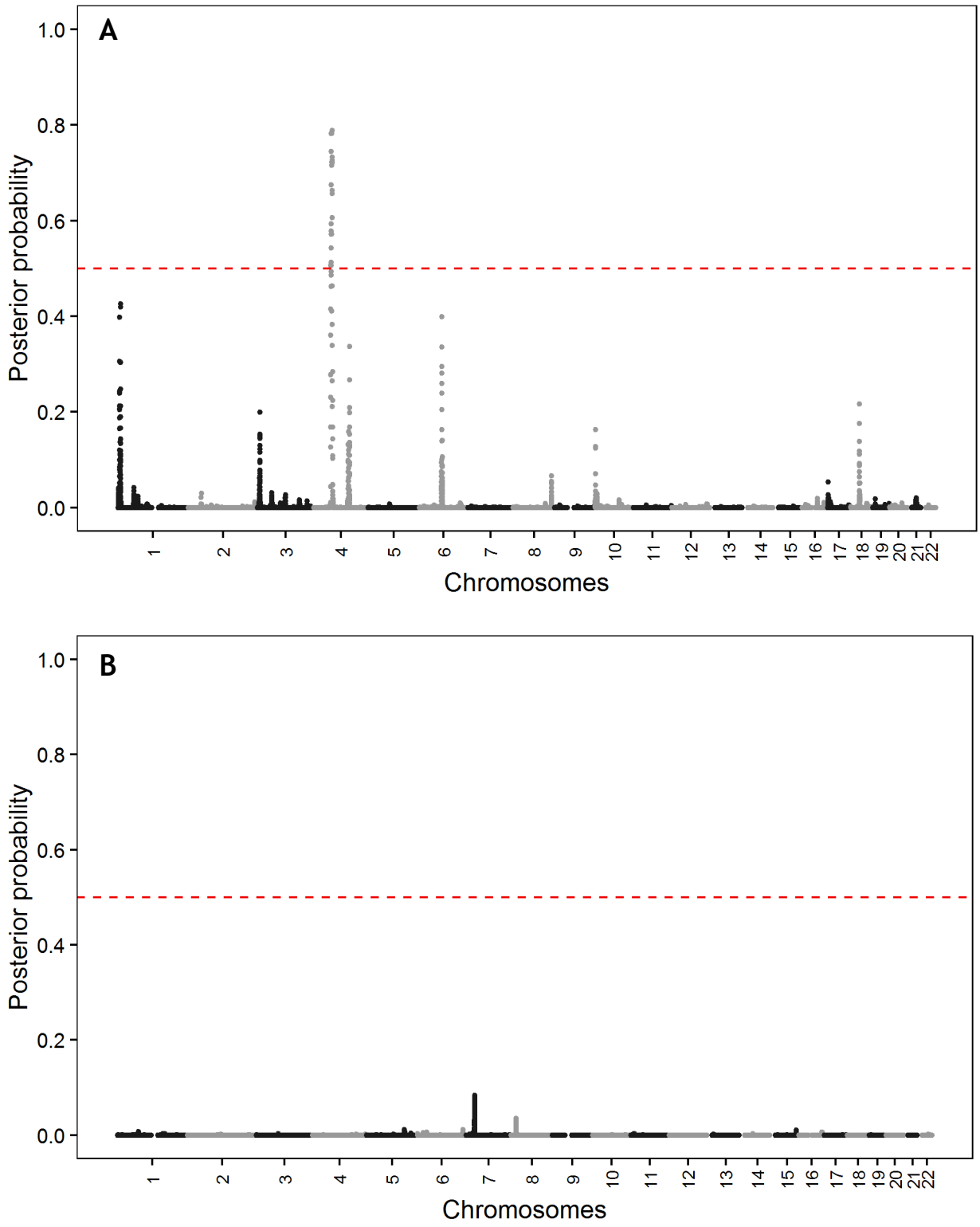


Figure 5.5 Manhattan plot depicting the posterior probabilities of admixture mapping in A) African-Americans and B) Jamaicans. Each point represents the posterior probability of an association test across the genome. Candidate admixture mapping peaks were defined as those with a probability of greater than 0.5, represented as the red dashed horizontal line.

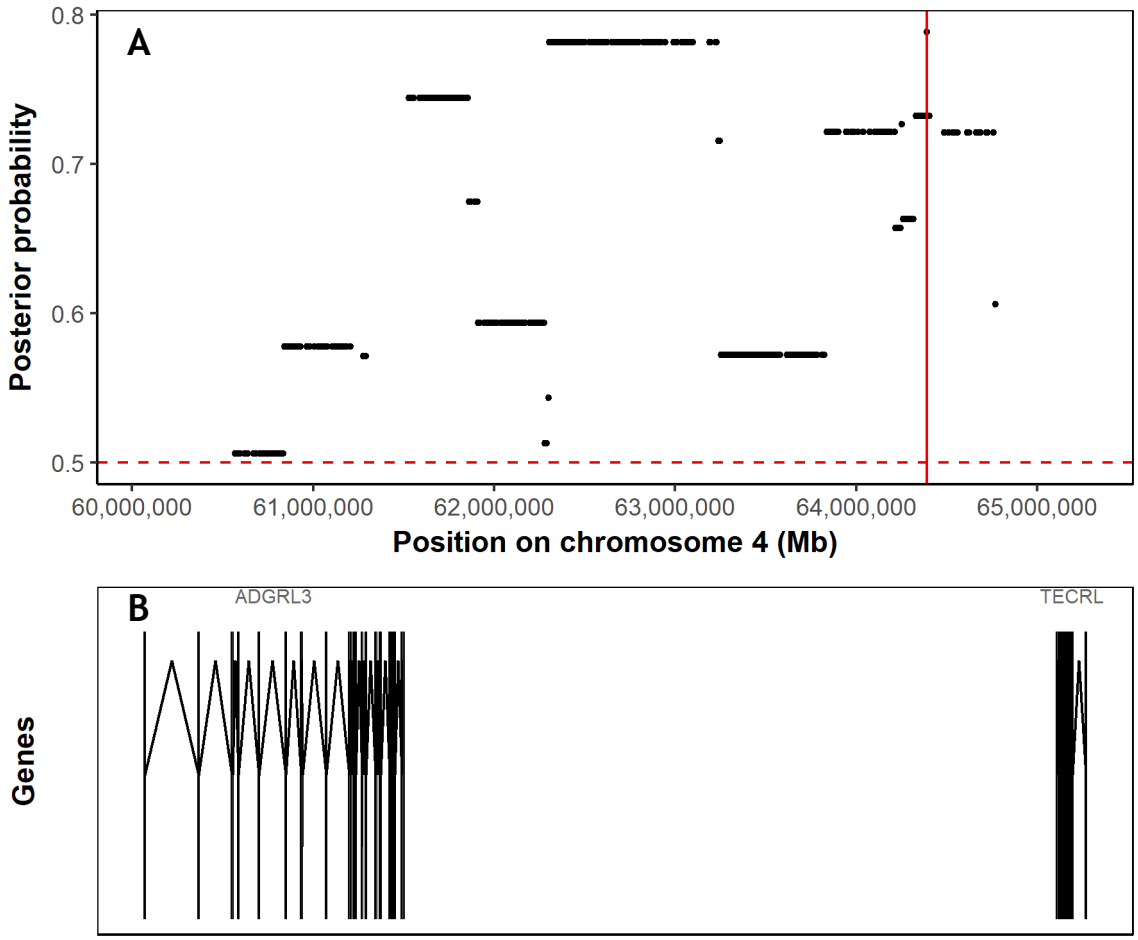


Figure 5.6 Significant admixture mapping peak at 4q13.1 in African-Americans. A) The dashed red line indicates the significance threshold while each point indicates the posterior probability of an association test. The red vertical line is the location of the index SNP rs13118127. B) The location of the two genes (*ADGRL3* and *TECRL*) shown in relation to the peak significantly associated with athlete status.

5.4 Discussion

The main findings of the investigations reported in this chapter are:

a) There were no significant differences in genome-wide African ancestry for either African-American athletes or Jamaican athletes comparted to their respective controls.
b) There were no significant differences in the number of generations since the initial admixture event in either African-American athletes and controls or Jamaican athletes and controls.
c) Admixture mapping identified a single significant genomic peak associated with athlete status in African-Americans at chromosome 4q13.1. The significant peak contained the *ADGRL3* gene.
d) There was no region of the genome that was significantly associated with athlete status in Jamaicans.

It was hypothesised that there would be significant differences in the distributions of genome-wide African ancestry and the estimated number of generations since admixture began between African-American athletes and controls. These hypotheses were rejected, and the results suggest there is no significant difference between the distribution of genome-wide African ancestry estimates or the estimated number of generations of admixture in African-American athletes and controls. It was also hypothesised that there would be no significant difference in the distributions of genome-wide African ancestry and the estimated number of generations of admixture between Jamaican athletes and controls. These hypotheses were confirmed. Additionally, it was hypothesised that regions of the genome would be associated with athlete status in African-Americans due to the overrepresentation of non-Sub-Saharan mtDNA haplogroups in African-American athletes, but not in Jamaicans due to poor statistical power in the uniparental lineage assessments. This hypothesis was also confirmed. This is the first study to show any association with local ancestry for elite athlete status.

Genome-wide ancestry estimates were calculated by comparing the allele frequencies in African-Americans and Jamaicans to the allele frequencies in presumptive source populations from Africa and Europe. These estimates were generated to compare the overall distribution of estimated African ancestry in athletes and controls. Mean genome-wide African ancestry in African-Americans

has been demonstrated to range between 73.0% and 93.0% (Bryc *et al.*, 2015). After excluding African-Americans with very little African ancestry, mean genome-wide African ancestry in African-American athletes (78.5%) and controls (81.8%) fell within the ranges reported in the literature. The Jamaican athletes (87.0%) and controls (86.2%) had very similar genome-wide mean estimates of African ancestry, both with values in the range found in the literature (76% to 94%; McKeigue *et al.*, 2000; Vergara *et al.*, 2009). There were no significant differences between the athletes and controls in either cohort, suggesting that any environmental factors influencing sprint athlete status, including access to better training facilities and coaching, are not linked to extent of African ancestry.

Eleven African-American athletes had less than 5% African ancestry, and since there were no controls with similarly extreme values, this would very likely have confounded any association for athlete status based on ancestry estimates. When these individuals are included, the result of the Mann-Whitney-U test suggested that the distributions of genome-wide African ancestry are significantly different between athletes and controls ($P = 0.022$). African-Americans are a self-reported ethnic group, and there are numerous published studies highlighting the possibility that self-reported African-Americans may have very low amounts of estimated genome-wide African ancestry (Sinha *et al.*, 2006; Yang *et al.*, 2008; Tishkoff *et al.*, 2009; Bryc *et al.*, 2010; Bryc *et al.*, 2015), though none of these works state that approximately 10% of their African-American cohort have less than 5% African ancestry as found here. African-Americans with very little African ancestry may be an artefact of hypodescent, a practice in the United States where mixed race children were not only categorised as African-American, regardless of their proportion of African ancestry, but also perceived by both European-Americans and African-Americans as belonging to the African-American ethnic group (Khanna, 2010). The African-American athlete data used in this thesis were anonymised, so athletes with predominantly European ancestry included in the original samples of buccal swabs sent from the United States may have been included because of human error.

There are many ways to estimate local ancestry dosage. Past results have shown that algorithm selection can greatly affect the outcome of admixture mapping (Chen *et al.*, 2014). Inferring local ancestry dosage using popular programs such as LAMP-LD (Baran *et al.*, 2012) and MULTIMIX (Churchhouse & Marchini, 2013)

involves grouping markers into windows and only allows ancestry switching between windows. This can be problematic if pre-specified windows do not line up with the observed ancestry switches. The admixture estimates here were calculated with ELAI, a program that uses both clustering at the source population level and clustering at the haplotype level to detect local ancestry. ELAI detects haplotypes in the source populations before comparing those haplotypes to the admixed population, freeing itself from the constraint of fixed window sizes. Combining the two types of estimates was shown to provide additional accuracy in simulated admixture data (Guan, 2014).

The number of detected ancestry switches along the genome was transformed into an estimated number of generations since admixture began. The trend in both populations shows an expected increase in the estimated number of generations since admixture as genome-wide African ancestry increases. The estimated 13.81 and 13.95 generations since admixture began in the African-American athletes and controls, respectively, falls within the six (Tian *et al.*, 2006) to 20 generations (Smith *et al.*, 2004) reported in the literature. These are the first estimates for the number of generations since admixture began for a sample of Jamaicans. These values, however, are likely to be dependent on the local ancestry algorithm used. The greater accuracy of ELAI when compared to older algorithms may mean that a greater number of switches are detected, potentially leading to estimates larger than some presented in the literature.

While only a single admixture event was assumed to have occurred in each individual's past to facilitate estimations, this assumption is unlikely to apply to each individual. Recent work on the distribution of ancestral chromosomal segment length suggests that African-Americans follow a gradual admixture model, with new contributions of genetic material from the source populations entering the admixed gene pool at a detectable rate at each generation (Jin *et al.*, 2014). This contrasts with a hybrid isolation model, used in this chapter, in which there is only a single admixture event, followed by multiple generations in isolation. Using a comparable single single event assumption, Jin *et al.* (2014) also calculated African-Americans to have undergone approximately 13 generations of admixture, after fitting an exponential distribution to the size of ancestral chromosomal segments from HapMap for African-Americans from the Southwest US (ASW). More recently, Ni *et al.* (2016) also developed a model to not only

estimate the number of generations since admixture using the distribution of ancestral chromosomal segment lengths, but also which admixture model fits the observed pattern of chromosomal ancestry segments. Using this new methodology, the authors also inferred a gradual admixture model over roughly 13 generations for African-Americans using the ASW cohort, suggesting continuous gene flow between African-Americans and people of European and African ancestry. Applying the methods described in Ni *et al.* (2016) would likely provide greater accuracy in the estimated number of generations since the initial admixture event for the African-Americans and Jamaicans in this thesis. Future work in this avenue of research should concentrate on applying model selection methodology described above to create more robust historical estimates of admixture in these populations.

The large λ inflation factor in the African-American plot (Figure 5.4A) indicates that there may be additional population stratification not considered in the association model. Many African-Americans have a non-negligible amount of Native American ancestry (Parra *et al.*, 2001; Tishkoff *et al.*, 2009; Baharian *et al.*, 2016) which may account for this pattern. There is also the possibility of unaccounted for population stratification within the African ancestral component. Previous research has shown that the African component of African-Americans can vary using estimates of genome-wide ancestry (Zakharia *et al.*, 2009; Bryc *et al.*, 2010). The Jamaican Q-Q plot (Figure 5.4B) showed fewer significant associations than would be expected by chance likely indicating inadequate statistical power. Analyses of mtDNA and NRY presented in the previous chapter hinted that the Jamaican athletes were ancestrally similar to the Jamaican controls, though there was not sufficient statistical power to conclude this for certain. This plot Q-Q plot, along with the admixture mapping analyses, may suggest that the Jamaican people may be more genetically homogeneous when compared to African-Americans, though without sufficient statistical power it is not possible to state this conclusively.

Athlete status was significantly associated with an overrepresentation of European ancestry at chromosome 4q13.1 in African-Americans, a genomic region a previous GWAS with the same cohort failed to associate with athlete status (Wang, 2013). That work did not find any SNP association for athlete status in African-Americans below the 5×10^{-8} threshold for significance. Different mapping methodologies

may account for this discrepancy. While association mapping tests for very localised associations, admixture mapping tests for broader associations over blocks of ancestry in linkage disequilibrium. Nevertheless, of the SNPs reported in Wang (2013) with a *P*-value below 5×10^{-5}, none were located on chromosome 4. This suggests that there may be differences between the overall designs of the experiments that variation in mapping methodology cannot readily explain.

The described peak significantly associated with athlete status was in chromosome 4q13.1 and contains the *ADGRL3* gene that encodes for latrophilin 3, a protein associated with both cell adhesion (Boucard *et al.*, 2014) and signal transduction (O'Sullivan *et al.*, 2012). Although there are known functional variants (Orsini *et al.*, 2016), *ADGRL3* shows its highest median expression in the basal ganglia and other regions of the brain in a sample of 53 tissues from 570 donors (Lonsdale *et al.*, 2013). Neonatal knockout mice created with an inactivated version of this gene showed elevated expression of genes involved in dopamine and serotonin transport, while adult knockout mice were hyperactive in an open-field test and were more sensitive to the locomotor stimulant effects of cocaine than wild-type mice (Wallis *et al.*, 2012). *ADGRL3* variants have also been associated with a predisposition to attention deficit hyperactivity disorder and sensitivity to stimulant medications in humans (Arcos-Burgos & Muenke, 2010; Acosta *et al.*, 2016).

This is the first reported association between the *ADGRL3* gene and elite sprint athlete status. Sprint running requires high levels of coordination and neuromuscular adaption (Majumdar & Robergs, 2011a), and this gene's role in the brain may influence this. Alternatively, the involvement of the gene with dopamine and serotonin transport may influence the perception of fatigue in voluntary muscle efforts (Meeusen *et al.*, 2006). However, these explanations are purely speculative, and replication with another cohort of elite sprinters or additional association tests for *ADGRL3* variants and the physiological factors known to influence sprint running will be necessary before any solid conclusions between the *ADGRL3* gene and sprint athlete status can be made.

Despite the 4.2 Mb peak containing the *ADGRL3* gene, it is possible, though unlikely due to linkage disequilibrium in the dosage blocks, that the variant responsible for the association with athlete status is not found in the peak at all.

The *TECRL* gene is located 114 kilobases downstream from the index SNP rs13118127 and has been associated with the regulation of lipid metabolism by peroxisome proliferator-activated receptor alpha (Lee *et al.*, 2003). Loci within this gene have been associated with post-traumatic stress disorder (Xie *et al.*, 2013) and ocular axial length (Cheng *et al.*, 2013) and the rare paediatric Kawasaki disease (Lee *et al.*, 2012b). *TECRL* is part of the steroid 5-alpha reductase family of genes, a gene family associated with the conversion of testosterone into the more potent androgen dihydrotestosterone (Cantagrel *et al.*, 2010), although the importance of circulating dihydrotestosterone on sprint performance is unclear (Smith *et al.*, 2013).

There were no significant peaks found with admixture mapping in Jamaican athletes and controls, though there was an evident lack of statistical power. These results are in line with those from the previous chapter, suggesting a more general dearth of statistical power when comparing Jamaican athletes and controls. Any difference in the genetics of Jamaican athletes and controls is likely to have an effect size too small to be detectable with the sample size and methodology used in this or the previous chapter.

It is possible to combine the local ancestry dosages with the genotype data directly to create a more statistically robust mapping method (Seldin *et al.*, 2011). These joint tests not only increase the power to detect phenotype associated loci, but also increase resolution to help further localise any signals by stratifying association models by local ancestry at each locus (Shriner, 2013). This has been accomplished investigating type 2 diabetes mellitus (Shriner *et al.*, 2011), breast cancer (Pasaniuc *et al.*, 2011) and blood pressure (Chen *et al.*, 2014). Despite using the first two steps of the six-step joint ancestry association test BMIX (Shriner *et al.*, 2011) for this current study, modifying the underlying model framework to a case-control binary outcome resulted in making the genotype association stratified by local ancestry intractable. To utilise this more powerful method, future work should concentrate on refining the joint methodology to accommodate the case-control experimental framework.

Admixture mapping has several limitations. The power to detect an association with local ancestry is reliant on the assumption that no directional selection has been acting on the admixed populations (Seldin *et al.*, 2011). While some have

suggested that there has indeed been directional selection acting on the genomes of African-Americans since admixture (Bhatia *et al.*, 2011; Jin *et al.*, 2012), a recent comprehensive study has discounted any directional selection since the process of admixture began because the authors either did not account for long-range linkage disequilibrium, used inappropriate statistics to measure genetic distance or set an appropriate threshold of significance (Bhatia *et al.*, 2014). No such work has been completed for Jamaicans, although given their different population histories as highlighted earlier in this thesis, the same assumptions may not apply. Admixture mapping also relies strongly on differences in allele frequencies of any causative variants between cases and controls. For example, the detection of any causative variants for sprint performance must differ between Europeans and Africans for admixture mapping to detect any association. Finally, this methodology can only guide in further hypothesis generation and cannot finely map variants associated with a phenotype.

Lastly, for admixture mapping to reliably detect an association between ancestry and athlete status, the athletes should be properly matched to controls. In general, the source population from which controls are sampled should be that from which athletes are also sampled (Cardon & Palmer, 2003). These criteria were violated, however, when it was hypothesised that there would be evidence of differences in the distributions of genome-wide African ancestry and the estimated number of generations since admixture began between African-American athletes and controls. While there were no significant differences found between the athletes and controls in either genome-wide ancestry or the estimated number of generation since admixture began, future studies should take appropriate case-control matching into account when designing future admixture mapping studies.

In conclusion, there were no significant differences between either the genome-wide African ancestry estimates or the estimated number of generations since admixture began in either African-American athletes and controls or Jamaican athletes and controls. Contrary to what was hypothesised, this suggests differences in mitochondrial haplogroup frequencies between African-American athletes and controls are not linked to environmental benefits in athlete development. The lack of significant difference between Jamaican athletes and controls was expected given the poor statistical power in the comparisons

between athletes and controls stated in the previous chapter. Admixture mapping revealed a significant peak in the African-American athletes and controls at 4q13.1. This suggests that the *ADGRL3* gene or neighbouring gene *TECRL* may be associated with athlete status in African-Americans, though this is unlikely to the stratification evident in the African-Americans. Further fine mapping and subsequent replication of these results is necessary before any definitive conclusions can be made about this presumptive gene association though. A similar peak was not found in the admixture mapping of Jamaican athletes and controls, though there was not sufficient statistical power to make any definitive conclusions and any difference between athletes and controls groups may have an effect size too small to detect with the current methodology.

6

General Discussion

The overall aim of this thesis was to investigate the influence that individual genetic ancestry may have on elite sprint athlete status in African-Americans and Jamaicans. To explore the overrepresentation of African-Americans and Jamaicans in sprint athletics, several research questions were addressed:

a) What are the likely sources in Africa of the Jamaican population?
b) How might past demographic pressures have shaped modern Jamaicans?
c) Do the distributions of maternal and paternal lineages of African-American and Jamaican athletes differ from matched controls?
d) Do the distributions of maternal and paternal lineages of African-American and Jamaican athletes differ from one another?
e) Do the distributions of maternal and paternal lineages of African-American and Jamaican controls differ from one another?
f) Does the proportion of genome-wide African ancestry differ significantly between African-American athletes and controls and Jamaican athletes and controls?
g) Are the estimated number of generations since admixture began significantly different between African-American athletes and controls and Jamaican athletes and controls?
h) Are regions of the genome enriched for either African or European ancestry in either African-American athletes and controls or Jamaican athletes and controls?

The main findings of this thesis are:

a) Modern Jamaicans are predominantly descended from slaves originating from the Gold Coast despite large influxes of slaves from the Bight of Biafra and West-central Africa before the end of the slave trade.
b) There appears to have been selective pressure acting on the slave population of Jamaica.
c) The maternal lineages of African-American athletes are significantly different from African-American controls. There are not significant differences between the paternal lineages of African-American athletes and controls, maternal lineages in Jamaican athletes and controls, or paternal lineages in Jamaican athletes and controls.
d) The maternal and paternal lineages of African-American athletes and Jamaican athletes are significantly different.
e) The maternal and paternal lineages of African-American controls and Jamaican controls are also significantly different.
f) The proportions of genome-wide African ancestry do not differ significantly between either African-American athletes and controls or Jamaican athletes and controls.
g) The estimated number of generations since admixture occurred does not differ significantly between either African-American athletes and controls or Jamaican athletes and controls.
h) An enrichment of European ancestry at 4q13.1 is significantly associated with athlete status in African-Americans. There are no significant loci associated with athlete status in Jamaicans.

The West African diaspora has been significantly overrepresented in international sprint athletics since the 1930s, yet despite the connection between the Trans-Atlantic slave trade and the ethnicity of elite sprint athletes, comparatively little research has been conducted to investigate this phenomenon. This thesis presents several novel investigations:

- This is the first study to look at the elite athlete status in the context of the Trans-Atlantic slave trade.
- This is the first study to examine the likely origins of modern Jamaicans.

- This is the first study to compare the maternal, paternal and autosomal ancestries of African-American and Jamaican athletes to their respective controls.
- This is the first study to make inferences about and compare the estimated number of generations since the initial admixture event in African-American and Jamaican athletes and their respective controls.
- This is the first study to use admixture mapping for elite athlete status.

6.1 Selection for athletic performance because of the slave trade

The results of Chapter 3 and elsewhere (see Section 1.6 above) appear to suggest that slavery may have produced detectable signals of selection pressure. Chattel slavery in the Western Hemisphere was characterised by the coerced relocation of millions of Africans to a new disease environment to undertake a torturous amount of labour, resulting in very high levels of slave mortality. The following section will present both sides of the argument that such selection resulted in the overrepresentation of the West African diaspora in elite sprint athletics.

As presented in Chapter 3, there does not appear to be evidence that slavery acted to actively select for elite sprint running phenotypes such as muscular strength or power output. While the selection mechanisms put forward in the hypotheses posited by Morrison and Cooper (2006) and Aiken (2011) are debatable (Section 1.6), selection driven by slavery resulting in the overrepresentation of elite sprint athletes in slave descendant populations is still a possibility. While there is unlikely to have been selection acting directly on the physiological traits associated with sprint running, that does not necessarily mean that selection did not act indirectly on those physiological traits. In an environment with elevated levels of mortality from infancy to adulthood, there may have been selective pressure to improve resistance to some of the mortality risks in the environment by either improving developmental stability or reducing susceptibility to tropical diseases.

The association of a single gene or variant with multiple phenotypes is known as pleiotropy. The *HBB* gene, which may influence elite sprint running, is a commonly cited example of pleiotropy effecting African-Americans and Jamaicans. While the sickling of red blood cells associated with the *HbA*/*HbS* genotype has been shown

to provide resistance to falciparum malaria, the *HbS*/*HbS* genotype causes a sickling in a large proportion of red blood cells, leading to complications commonly known as sickle cell disease (Lopez *et al.*, 2010). These sickled blood cells are responsible for a wide range of health problems: fibrosis of the spleen due to the accumulation of red blood cells; damage to the brain, heart, lungs and kidneys because of sickled cells clumping and causing interference with blood circulation; and anaemia, which results from the rapid destruction of sickled cells leaving too few blood cells in the blood stream to carry adequate amounts of haemoglobin (Ballas, 2011).

Despite the reduced functionality of the protective genotype at circulating haemoglobin (Chirico *et al.*, 2016), the *HbA*/*HbS* genotype has also been linked to greater surface area of the type IIx muscle fibre, though the mechanisms underlying this association are not known (Vincent *et al.*, 2010). As hypothesised by Morrison and Cooper (2006), it may be possible that the mechanisms evolved to provide resistance to malaria in the African ancestors of African-Americans and Jamaicans may have coincided with an increased frequency of variants responsible for sprint running. However, considering the broad geographic range association with the sickle cell trait (Gething *et al.*, 2011), this theory does not explain the overrepresentation of the West African diaspora in sprint athletics when compared to other regions where the sickle cell trait is prominent such as North Africa and South Asia.

Slaves born into slavery in North America or the Caribbean may have developed genetic mechanisms to resist pathogens that had pleiotropic effects for physiological traits associated with elite sprint running. However, there is the very unlikely possibility that instead of selection for disease resistance working as a pleiotropic effect enriching the population for these physiological traits, the relaxed selection pressure over the last hundred years may have increased the number of deleterious mutations, increasing the overall susceptibility to diseases. Relaxed selection acting on these deleterious alleles may be in linkage with alleles responsible for the physiological traits associated with elite sprint running. It would be difficult to test for a pleiotropic effect in a sample of elite sprint athletes though because possibly localised and individual adaptations to slavery. Finally, any adaptations would have needed to carry on through generations of gene flow

between newly arrived Africans and Europeans without diluting the effect of the pleiotropically-selected alleles.

Pleiotropic effects are likely to be rare variants that have large effect sizes. Under a mutation-selection balance model, however, it may have been possible for different novel rare variants to have the same outcome. The mutation-selection balance model is used explain the number of new mutations created as deleterious alleles are eliminated by selection (Lynch, 2016). Since selection prunes deleterious alleles, these new rare variants are likely to have originated much more recently than common variants. Many rare alleles with modest effect sizes may influence sprint performance. It is possible that an individual athlete may only have a unique subset of those alleles present in the entire population.

Positive selection is traditionally thought to decrease genetic diversity by increasing the frequency of an advantageous allele. Many generations of admixture, on the contrary, would have increased genetic diversity. The observed advantage in elite sprint running may be the result of increased levels of heterozygosity because of the introgression of European genes into the African genomes in addition to the admixture between many previously separated African sub-populations. This increased genetic diversity may influence variants with large effect sizes for the physiological traits associated with elite sprint running. There is evidence that selection can act in concert with increased diversity through selecting multiple genetic pathways to encode the same phenotype. This has been shown in Africa (Ingram *et al.*, 2009) and in admixed Brazilians (Friedrich *et al.*, 2012) for lactase persistence. Future work addressing the levels of heterozygosity in athletes and controls may prove insightful to identify variants under selection.

Despite the apparent evidence of selection pressures caused by slavery, it has already been argued that there does not appear to be evidence of selection for physiological traits that would be advantageous to sprint performance. The results of Chapter 3 suggest that Jamaicans are more likely to originate from the Gold Coast as opposed to the Bight of Biafra and West-central Africa. Chapters 4 and 5 suggest that there is little evidence that the genetic constitution of Jamaican athletes differs significantly from Jamaican controls. Using transitive logic, these results suggest that the matrilineal origin of most Jamaican athletes is from the Gold Coast. This means that, contrary to hypotheses posited by Morrison and

Cooper (2006) and Aiken (2011) and introduced in Section 1.6, there does not seem to be evidence that the environmental selective pressures of the Middle Passage and chattel slavery were strong enough to select for traits associated with elite sprint performance, as modern Jamaican elite-athletes are descended from those earlier on in the slave trade, rather than those enduring the worst conditions later on in the slave trade. While both hypotheses are supported with observations and previous research, the connections both hypotheses make to the Trans-Atlantic slave trade are tenuous.

Despite the difficult environment and high levels of slave mortality experienced in North America and the Caribbean, slaves in both regions were able to reproduce, though only in North America did the slave population reproduce above replacement. This begs the question: what selective pressures were influencing levels of slave fecundity? Historical evidence suggests that slave women in lighter duties had more children, suggesting that domestics were more likely to have children when compared to women working in the fields. In Jamaica, there is also evidence that women on non-sugar cane plantations e.g. livestock pens, had more children compared with women working on the sugar plantations (Higman, 1976). There is also evidence that the difficulties of work can explain the differences in fecundity. In a comparison between the sugar plantations in the Caribbean and Louisiana, Tadman (2000) noted very similar levels of mortality and reproduction despite the slaves in Louisiana largely arriving acclimatised to slavery from the Upper South. Similarly, Craton (1978) noted that slaves on cotton plantations in the Bahamas reproduced at similar levels to plantations in North America when compared to slaves on sugar plantations in Jamaica. Female slaves assigned lighter duties would have been able to devote a greater amount of their resources to child bearing than slaves tasked with more demanding work in sugar cane fields. This may have led directly to improved fertility in those slaves.

If the environmental and selection pressures on slave populations in Jamaica were as strong as Morrison and Cooper (2006) and Aiken (2011) suggested, one would have expected to see a distribution of mtDNA haplogroup profiles like those initially hypothesised in Chapter 3. The strong mortality associated with sugar cane production ensured that an influx of slaves from Africa was always needed. If the high levels of mortality were coupled with a slave owner's coercion of the reproduction of his slaves in the eugenic effects suggested by Morrison and Cooper

(2006) and those in the popular media (Kane, 1971; Entine, 2000; Hughey, 2014), one would expect to see a very distinctive signal from the newly arrived Africans just as the slave trade ended. The positive selection associated with the eugenically selected slaves would have had to outweigh the newly arrived Africans by such an enormous amount that the genetic effects that now manifest themselves as sprint running ability would need to have reached fixation within the entire population of modern Jamaicans despite the removal of selective breeding at the end of slavery for this to be the case.

Finally, the idea of systematic selective slave breeding by slave owners for traits presently associated with sprint performance is in direct conflict with the patrilineal evidence presented in Chapter 4 and elsewhere (Simms *et al.*, 2012; Torres *et al.*, 2012) showing relatively high levels of European paternal ancestry. Were selective breeding to have been the basis of the elite sprint phenotype, one would expect entirely African paternal ancestry in athletes. Despite most European settlers abandoning Jamaica after slavery collapsed in 1834 (Sewell, 1968), the European component of the paternal lineage pool in Jamaica is roughly 41.1% (95% CI [20.8-61.4%]; Torres et al., 2012). Historic evidence suggests many European men in Jamaica took slaves as concubines during the slave era (Burnard, 2006). If selective breeding were to take place in Jamaica, the gene flow between Europeans and slaves during the slave era would certainly have negated any population-level effect.

6.2 Limitations

Despite efforts to mitigate their effects, the research presented in this thesis may have been hindered by numerous limitations.

6.2.1 Selection of truly representative control samples

As a consequence of the varying amounts of genetic ancestry, case-control studies in admixed populations must carefully select representative control populations to avoid arriving at spurious conclusions as a result of type I error. African-American individuals self-identify as belonging to the African-American community. Because of varying levels of admixture, their amount of African and

European genetic ancestry can vary dramatically, from individuals inferred to have entirely African ancestry to individuals inferred to have entirely European ancestry. This is likely the result of the one-drop rule, a rule unique to the United States in which a single African ancestor was enough for an individual to be categorised as African-American. Despite the relative cultural homogeneity of Jamaica, individual admixture is still distributed widely around the mean. This variability in the levels of ancestry has the potential to confound any case-control analyses if it is not accounted for.

Controlling for population stratification such as continent-level ancestry in case-controls studies, as done in the admixture mapping presented in Chapter 5, unfortunately may not be enough to control for confounding effects. Naïve procedures designed to capture the genetic variation present in each sample using dimensionality-reducing techniques such as principal component analysis or clustering techniques like those used by the programs STRUCTURE (Pritchard *et al.*, 2000) and EIGENSTRAT (Price *et al.*, 2006) work well to identify population stratification even when the origins of the admixed cohort are unknown. Future work should use these more robust naïve estimates of population stratification over inferred continental ancestry to thoroughly control for spurious associations.

From the early 18th century, slaves in the United States began to reproduce above the level of replacement. The resulting surplus in labour led to a waning of demand for the Trans-Atlantic trade, and demand for slaves from territories further west including not only Tennessee and Kentucky, but also Louisiana and Texas were filled by slaves from the Upper South in Maryland and Virginia. While the resulting increased homogeneity would have had the effect of muting any regional variability in the African ancestral component of the descendants of these slaves, levels of admixture have still been shown to differ between regions. The rate of gene flow into slaves was largely dependent on the density of European and Native American groups in the area, each with their own geographically-specific set of genotypes. After emancipation, African-Americans then began migrating across the country in search of work in the Northeast, Midwest and California. This is referred to as the Great Migration. Although there is genetic evidence of regional associations between their regions of departure and destination (Baharian *et al.*, 2016), this movement of people further complicates the notion of a single African-American macro-ethnic group.

The resulting mosaic of differential admixture between a wide variety of groups has led to the identification of micro-ethnic groups within African-Americans. These groups show enough variability that the risk of certain non-communicable diseases has been shown to vary regionally (Jackson, 2008b). Links between purported source populations in Africa and micro-ethnic African-Americans have been demonstrated for the risk of breast cancer (Jackson, 2008a). No attempt was made to test for the presence of an overrepresentation of micro-ethnic groups present in the elite African-American sprint athletes or controls, although the failure to account for population stratification at a greater resolution may have led to the pattern observed in the admixture mapping Q-Q plots for African-Americans (Figure 5.4).

6.2.2 Statistical and analytical limitations

Numerous studies in this thesis were hampered by poor statistical power, severely limiting the case-control comparisons presented. Statistical power is the probability that a false null hypothesis will be rejected, and is essentially the likelihood that a study will detect an effect when there is an effect there to be detected. The poor statistical power in Chapters 4 and 5 means that it is not possible to say with the current results whether a difference exists between the athletes and controls.

Tying in with the previous point of statistical power, another limitation of this thesis involved the characterisation of elite sprint performance and hence, elite sprint athlete status. To draw more statistically powerful and meaningful inferences within the case-control framework, a greater number of athletes would need to be sampled. If one assumes that the pool of elite athletes is fixed, this introduces a theoretical paradox. The term "elite" is comparative and shifting the inclusion threshold for elite downward to increase the number of elite athletes would require broadening the very definition of elite. While likely improving statistical power, this may cause any potential genetic signals of truly phenomenal athletes to become less apparent or masked entirely. By studying high-calibre athletes from populations overrepresented in sprint athletics, it was hoped that any genetic signatures would be distilled to their most pronounced.

Issues regarding statistical power can be improved in three other ways. Firstly, the pool of populations can be broadened. As mentioned in the introduction, the West African diaspora is overrepresented in elite sprint performance, and although only African-Americans and Jamaicans were researched in this thesis, there are many more nations represented on the world stage. Large nations including Great Britain and Canada and small nations such as Grenada and St. Kitts and Nevis have all produced world champions in sprint athletics from the West African diaspora. Sampling from this expanded pool of athletes while controlling for confounders such as genome-wide admixture would increase statistical power, though the number of additional samples would depend on the observed genetic variance and the inclusion of a more diverse set of athletes may introduce additional levels of population stratification.

Conversely, a more protracted and logistically more complex sampling period would be necessary to maximise the number of athletes sampled. As with all physically demanding sports, sprint athletes will only be competitive for a finite amount of time. Since the genetic make-up of these athletes will remain unchanged after retirement, the most likely route for broader sampling of the best sprint athletes would be by sampling elite athletes after their careers in athletics had ended.

Finally, coding athlete status on a continuum in a cases-only framework may also increase statistical power since there is no statistical noise introduced by controls (Balding, 2006). To normalise differences between men and women and account for different disciplines, personal bests can be converted into performance scores objectively using the IAAF Scoring Table for Athletics (Spiriev, 2011). A generalised linear model could then be constructed to regress these performance scores against genotype. Additional covariates including the year the personal best was attained may also help control for the influence of improved technology on the time attained. Only 56 African-American athletes and 91 Jamaican athletes had accompanying personal best data in this thesis and undoubtedly the personal bests and genotype data of many more athletes would be needed to detect any significant effect.

The degree to which an inherited allele is expressed in a phenotype in a population is known as penetrance. While certain genotypes have high penetrance, resulting

in a near perfect relationship between genotype and phenotype, the alleles associated with sprint athlete status are likely to have low penetrance, meaning that the alleles are much more susceptible to environmental factors. Nevertheless, any genetic association with athlete status is predicated on the assumption that athletes and controls can be delineated genetically by the causative element of elite sprint athlete status. As presented in Chapter 4 and Chapter 5, there is no evidence that elite Jamaican sprint athletes differ significantly from Jamaican controls. Despite an individual having the ideal genetic make-up for elite sprint running, many socioeconomic, behavioural or biological reasons may be keeping them from either reaching their potential or taking up the sport to begin with. Incorporating environmental factors into a case-control model can help distinguish between the effects of genotype and the environment. However, adding these additional interactions does require a substantially larger sample size because of the large number of comparisons. For example, adding an interaction for a single genotype with a dichotomous environmental variable would entail twice as many tests.

6.2.3 Sprint athlete confounders

An important assumption critical to the associations with athlete status tested in this thesis was that elite sprinters are homogeneous sub-populations within their respective populations. Despite the apparent simplicity in covering a fixed amount of distance in the shortest possible time, sprinting is a multifactorial task with many potential combinations of genetic, biomechanical and environmental factors (Sections 1.2, 1.3 and 1.4). It is very likely there are different combinations of factors responsible for success in sprint athletics at the individual level. These combinations may lead to additional stratification that needs to be accounted for, including characteristics such as gender, but also morphological characteristics such as height. Future work should attempt to control for as many confounders as possible to better understand the aetiology of elite sprint performance in humans.

When compared to other sports, athletics is not a financially lucrative sport for many its participants, even at the professional level. Naturally gifted sprint athletes may pursue other sports from a young age where possible, or simply give up before their potential within athletics is realised due to other time

commitments, financial limitations, access to training facilities or injury. In the United States, many male Olympic-calibre sprint athletes come from a background in American football (Pitsiladis *et al.*, 2011). While a surprising number can compete to a high level in both sports (e.g. 1964 double gold medallist and NFL Hall of Fame wide receiver Bob Hayes) it is more likely that top male sprint athletes would concentrate on the more lucrative sports. Elite female sprint athletes, however, are unlikely to be drawn away from athletics because there is not a comparably lucrative sport for which elite female sprinter could be drawn away toward during their development. Future research investigating the genetics of elite sprint athletes could control for this effect by adding sex as an explanatory factor.

Athletics as a sport requires heavy training loads, repetitive training sessions and time-intensive supplementary weight training. From a psychological viewpoint, athletics requires highly motivated athletes who are capable of practising both individually and within a group and of leading an appropriate lifestyle with optimal recovery and nutrition (Dosil, 2006). Elite junior athletes may also fail to make the transition to senior level competition due to competing demands from their social, academic or professional lives (Hollings *et al.*, 2014). While top-level international junior athletes were more likely to remain in elite sport as a senior athlete than those at a lower standard of performance, there was not a perfect correlation (Bussmann *et al.*, 1994). It may be a case that even the athletes with high performance potential do not stay in the sport because of academic, social or professional pressures. The implications of this are that individuals may be considered a control when in fact they have the genetic profile of an athlete.

6.3 Recommendations for further research

There are several promising directions for future research into the genetics of elite sprint running performance, and the limits of human performance more generally. However, all future research in this area should put a concerted effort into improving sample sizes. This is the most important improvement that can be made to better increase the statistical power and reproducibility of association studies. Without increasing sample size, even the most promising results are likely to be false positives.

To determine the genetic factors responsible for elite sprint running, a broad shift in study design may be necessary. Current GWAS test for association between a phenotype and common SNPs. Common SNPs have been shown to be unable to explain extreme performance phenotypes (Bouchard, 2015). Common SNPs have a wider geographic distribution and have been exposed to selection pressures for thousands of generations while rare variants have arisen relatively recently (Slatkin, 2000; Frazer *et al.*, 2009). Because they only affect a small proportion of the sample population, significant associations with rare variants indicate that the effect is large enough to overcome infrequency in cases when compared to controls. Were rare variants to have a small effect, a prohibitively large sample size would be needed to find a significant association after controlling for multiple testing. It may well be that a source of elite performance lies in these rare variants, although improved genotyping resolution and sample sizes would be needed to assess whether rare variants are critical to explaining the genetics of elite sprint performance.

Testing for the physiological measurements associated with elite sprint running instead of simply for elite athlete status is also vital. Elite athlete status is a multifaceted phenotype and there may be many different pathways to elite performance, and thus, any associations found would likely have limited generalisability. By shifting to testing for physiological traits associated with elite sprint running such as muscle activation and joint power, it may be possible to narrow down a portion of the biological pathways responsible for elite sprint running. Because of the complexity of sprint running, however, such a shift will only be viable in the context of a very large sample sizes.

Elite athletic performance is a good example of a genotype × environment interaction. An approach that may prove insightful in determining the environmental and lifestyle factors underpinning elite sprint performance would be to perform an environmental-wide association analysis. It is possible to search for association with unique environmental variables in a hypothesis-free manner using methodology analogous to a GWAS. This method has been successful in determining the environmental underpinning of type 2 diabetes mellitus (Patel *et al.*, 2010) and metabolic syndrome (Lind *et al.*, 2013). Introducing environmental interaction into a genetic association model does require a substantially larger

sample size because of the large number of comparisons made. Increasing the number of athletes has its own limitations, as discussed above.

A shift in the traditional methods for investigating the genomes of elite athletes also needs to occur if there is to be any further insight into the genetic mechanisms behind elite performance. Because of limited sample sizes, commercially available genotyping arrays containing common SNPs have been unable to explain extreme performance phenotypes (Bouchard, 2015). It is reasonable to hypothesise that the rare and extreme phenotypes of elite athletes can be explained by rare functional variations or mutations that emerged and were selected rather recently in human history, so that these SNPs or mutations are detectable only in certain regions, although a much larger sample size than currently available would need to be genotyped. Future work in understanding the genetic basis for human performance should concentrate not only on broadening the experimental techniques used to test association with elite phenotypes, but also focus on association tests free from underlying assumptions for biological candidate loci. Combined genomic, transcriptomic, metabolomic, proteomic, and epigenomic solutions with much larger sample sizes have the potential to greatly facilitate discovery of the genetic influences on sporting performance, training response, injury predisposition, and other potential determinants of successful human performance. In an attempt to enhance collaboration and replication among research groups at the technological frontier of elite athletic performance research, the aim of the recently announced Athlome Consortium is to combine resources from individual studies and consortia worldwide to collectively study the genotype and phenotype data available on elite athletes, the adaptation to exercise training (in both human and animal models), and the determinants of exercise-related musculoskeletal injuries (Pitsiladis *et al.*, 2016). Additionally, by releasing their data to the public, the consortium will likely accelerate the understanding of the biology behind not only elite sprint performance, but also health-related fitness.

6.4 Conclusion

In conclusion, the main findings of this thesis were:

a) Modern Jamaicans are mostly descended from slaves originating from the Gold Coast of Africa, despite large influxes of slaves from the Bight of Biafra and West-central Africa before the end of the slave trade.
b) There appears to have been selective pressure acting on the slave population of Jamaica. Differences between the presumptive origins and the stable population model suggested varying levels of mortality and fecundity within the slave population, consistent with earlier ethnographic and linguistic studies.
c) The maternal lineage distributions in the African-American athletes were significantly different from that of African-American controls while there were no significant differences in maternal lineage distributions between Jamaican athletes and controls. There was insufficient statistical power to infer any differences between the paternal lineages of African-American athletes and controls or the Jamaican athletes and controls. This suggests that maternal ancestry may be a factor in elite sprint athlete status for African-Americans, while Jamaican maternal lineages are homogeneous with regards to elite sprint athlete status. There was insufficient statistical power to arrive at a similar conclusion regarding the paternal lineages.
d) The maternal lineages of African-American athletes and Jamaican athletes were significantly different, although there was insufficient statistical power to determine if there were any differences between the paternal lineages of African-American athletes and Jamaican athletes. This suggests that the same maternal lineage distribution is not associated with sprint athlete status in the two populations, while there is insufficient evidence to make a similar claim regarding paternal lineages.
e) The maternal lineages of African-American controls and Jamaican controls were also significantly different, although there was insufficient statistical power to conclude whether significant difference exists in the paternal lineages of African-American controls and Jamaican controls. These results suggest that there is some evidence that the population histories of African-Americans and Jamaicans are significantly different despite the lack of evidence from the paternal lineages.

f) The proportion of genome-wide African ancestry did not differ significantly between either African-American athletes and controls or Jamaican athletes and controls. This suggests that environmental factors typically associated with higher levels of African ancestry in these populations (e.g. lower socioeconomic status, diminished access to healthcare) are not directly linked with elite athlete status.

g) The estimated number of generations since admixture occurred did not differ significantly between athletes and controls for either African-Americans or Jamaicans. This suggests that athletes were not more likely than controls to have had European ancestors in the recent past, thereby having greater access to resources.

h) Admixture mapping was used to detect an enrichment of European ancestry at chromosome 4q13.1 that is significantly associated with athlete status in African-Americans. There were no significant loci associated with athlete status in Jamaicans. Although there was insufficient statistical power to draw any significant conclusions from the Jamaican data, this suggests that the regions of the genome influencing sprint athlete status may be different in the two populations.

7

Appendix

7.1 HVS-1 SNP motifs of the Jamaican individuals analysed in 3.1.2

This table shows the mitochondrial SNP motifs from HVS-1 used to generate mtDNA haplogroup data in the third and fourth chapter of this thesis. Motifs are presented as differences from the revised Cambridge reference sequence, subtracted by 16000. Typographical notes indicate other supplementary phylogenetic data generated for a selection of samples through either RLFP or micro array (described in Section 5.2.1).

* Sample typed at positions 10398, 10400 and 13803 by RFLP
† Sample types at 26 coding region mtSNPs on the Illumina Omni1-Quad BeadChip technology (Illumina, San Diego CA, USA)
‡ Sample types at 135 coding region mtSNPs on the Illumina Human 660W-Quad BeadChip technology (Illumina, San Diego CA, USA)

Sample ID	HVS-1 motif	Haplogroup
JAM001	093 223 264 278 390	L2c2
JAM002	223 264 278 390	L2c2
JAM003	124 189 223 278 311 362	L3b2
JAM004*	183C 189 278 318 390	L2c
JAM005	223 278 294 309 319 390	L2a1
JAM006	189 223 278 294 309 390	L2a1
JAM007	223 278 294 309 390	L2a1
JAM008	223 278 294 309 319 390	L2a1
JAM009	189 223 278 294 309 390	L2a1
JAM010‡	189 223 278 294 309 390	L2a1c
JAM011	126 178 189 223 264 270 278 293 311	L1b1
JAM012‡	086 114 172 223 278 294 309 390	L2a1c
JAM013	124 223 278 362	L3b
JAM014	066 124 145 183C 189 223 278 362	L3b2
JAM015	189 223 278 294 309 390	L2a1
JAM016‡	187 189 223 278 294 309 390	L2a1
JAM017	114A 129 213 223 278 355 362 390	L2b1a
JAM018	223 278 294 309 319 390	L2a1
JAM019†	223 230 399	L3e2a1b1
JAM020	126 187 189 223 264 270 278 293 311	L1b1
JAM021	093 223 264 278 390	L2c2
JAM022	078 129 183 187 189 223 265C 286A 294 311 320 360	L1c2b2
JAM023‡	066 129 223 278 362	L3b
JAM024	223 265T	L3e3
JAM025	126 187 189 223 264 270 278 311	L1b1
JAM026	189 223 278 294 309 390	L2a1
JAM027‡	172 223 278 294 309 390	L2a1e

JAM028	187 189 223 278 294 390	L2a1'2
JAM029‡	223 278 294 309 325 390	L2a1c
JAM030	189 223 278 294 309 390	L2a1
JAM031*	223	M*
JAM032†	066 124 223 278 362	L3b
JAM033*†	126 187 189	J
JAM034	086 114 172 223 278 294 309 390	L2a1c1
JAM035‡	124 209 223 278 362	L3b
JAM036†	093 223 264 278 390	L2c2
JAM037†	223 256 278 294 309 390	L2a1
JAM038	189 223 278 294 309 390	L2a1
JAM039	093 223 278 294 309 390	L2a1a3
JAM040†	223 278 294 309 319 390	L2a1
JAM041	189 223 278 294 309 390	L2a1
JAM042	189 223 278 294 309 390	L2a1
JAM043‡	183C 189 217	B2
JAM044	189 223 278 294 309 390	L2a1
JAM045	093 223 264 278 390	L2c2
JAM046	086 223 278 294 309 390	L2a1c1
JAM047‡	223 278 355 362	L3b
JAM048‡	086 223 278 294 309 390	L2a1c
JAM049‡	189 192 223 278 294 309 390	L2a1
JAM050	017 126 187 189 223 264 270 278 293 311	L1b1
JAM051	051 111 124 183C 189 223	L3d2
JAM052	129 187 189 223 265C 278 286G 294 311 360	L1c2
JAM053	172 187 189 223 265C 278 286G 294 311 360	L1c2
JAM054‡	189 223 278 294 309 390	L2a1c
JAM055	126 187 189 223 264 270 278 311	L1b1
JAM056*†	223 266 278 287 340 390	L2
JAM057‡	189 223 278 294 309 390	L2a1c
JAM058‡	092 223 278 294 309 390	L2a1a
JAM059	126 187 189 223 264 270 278 293 311	L1b1
JAM060	126 187 189 223 264 270 278 293 294 309 311	L1b1
JAM061	114A 129 213 223 278 311 362 390	L2b1
JAM062	172 183C 189 223 320	L3e2b
JAM063	223 278 294 309 319 390	L2a1
JAM064	124 223	L3d
JAM065	126 187 189 223 264 270 278 293 311	L1b1
JAM066	189 223 278 294 309 390	L2a1
JAM067	223 278 294 309 319 390	L2a1
JAM068	189 223 278 294 309 390	L2a1
JAM069	223 278 294 309 319 390	L2a1
JAM070	223 320 399	L3e2a1b1
JAM071	172 189 223 320	L3e2b
JAM072	093 223 264 278 390	L2c2
JAM073	124 223	L3d
JAM074	189 223 278 294 309 390	L2a1
JAM075	126 187 189 223 264 270 278 293 311	L1b1
JAM076	126 187 189 223 264 270 278 293 311	L1b1
JAM077	182C 183C 189 218 223 234 278 294 309 357 390	L2a1k
JAM078	223 278 294 309 319 390	L2a1
JAM079	086 114 172 223 278 294 309 390	L2a1c1
JAM080	126 187 189 223 264 270 278 293 311	L1b1
JAM081‡	126 187 189 223 264 270 278 293 311	L1b
JAM082‡	223 278 294 309 390	L2a1a
JAM083	093 223 264 278 390	L2c2
JAM084	223 278 294 309 319 390	L2a1
JAM085	129 189 278 300 390 399	L2d1
JAM086	223 270 320 399	L3e2a1b1
JAM087	189 223 278 294 309 390	L2a1
JAM088	129 148 168 172 187 188G 189 223 230 278 293 311 320	L0a1b
JAM089	189 192 223 278 294 309 390	L2a1l

JAM090†	126 187 189 223 264 270 278 293 311	L1b1
JAM091	124 223 278 311 362	L3b1a3
JAM092†	124 223 278 292 362	L3b
JAM093‡	223 265T	L3e3
JAM094	223 278 294 309 390	L2a1
JAM095‡	037 126 172 187 189 223 264 270 278 293 301	L1b
JAM096	129 223 278 294 390	L2a1'2
JAM097	189 223 248 278 294 309 390	L2a1k
JAM098	129 184 189 215 223 278 294 311 360	L1c3a
JAM099	129 209 223 292 295 311	L3f1b1
JAM100‡	223 327	L3e1
JAM101	223 291 320 399	L3e2a1b1
JAM102	126 187 189 223 264 270 278 293 311	L1b1
JAM103	126 187 189 223 264 270 278 293 311	L1b1
JAM104	235 291 293	H2a2b1
JAM105‡	129 183 187 189 223 265C 286A 294 311 320 360	L1c2b2
JAM106	126 187 189 223 264 270 278 311	L1b1
JAM107	114A 129 213 223 390	L2b
JAM108‡	017 129 163 187 189 223 278 293 294 311 360	L1c3b1
JAM109	126 187 189 223 264 270 278 293 311	L1b1
JAM110	093 129 187 189 223 265C 278 294 311 360	L1c2
JAM111	093 129 187 189 223 278 290 293 294 311 360 368	L1c1
JAM112	223 299 320 399	L3e2a1b1
JAM113	086 223 278 294 309 390	L2a1c1
JAM114*	223 261 278 390	L2c
JAM115‡	093 187 189 223 230 265C 278 294 311	L1c1'2'4'6
JAM116	189 278 294 309 390	L2a1k
JAM117	124 166 223	L3d1c
JAM118‡	172 189 223 278 294 309 362 390	L2a1
JAM119‡	124 223	L3d1b
JAM120	223 278 286 294 309 390	L2a1a2
JAM121‡	093 148 223 265T 311	L3e3
JAM122‡	086 102 129 148 172 184 187 189 223 261 278 311	L1c1'2'4'6
JAM123	093 189 223 278 294 309 390	L2a1a3
JAM124‡	124 223 278 362	L3b
JAM125	051 219 223 264	L3e4
JAM126	086 187 189 223 278 293 294 311 360	L1c1b'c'd
JAM127‡	172 183C 189 223 311 320	L3e2b
JAM128	129 187 189 223 274 278 293 294 311 360	L1c1a2
JAM129	189 223 278 294 309 390	L2a1
JAM130	126 187 189 223 264 270 278 311	L1b1
JAM131	185 189 223 264 327	L3e1a
JAM132‡	223 320 399	L3e2a1b1
JAM133‡	223 278 294 309 390	L2a1c
JAM134	129 145 189 223 234 278 294 311 360 368	L1c
JAM135	114A 129 187 189 213 223 278 293 355 362 390	L2b1a
JAM136	038 086 129 187 189 223 278 284 293 294 311 360	L1c1d
JAM137	126 172 187 189 223 264 270 311 368	L1b1
JAM138‡	172 183C 189 223 266 320	L3e2b
JAM139	124 223 256	L3d
JAM140	124 223 278 362	L3b
JAM141	126 187 189 223 264 270 278 292 311	L1b1
JAM142	172 184 219 234 278 311	U6a
JAM143	129 148 168 172 187 188G 189 223 230 278 293 311 320	L0a1b
JAM144‡	223 278 294 309 390	L2a1c
JAM145	223 278 294 309 390	L2a1
JAM146	129 209 223 291 292 295 311	L3f1b1
JAM147	223 278 294 309 390	L2a1
JAM148	172 183C 189 219 240 278	U6a1
JAM149‡	129 209 223 292 295 311	L3f1b1
JAM150	129 223 278 294 309 390	L2a1c5
JAM151	114A 129 213 223 278 390	L2b

JAM152	136 172 183C 189 223 239 311 320	L3e2b
JAM153	093 129 189 278 300 311 354 390 399	L2d1a
JAM154	124 223 278 362	L3b
JAM155‡	129 187 189 223 274 278 293 294 311	L1c1a2
JAM156	209 223 311	L3f1b4a
JAM157	129 189 209 223 292 295 311	L3f1b1
JAM158	278 287 294 309 390	L2a1
JAM159	172 189 223 278 320	L3e2b
JAM160	124 223 278 362	L3b
JAM161	129 187 189 223 265C 278 286G 294 311 360	L1c2
JAM162	182C 183C 184 189 192 223 278 290 294 309 390	L2a1b1
JAM163	124 223 278 325	L3b
JAM164	129 223 278 294 309 390	L2a1c5
JAM165*	129 223 254 266 278 344 368 390	L2
JAM166	124 223 278 293 362 399	L3b
JAM167	223 278 390	L2
JAM168	189 223 278 294 390	L2a1'2
JAM169*	129 223	M5
JAM170‡	111A 145 184 239 278 292 311 355 390 399 400	L2e
JAM171	223 274 278 290 294 390	L2a1'2
JAM172	051 206C 291 359	U2
JAM173‡	172 223 278 309 327	L3e1
JAM174*†	223 278 390	L2c
JAM175†	189 223 278 294 309 390	L2a1
JAM176‡	086 223 278 294 309 390	L2a1c
JAM177	092 129 148 168 172 187 188G 189 223 230 311	L0a1'4
JAM178‡	189 223 278 294 309 390	L2a1
JAM179‡	185 223 264 327	L3e1
JAM180	093 148 223 265T	L3e3
JAM181	223 278 294 309 390	L2a1
JAM182*	223 278 318 363 390	L2c
JAM183‡	051 223 264	L3e4
JAM184	129 148 172 187 188G 189 223 230 311 320 362	L0a
JAM185	223 320	L3e2
JAM186	126 187 189 223 264 270 278 293 311	L1b1
JAM187	114A 129 189 223 278 390	L2b
JAM188†	126 187 189 223 264 270 278 293 311	L1b1
JAM189‡	185 223 327	L3e1a3
JAM190	124 223	L3d
JAM191	223 278 390	L2
JAM192	185 189delT 223 264 327	L3e1a
JAM193‡	179 183C 189 223 239 311 320 362	L4b1
JAM194	086 129 148 183C 189 278 300 354 390 399	L2d1
JAM195	124 223 278 362	L3b
JAM196‡	223 278 390	L2
JAM197‡	124 209 223 278 362	L3b
JAM198	148 172 187 188G 189 223 230 256 311 320	L0a2
JAM199‡	223 278 294 390	L2a1
JAM200	209 223 292 311	L3f1b
JAM201‡	172 189 223 311 320	L3e2b1a
JAM202*†	223 278 390	L2c
JAM203	126 187 189 223 264 270 278 293 301 311	L1b1
JAM204‡	172 189 223 320	L3e2b1a
JAM205‡	209 223 292 295 311	L3f1
JAM206‡	129 209 223 292 295 311	L3f1b1
JAM207	114A 129 213 223 278 355 362 390	L2b1a
JAM208	111 126 187 189 223 239 270 278 293 311	L1b
JAM209	111 223 290 299 319 362	A2
JAM210	093 189 192 223 278 294 309 390	L2a1l
JAM211	129 148 168 172 187 188G 189 223 230 278 293 311 320	L0a1b
JAM212‡	213 223 239 278 294 309 390	L2a1c2
JAM213‡	223 278 294 309 390	L2a1a

JAM214‡	148 172 187 188G 189 223 230 311 320	L0a2
JAM215	093 189 223 264 270 278 294 309	L2a1a3
JAM216	129 187 189 223 265C 278 286G 288 294 311 360	L1c2
JAM217‡	126 187 189 223 264 270 278 293 311 362 400	L1b
JAM218	129 187 189 223 265C 278 286G 290 311 359 360	L1c2
JAM219‡	086 223 278 294 309 390	L2a1e
JAM220‡	223 325delT 327	L3e1
JAM221	037 093 129 140 184 187 189 223 278 294 301 311 360 390	L1c4a
JAM222‡	129 148 168 172 187 188G 189 223 230 311 320 362	L0a
JAM223	069 126	J
JAM224	129 209 223 292 295 311	L3f1b1
JAM225	129 184 189 215 223 278 294 311 360	L1c3a
JAM226	126 166 187 189 223 264 270 278 293 311	L1b1
JAM227	172 183C 189 223 265 320	L3e2b
JAM228	172 182C 183C 189 223 320	L3e2b
JAM229	172 183C 189 223 239 320	L3e2b
JAM230	111 124 223	L3d2
JAM231	126 187 189 223 239 264 270 278 310	L1b1
JAM232‡	172 183C 189 223 320	L3e2b
JAM233	038 086 129 187 189 223 278 284 293 294 311 360	L1c1d
JAM234	093 124 223 256 265C	L3d
JAM235	129 189 209 223 292 295 311	L3f1b1
JAM236	209 223 292 311	L3f1b
JAM237	126 187 189 223 264 270 278 293 311	L1b1
JAM238	148 167 172 187 188G 189 223 230 311 320	L0a2
JAM239*	114A 189 223 278 294 362 390	L2a1i
JAM240	038 086 129 187 189 223 278 293 294 311 360 362	L1c1d
JAM241	223 265T	L3e3
JAM242	129 209 223 292 295 311	L3f1b1
JAM243	213 223 278 294 309 390	L2a1
JAM244*	051 223 264 290	L3e4
JAM245	124 223 278 311 362	L3b1a3
JAM246	129 145 189 215 223 278 294 311 319 360	L1c3a
JAM247*	093 223 278 294 368 390	L2a1a1
JAM248‡	172 183C 189 223 320	L3e2b
JAM249	189 192 223 274 278 294 309 390	L2a1l
JAM250	223 278 294 309 390	L2a1
JAM251‡	129 223 278 294 309 390	L2a1c
JAM252‡	126 187 189 223 264 270 278 311 327	L1b
JAM253	172 183C 189 223 320	L3e2b
JAM254	213 223 239 278 294 309 390	L2a1c2
JAM255	189 223 278 294 309 390	L2a1
JAM256	172 183C 189 223 320	L3e2b
JAM257	223 265T	L3e3
JAM258‡	126 187 189 223 264 270 278 293 301 311	L1b
JAM259	111 223 278 294 309 390	L2a1
JAM260	223 264 278 390	L2c2
JAM261	189 192 223 278 294 309 390	L2a1l
JAM262	051 223 264 319	L3e4
JAM263	223 320 399	L3e2a1b1
JAM264	093 129 148 168 172 187 188G 189 223 230 278 293 311 320	L0a1b2
JAM265*†	223 278 390	L2c
JAM266‡	129 148 168 172 187 188G 189 223 230 311 320	L0a
JAM267	189 192 223 278 294 362 390	L2a1i
JAM268	124 223	L3d
JAM269	148 168 172 187 188G 189 223 230 311 320 362	L0a1'4
JAM270	124 223 256 278 362	L3b
JAM271	124 223 362	L3d5
JAM272	189 223 234 249 278 294 295 390	L2a1'2
JAM273	114A 129 213 223 278 355 362 390	L2b1a

JAM274	172 183C 189 223 320	L3e2b
JAM275	126 187 189 223 264 270 278 311	L1b1
JAM276†	129 209 223 292 295 311	L3f1b1
JAM277	129 148 172 187 188G 189 223 230 311 320 362	L0a
JAM278	114A 129 213 223 278 355 356 362 390	L2b1a
JAM279	126 187 189 223 264 270 278 311	L1b1
JAM280*	145 166 223 224 278 295 390	L2
JAM281	124 166 223	L3d1c
JAM282‡	093 148 223 265T	L3e3
JAM283	172 183C 189 223 311 320	L3e2b
JAM284	093 126 187 189 223 264 270 278 311	L1b1a6
JAM285	189 223 278 294 309 390	L2a1
JAM286	189 192 223 278 294 362 390	L2a1i
JAM287	126 187 189 223 264 270 278 311	L1b1
JAM288	147 172 183C 189 223 320	L3e2b
JAM289	189 192 223 278 294 362 390	L2a1i
JAM290	129 187 189 223 265C 278 286G 288 294 311 360	L1c2
JAM291	129 148 168 172 187 188G 189 223 230 311 320 362	L0a1'4
JAM292	129 148 168 172 187 188G 189 223 230 311 320 325 362	L0a1'4
JAM293	172 183C 189 223 320	L3e2b
JAM294	223 320 399	L3e2a1b1
JAM295	129 148 168 172 187 188G 189 223 230 311 320	L0a1'4
JAM296	124 223 278 362	L3b
JAM297‡	169 195 223 235 243 244 311	L3x2b
JAM298	038 086 129 187 189 223 278 284 293 294 311 325 360	L1c1d
JAM299	114A 129 213 223 278 355 362 390	L2b1a
JAM300*	223 242 249 278 390	L2c
JAM301	129 172 174A 184 187 189 223 261 278 290 293 311 360	L1c1
JAM302	126 187 189 223 264 270 278 293 311	L1b1
JAM303	176 223 311 327	L3e1d
JAM304	172 182C 183C 189 223 320	L3e2b
JAM305‡	172 183C 189 223 320	L3e2b
JAM306	189 192 223 278 294 309 390	L2a1l
JAM307	075 185 189 223 278 294 309 390	L2a1k
JAM308	126 187 189 223 264 268 270 278 293 311	L1b1
JAM309	172 183C 189 223 261 309 320	L3e2b
JAM310	114A 129 213 223 265 278 311 368 390	L2b
JAM311	223 278 294 309 390	L2a1
JAM312	124 223 256	L3d
JAM313	148 218 223 278 320	L3e2
JAM314	223 291	D4j2
JAM315	126 187 189 223 264 270 278 293 311	L1b1
JAM316	126 187 189 223 264 270 278 311	L1b1
JAM317	126 187 189 223 264 270 278 293 311	L1b1
JAM318	189 223 278 294 304 309 390	L2a1
JAM319	189 192 223 278 294 390	L2a1l
JAM320	183C 189 223 278 294 362 390	L2a1i
JAM321‡	126 187 189 223 264 270 278 311	L1b
JAM322‡	172 183C 189 219 240 278	U6a
JAM323	092 182C 183C 189 223 278 294 390	L2a1'2
JAM324*	172 223 278 311 318 390	L2c
JAM325	209 223 292 311	L3f1b
JAM326	129 209 223 292 295 311	L3f1b1
JAM327	183C 189 223 278 294 309 362 390	L2a1i
JAM328	223 265T 316	L3e3
JAM329	129 209 223 266 292 295 311	L3f1b1
JAM330	223 278 286 294 390	L2a1a2
JAM331	223 278 294 309 368 390	L2a1a1
JAM332	179 223 327	L3e1
JAM333*	355	L3*
JAM334	185 223 311 327	L3e1a1
JAM335	129 187 189 213 223 265C 274 278 286G 294 311 360	L1c2a1

JAM336	223 278 294 309 390	L2a1
JAM337	114A 129 213 223 278 390	L2b
JAM338	124 192 223 271 278 362	L3b
JAM339	126 187 189 223 264 270 278 311	L1b1
JAM340	189 223 264 278 390	L2c2
JAM341	114A 129 213 223 278 327 355 356 362 390	L2b1a
JAM342	223 274 278 294 390	L2a1'2
JAM343	038 086 108 129 187 189 223 278 284 293 294 311 360	L1c1d
JAM344	129 189 223 265C 278 294 311 360	L1c2
JAM345	129 187 189 223 265C 278 286G 294 311 360	L1c2
JAM346	126 189 223 264 270 278 293 311	L1b1
JAM347*	223 278 362 390 399	L2c
JAM348‡	223 278 304 355 362	L3b
JAM349	172 183C 189 223 320	L3e2b
JAM350	209 223 311	L3f1b4a
JAM351	209 223 292 295 311	L3f1b1
JAM352	129 189 192 209 223 278 294 309 390	L2a1
JAM353‡	124 223	L3d1b
JAM354	126 187 189 223 264 270 278 293 311	L1b1
JAM355	172 183C 187 189 223 320	L3e2b
JAM356	048 124 189 223 278 362	L3b2
JAM357	223 278 294 309 390	L2a1
JAM358	209 223 292 311	L3f1b
JAM359	166delA 223 356	M30c1
JAM360	048 124 223 274 278 362	L3b
JAM361	086 209 223 292 311	L3f1b
JAM362‡	172 183C 189 223 320	L3e2b
JAM363	129 209 223 292 295 311	L3f1b1
JAM364*	129 223 362	D4a
JAM365	129 148 168 172 187 188G 189 223 230 294 311 320 362	L0a1'4
JAM366	038 086 108 129 187 189 223 278 284 293 294 311 360	L1c1d
JAM367	189 192 223 278 294 309 390	L2a1l
JAM368	172 183C 189 223 320	L3e2b
JAM369	126 187 189 223 264 270 278 293 311	L1b1
JAM370	114A 129 213 223 278 311 362 390	L2b1
JAM371	129 209 223 292 295 311	L3f1b1
JAM372	209 223 292 311	L3f1b
JAM373	124 223 399	L3d
JAM374	172 183C 189 209 223 311 320	L3e2b
JAM375	086 223 270 278 294 309 368 390	L2a1a1
JAM376	124 223	L3d
JAM377	129 209 223 292 295 311	L3f1b1
JAM378‡	223 278 390	L2
JAM379	209 223 235 274 292 311	L3f1b
JAM380	093 148 172 187 188G 189 223 230 311 320	L0a2
JAM381	086 183C 189 223 264 278 294 309 390	L2a1c1
JAM382*	114A 189 192 223 278 294 362 390	L2a1i
JAM383	114A 129 213 223 278 362 390	L2b1
JAM384	189 192 223 278 294 309 390	L2a1l
JAM385	172 183C 189 223 320	L3e2b
JAM386	223 278 286 294 309 390	L2a1a2
JAM387	129 148 168 172 187 188G 189 223 230 311 320	L0a1'4
JAM388‡	093 129 148 168 172 187 188G 189 223 230 278 311 320	L0a
JAM389	223 399	L3h1a2a
JAM390	223 278 294 309 390	L2a1
JAM391‡	126 187 189 223 264 270 278 311	L1b
JAM392*	114A 189 223 278 294 362 390	L2b
JAM393	223 320	L3e2
JAM394‡	223 278 294 309 368 390	L2a1e
JAM395‡	223 278 294 309 390	L2a1c
JAM396	223 325delT 327	L3e1b
JAM397	124 223	L3d

JAM398	124 223 278 355 362	L3b
JAM399	129 187 189 223 265C 278 286G 294 311 359 360	L1c2
JAM400‡	223 278 318 390	L2

7.2 African Mitochondrial DNA HVS-1 sequences included in 3.2.4.

Ethnicities and country of origin are shown along with the appropriate reference. Each groups either residing in the Sahel region of Africa or belonging to pygmy groups of Central Africa were noted as such. "Multiple groups" denotes incidences when the authors do not distinguish the ethnicities for a particular sample.

Ethnicity	Country	*N*	Sahel	Pygmy	Reference
Senegambia					
Bijago	Guinea-Bissau	22	no	no	Rosa *et al.*, 2004
Multiple groups	Guinea-Bissau	17	no	no	Rosa *et al.*, 2004
Wolof	Senegal	131	yes	no	Watson *et al.*, 1997; Rando *et al.*, 1998; Stefflova *et al.*, 2011; Ely *et al.*, unpublished
Mandenka	Senegal	110	yes	no	Watson *et al.*, 1997
Malinke	Senegal	93	yes	no	Gonzalez *et al.*, 2006; Ely *et al.*, unpublished
Balanta	Guinea-Bissau	70	yes	no	Rosa *et al.*, 2004; Ely *et al.*, unpublished
Joola	Senegal	49	yes	no	Ely *et al.*, unpublished
Fula	Guinea-Bissau	38	yes	no	Rosa *et al.*, 2004
Peul	Senegal	32	yes	no	Gonzalez *et al.*, 2006; Stefflova *et al.*, 2011; Ely *et al.*, unpublished
Serer	Senegal	30	yes	no	Rando *et al.*, 1998; Ely *et al.*, unpublished
Manjaco	Guinea-Bissau	27	yes	no	Rosa *et al.*, 2004
Malu	Guinea-Bissau	26	yes	no	Rosa *et al.*, 2004
Papel	Guinea-Bissau	24	yes	no	Rosa *et al.*, 2004; Ely *et al.*, unpublished
Tuareg	Senegambia	24	yes	no	Watson *et al.*, 1997; Gonzalez *et al.*, 2006
Mancanha	Guinea-Bissau	20	yes	no	Rosa *et al.*, 2004; Ely *et al.*, unpublished
Beafada	Guinea-Bissau	19	yes	no	Rosa *et al.*, 2004
Fula-Preto	Guinea-Bissau	19	yes	no	Rosa *et al.*, 2004
Futa-Fula	Guinea-Bissau	19	yes	no	Rosa *et al.*, 2004
Djola	Guinea-Bissau	18	yes	no	Rosa *et al.*, 2004
Mansonca	Guinea-Bissau	18	yes	no	Rosa *et al.*, 2004
Tukulor	Senegal	14	yes	no	Rando *et al.*, 1998; Stefflova *et al.*, 2011
Diola	Senegal	11	yes	no	Rando *et al.*, 1998; Stefflova *et al.*, 2011
Multiple groups	Senegal	16	yes	no	Ely *et al.*, unpublished
Multiple groups	Senegal	8	yes	no	Rando *et al.*, 1998
Soninke	Senegal	8	yes	no	Rando *et al.*, 1998; Gonzalez *et al.*, 2006
Multiple groups	Guinea-Bissau	15	yes	no	Rosa *et al.*, 2004
Multiple groups	Senegal	7	yes	no	Stefflova *et al.*, 2011
Lebou	Senegal	4	yes	no	Stefflova *et al.*, 2011; Ely *et al.*, unpublished
Maure	Senegal	3	yes	no	Stefflova *et al.*, 2011; Ely *et al.*, unpublished
Total		**892**	**39**	**892**	

Sierra Leone						
	Ahizi	Ivory Coast	128	no	no	Brucato *et al.*, 2010
	Temne	Sierra Leone	121	no	no	Jackson *et al.*, 2005
	Sierra Leone	Sierra Leone	109	no	no	Monson *et al.*, 2002
	Bambara	Mali	79	no	no	Gonzalez *et al.*, 2006; Stefflova *et al.*, 2011; Ely *et al.*, unpublished
	Limba	Sierra Leone	67	no	no	Jackson *et al.*, 2005; Ely *et al.*, unpublished
	Yacouba	Ivory Coast	62	no	no	Brucato *et al.*, 2010
	Mende	Sierra Leone	59	no	no	Jackson *et al.*, 2005
	Loko	Sierra Leone	29	no	no	Jackson *et al.*, 2005
	Bamana	Mali	3	no	no	Ely *et al.*, unpublished
	Baoulé	Ivory Coast	1	no	no	Brucato *et al.*, 2010
	Kru	Liberia	1	no	no	Ely *et al.*, unpublished
	Fulani Banfora	Burkina Faso	59	yes	no	Cerny *et al.*, 2006
	Mandinga	Guinea-Bissau	30	yes	no	Rosa *et al.*, 2004
	Manjak	Guinea-Bissau	22	yes	no	Ely *et al.*, unpublished
	Senoufo	Mali	12	yes	no	Gonzalez *et al.*, 2006; Ely *et al.*, unpublished
	Mandingo	Guinea-Bissau	11	yes	no	Ely *et al.*, unpublished
	Bobo	Mali	8	yes	no	Gonzalez *et al.*, 2006; Ely *et al.*, unpublished
	Multiple groups	Mali	8	yes	no	Ely *et al.*, unpublished
	Dogon	Mali	7	yes	no	Gonzalez *et al.*, 2006; Ely *et al.*, unpublished
	Sonrhai	Mali	6	yes	no	Gonzalez *et al.*, 2006
	Burkinabè	Burkina Faso	1	yes	no	Brucato *et al.*, 2010
Total			**823**	**659**	**823**	

Gold Coast						
	Akan	Ghana	275	no	no	Veeramah *et al.*, 2010; Ely *et al.*, unpublished
	Ewe	Ghana	115	no	no	Veeramah *et al.*, 2010; Ely *et al.*, unpublished
	Gaa-Adangbe	Ghana	77	no	no	Ely *et al.*, unpublished
	Multiple groups	Ghana	24	no	no	Ely *et al.*, unpublished
	Songhai	Burkina Faso	10	yes	no	Watson *et al.*, 1997
	Mossi	Burkina Faso	2	yes	no	Ely *et al.*, unpublished
	Songhoy	Burkina Faso	2	yes	no	Ely *et al.*, unpublished
Total			**505**	**491**	**505**	
Bight of Benin						
	Yoruba	Nigeria	156	no	no	Watson *et al.*, 1997; Brucato *et al.*, 2010; Ely *et al.*, unpublished
	Fon	Benin	79	no	no	Brucato *et al.*, 2010
	Aïzo	Benin	11	no	no	Brucato *et al.*, 2010
	Goun	Benin	11	no	no	Brucato *et al.*, 2010
	Multiple groups	Benin	40	no	no	Brucato *et al.*, 2010
	Hausa	Nigeria/Niger	68	yes	no	Watson *et al.*, 1997; Brucato *et al.*, 2010; Ely *et al.*, unpublished
	Fulani Tindangou	Burkina Faso	54	yes	no	Cerny *et al.*, 2006
	Multiple groups	Benin	2	yes	no	Brucato *et al.*, 2010
Total			**421**	**297**	**421**	

Bight of Biafra

Ibibio	Nigeria	509	no	no	Veeramah *et al.*, 2010
Igbo	Nigeria	293	no	no	Veeramah *et al.*, 2010
Efik	Nigeria	145	no	no	Veeramah *et al.*, 2010
Ejagham	Nigeria	133	no	no	Veeramah *et al.*, 2010
Ngumba	Cameroon / Equatorial Guinea	132	no	no	Batini *et al.*, 2007; Quintana-Murci *et al.*, 2008
Fang	Cameroon	116	no	no	Pinto *et al.*, 1996; Quintana-Murci *et al.*, 2008
Aghem	Cameroon	115	no	no	Veeramah *et al.*, 2010
Bamun	Cameroon	109	no	no	Veeramah *et al.*, 2010; Ely *et al.*, unpublished
Annang	Nigeria	107	no	no	Veeramah *et al.*, 2010
Oron	Nigeria	98	no	no	Veeramah *et al.*, 2010
Fulbe	Niger / Cameroon	94	no	no	Watson *et al.*, 1997;Coia *et al.*, 2005
Ewondo	Cameroon	81	no	no	Coia *et al.*, 2005; Quintana-Murci *et al.*, 2008; Ely *et al.*, unpublished
Fali	Cameroon	81	no	no	Batini *et al.*, 2007
Bamileke	Cameroon	53	no	no	Coia *et al.*, 2005; Ely *et al.*, unpublished
Bakaka	Cameroon	50	no	no	Coia *et al.*, 2005
Benga	Gabon	50	no	no	Quintana-Murci *et al.*, 2008
Bassa	Cameroon	46	no	no	Coia *et al.*, 2005
Bubi	Equatorial Guinea	45	no	no	Mateu *et al.*, 1997
Podoko	Cameroon	43	no	no	Coia *et al.*, 2005; Ely *et al.*, unpublished
Mandara	Cameroon	37	no	no	Coia *et al.*, 2005
Uldeme	Cameroon	28	no	no	Coia *et al.*, 2005
Tupuri	Cameroon	25	no	no	Coia *et al.*, 2005
Daba	Cameroon	20	no	no	Coia *et al.*, 2005
Tali	Cameroon	20	no	no	Coia *et al.*, 2005
Multiple groups	Cameroon	34	no	no	Ely *et al.*, unpublished
Baka	Cameroon	177	no	yes	Batini *et al.*, 2007; Quintana-Murci *et al.*, 2008
Bakola	Cameroon	137	no	yes	Batini *et al.*, 2007; Quintana-Murci *et al.*, 2008
Biaka	Central Africa Republic	73	no	yes	Watson *et al.*, 1997; Quintana-Murci *et al.*, 2008
Tikar	Cameroon	69	no	yes	Quintana-Murci *et al.*, 2008; Veeramah *et al.*, 2010
W.Mbenzele	Cameroon	57	no	yes	Destro-Bisol *et al.*, 2004
Bakoya	Gabon	31	no	yes	Quintana-Murci *et al.*, 2008
Fulani Bongor	Chad	105	yes	no	Cerny *et al.*, 2006; Cerny *et al.*, 2007
Chadic-Cameroon	Cameroon	104	yes	no	Cerny *et al.*, 2004
Fulani Tcheboua	Cameroon	86	yes	no	Cerny *et al.*, 2006; Cerny *et al.*, 2007
Kotoko	Cameroon	56	yes	no	Cerny *et al.*, 2007
Kanembu	Lake Chad	50	yes	no	Cerny *et al.*, 2007
Kanuri	Nigeria / Cameroon / Sudan / Niger	49	yes	no	Watson *et al.*, 1997; Cerny *et al.*, 2007; Ely *et al.*, unpublished
Arabs Shuwa	Lake Chad	38	yes	no	Cerny *et al.*, 2007
Mafa	Lake Chad	32	yes	no	Cerny *et al.*, 2007
Masa	Lake Chad	32	yes	no	Cerny *et al.*, 2007

	Buduma	Lake Chad	30	yes	no	Cerny *et al.*, 2007
	Arabs Chad	Lake Chad	27	yes	no	Cerny *et al.*, 2007
	Hide	Lake Chad	23	yes	no	Cerny *et al.*, 2007
	Arab Choa	Cameroon	1	yes	no	Ely *et al.*, unpublished
Total			**3641**	**3008**	**3097**	
West-central Africa						
	Nyaneka-Nkhumbi	Angola	153	no	no	Coelho *et al.*, 2009
	Cabinda	Angola	110	no	no	Beleza *et al.*, 2005
	Ovimbundu	Angola	92	no	no	Coelho *et al.*, 2009
	Mitsogo	Gabon	64	no	no	Quintana-Murci *et al.*, 2008
	Nzebi	Gabon	63	no	no	Quintana-Murci *et al.*, 2008
	Kota	Gabon	56	no	no	Quintana-Murci *et al.*, 2008
	Ateke	Gabon	54	no	no	Quintana-Murci *et al.*, 2008
	Kuvale	Angola	54	no	no	Coelho *et al.*, 2009
	Punu	Gabon	52	no	no	Quintana-Murci *et al.*, 2008
	Galoa	Gabon	51	no	no	Quintana-Murci *et al.*, 2008
	Shake	Gabon	51	no	no	Quintana-Murci *et al.*, 2008
	Bateke	Democratic Republic of Congo	50	no	no	Batini *et al.*, 2007
	Akele	Gabon	48	no	no	Quintana-Murci *et al.*, 2008
	Duma	Gabon	47	no	no	Quintana-Murci *et al.*, 2008
	Obamba	Gabon	47	no	no	Quintana-Murci *et al.*, 2008
	Angoloan Others	Angola	45	no	no	Coelho *et al.*, 2009
	Makina	Gabon	45	no	no	Quintana-Murci *et al.*, 2008
	Mbundu	Angola	43	no	no	Plaza *et al.*, 2004
	Eshira	Gabon	40	no	no	Quintana-Murci *et al.*, 2008
	Ndumu	Gabon	39	no	no	Quintana-Murci *et al.*, 2008
	Eviya	Gabon	38	no	no	Quintana-Murci *et al.*, 2008
	Sanga	Central African Republic	30	no	no	Batini *et al.*, 2007
	Ganguela	Angola	21	no	no	Coelho *et al.*, 2009
	Orungu	Gabon	20	no	no	Quintana-Murci *et al.*, 2008
	Bakongo	Angola	1	no	no	Plaza *et al.*, 2004
	Babongo	Gabon	45	no	yes	Quintana-Murci *et al.*, 2008
	Babinga	Democratic Republic of Congo	44	no	yes	Batini *et al.*, 2007
Total			**1403**	**1403**	**1314**	

South-east Africa						
	Bantu-speaking	Mozambique	109	no	no	Pereira *et al.*, 2001
	!Kung	Namibia	96	no	no	Watson *et al.*, 1997; Chen *et al.*, 2000; Ely *et al.*, unpublished
	Shona	Zimbabwe	76	no	no	Salas *et al.*, 2002; Castri *et al.*, 2009
	Antandroy	Madagascar	59	no	no	Tofanelli *et al.*, 2009
	Antanosy	Madagascar	54	no	no	Tofanelli *et al.*, 2009
	Khwe	Namibia	31	no	no	Chen *et al.*, 2000
	Chopi	Mozambique	27	no	no	Salas *et al.*, 2002
	Ronga	Mozambique	22	no	no	Salas *et al.*, 2002
	Shangaan	Mozambique	22	no	no	Salas *et al.*, 2002
	Sena	Mozambique	21	no	no	Salas *et al.*, 2002
	Sukuma	Tanzania	21	no	no	Knight *et al.*, 2003
	Chwabo	Mozambique	20	no	no	Salas *et al.*, 2002
	Lomwe	Mozambique	20	no	no	Salas *et al.*, 2002
	Makhuwa	Mozambique	20	no	no	Salas *et al.*, 2002
	Nyanja	Mozambique	20	no	no	Salas *et al.*, 2002
	Nyungwe	Mozambique	20	no	no	Salas *et al.*, 2002
	Tonga	Mozambique	20	no	no	Salas *et al.*, 2002
	Herero	Botswana	19	no	no	Ely *et al.*, unpublished
	Makonde	Mozambique	19	no	no	Salas *et al.*, 2002
	Ndau	Mozambique	19	no	no	Salas *et al.*, 2002
	Tswa	Mozambique	19	no	no	Salas *et al.*, 2002
	Antaisaka	Madagascar	11	no	no	Tofanelli *et al.*, 2009
	Nguni	Mozambique	11	no	no	Salas *et al.*, 2002
	Yao	Mozambique	10	no	no	Salas *et al.*, 2002
	Merina	Madagascar	9	no	no	Tofanelli *et al.*, 2009
Total			**775**	**775**	**775**	

East Africa						
	Amhara	Ethiopia	127	no	no	Thomas *et al.*, 2002
	Kenya	Kenya	100	no	no	Brandstatter *et al.*, 2004
	Luhya	Kenya	94	no	no	Ely *et al.*, unpublished
	Hadza	Tanzania	66	no	no	Knight *et al.*, 2003; Ely *et al.*, unpublished
	Oromo	Ethiopia	51	no	no	Quintana-Murci *et al.*, 1999; Kivisild *et al.*, 2004
	Dinka	Sudan	47	no	no	Krings *et al.*, 1999
	Tigrai	Ethiopia	45	no	no	Kivisild *et al.*, 2004
	Hutu	Rwanda	42	no	no	Castri *et al.*, 2009
	Turkana	Kenya	37	no	no	Watson *et al.*, 1997
	Somali	Somalia	27	no	no	Watson *et al.*, 1997
	Kikuyu	Kenya	24	no	no	Watson *et al.*, 1997
	Gurage	Ethiopia	21	no	no	Kivisild *et al.*, 2004
	Datoga	Tanzania	18	no	no	Kivisild *et al.*, 2004
	Afar	Ethiopia	16	no	no	Kivisild *et al.*, 2004
	Iraqw	Tanzania	12	no	no	Knight *et al.*, 2003
	Nuer	Sudan	11	no	no	Krings *et al.*, 1999
	Eritrean	Eritrea	8	no	no	Kivisild *et al.*, 2004
	Shilluk	Sudan	7	no	no	Krings *et al.*, 1999
	Mbuti	Democratic Republic of Congo	52	no	yes	Watson *et al.*, 1997; Quintana-Murci *et al.*, 2008
Total			**805**	**805**	**753**	

7.3 Mitochondrial haplogroup frequencies for Jamaica and each African coast used for the continental admixture analyse presented in Chapter 3.

Jamaica	L0a	L1b	L1c	L2a	L2b	L2c	L2d	L3b	L3d	L3e	L3f	L4/L7	non-L/U6	other L0'1'2	other L3	U6a
	0.05	0.12	0.08	0.27	0.05	0.05	0.06	0.04	0.15	0.06	0.01	0.00	0.02	0.03	0.02	0.01

All	L0a	L1b	L1c	L2a	L2b	L2c	L2d	L3b	L3d	L3e	L3f	L4/L7	non-L/U6	other L0'1'2	other L3	U6a
Bight of Benin	0.04	0.11	0.05	0.23	0.03	0.05	0.01	0.13	0.08	0.17	0.05	0	0.02	0	0.01	0.01
Bight of Biafra	0.06	0.07	0.24	0.15	0.03	0.02	0.02	0.07	0.05	0.17	0.07	0.01	0.02	0	0.01	0
East Africa	0.13	0.01	0.01	0.17	0.01	0	0	0.04	0.02	0.03	0.05	0.15	0.22	0.03	0.11	0.01
Gold Coast	0.02	0.08	0.08	0.33	0.03	0.05	0.01	0.1	0.06	0.17	0.04	0	0.01	0	0.01	0.01
Senegambia	0.03	0.13	0.04	0.17	0.08	0.17	0.02	0.1	0.08	0.08	0.03	0	0.03	0	0.02	0.02
sierra Leone	0.03	0.13	0.06	0.21	0.04	0.11	0.03	0.11	0.07	0.1	0.05	0	0.01	0	0.02	0.02
South East Africa	0.18	0.01	0.04	0.22	0.02	0.01	0	0.04	0.07	0.13	0.01	0.01	0.11	0.14	0.01	0
West-central Africa	0.11	0.04	0.32	0.13	0.02	0.02	0.01	0.02	0.04	0.17	0.06	0	0	0.02	0.01	0

Without Pygmies	L0a	L1b	L1c	L2a	L2b	L2c	L2d	L3b	L3d	L3e	L3f	L4/L7	non-L/U6	other L0'1'2	other L3	U6a
Bight of Benin	0.04	0.11	0.05	0.23	0.03	0.05	0.01	0.13	0.08	0.17	0.05	0	0.02	0	0.01	0.01
Bight of Biafra	0.06	0.09	0.13	0.17	0.03	0.03	0.02	0.08	0.06	0.2	0.08	0.01	0.02	0	0.02	0
East Africa	0.11	0.01	0.01	0.14	0.01	0	0	0.05	0.02	0.04	0.05	0.15	0.24	0.04	0.12	0.01
Gold Coast	0.02	0.08	0.08	0.33	0.03	0.05	0.01	0.1	0.06	0.17	0.04	0	0.01	0	0.01	0.01
Senegambia	0.03	0.13	0.04	0.17	0.08	0.17	0.02	0.1	0.08	0.08	0.03	0	0.03	0	0.02	0.02
sierra Leone	0.03	0.13	0.06	0.21	0.04	0.11	0.03	0.11	0.07	0.1	0.05	0	0.01	0	0.02	0.02
South East Africa	0.18	0.01	0.04	0.22	0.02	0.01	0	0.04	0.07	0.13	0.01	0.01	0.11	0.14	0.01	0
West-central Africa	0.12	0.05	0.29	0.14	0.03	0.03	0.01	0.02	0.05	0.18	0.06	0.01	0	0.02	0.01	0

Without Sahelian	L0a	L1b	L1c	L2a	L2b	L2c	L2d	L3b	L3d	L3e	L3f	L4/L7	non-L/U6	other L0'1'2	other L3	U6a
Bight of Benin	0.05	0.11	0.05	0.28	0.03	0.04	0.01	0.09	0.08	0.19	0.06	0.01	0.01	0	0.01	0.01
Bight of Biafra	0.06	0.07	0.29	0.15	0.02	0.02	0.02	0.05	0.04	0.18	0.06	0.01	0.01	0	0.01	0
East Africa	0.13	0.01	0.01	0.17	0.01	0	0	0.04	0.02	0.03	0.05	0.15	0.22	0.03	0.11	0.01
Gold Coast	0.01	0.08	0.08	0.33	0.03	0.05	0.01	0.09	0.07	0.18	0.04	0	0	0	0.01	0.01
Senegambia	0.05	0.21	0	0.13	0.1	0.18	0	0.13	0.05	0.05	0.05	0	0	0	0.05	0
sierra Leone	0.03	0.13	0.08	0.23	0.03	0.1	0.03	0.11	0.06	0.1	0.05	0	0	0	0.02	0.01
South East Africa	0.18	0.01	0.04	0.22	0.02	0.01	0	0.04	0.07	0.13	0.01	0.01	0.11	0.14	0.01	0
West-central Africa	0.11	0.04	0.32	0.13	0.02	0.02	0.01	0.02	0.04	0.17	0.06	0	0	0.02	0.01	0

Without Pygmies nor Sahelian	L0a	L1b	L1c	L2a	L2b	L2c	L2d	L3b	L3d	L3e	L3f	L4/L7	non-L/U6	other L0'1'2	other L3	U6a
Bight of Benin	0.05	0.11	0.05	0.28	0.03	0.04	0.01	0.09	0.08	0.19	0.06	0.01	0.01	0	0.01	0.01
Bight of Biafra	0.07	0.09	0.15	0.18	0.03	0.03	0.02	0.06	0.05	0.22	0.07	0.01	0.01	0	0.02	0
East Africa	0.11	0.01	0.01	0.14	0.01	0	0	0.05	0.02	0.04	0.05	0.15	0.24	0.04	0.12	0.01
Gold Coast	0.01	0.08	0.08	0.33	0.03	0.05	0.01	0.09	0.07	0.18	0.04	0	0	0	0.01	0.01
Senegambia	0.05	0.21	0	0.13	0.1	0.18	0	0.13	0.05	0.05	0.05	0	0	0	0.05	0
sierra Leone	0.03	0.13	0.08	0.23	0.03	0.1	0.03	0.11	0.06	0.1	0.05	0	0	0	0.02	0.01
South East Africa	0.18	0.01	0.04	0.22	0.02	0.01	0	0.04	0.07	0.13	0.01	0.01	0.11	0.14	0.01	0
West-central Africa	0.12	0.05	0.29	0.14	0.03	0.03	0.01	0.02	0.05	0.18	0.06	0.01	0	0.02	0.01	0

7.4 Number of shared haplotypes between the different regions considered in 3.1.2. Numbers in the diagonal are the number of different haplotypes per region. In brackets is the proportion of the Jamaican haplotypes that are shared with each of the African regions considered.

	Jamaica	Bight of Benin	Bight of Biafra	East Africa	Gold Coast	Senegambia	Sierra Leone	South East Africa	West-central Africa
Jamaica	231	41 (18%)	60 (26%)	21 (9%)	43 (19%)	35 (15%)	41 (18%)	23 (10%)	38 (16%)
Bight of Benin		249	77	23	38	58	60	22	38
Bight of Biafra			1146	46	91	74	83	67	108
East Africa				482	15	25	19	20	32
Gold Coast					300	48	61	27	26
Senegambia						439	89	18	29
Sierra Leone							468	23	23
South East Africa								291	39
West-central Africa									500

7.5 Admixture mapping results with the 11 excluded African-American athletes reintroduced

After inferring the genome-wide African ancestry in African-American athletes and controls in Chapter 5, 11 athletes had less than 5% African ancestry. To make the results more comparable, these 11 athletes were excluded from the analyses presented in that study. Presented here are the same methods applied to the complete dataset from the African-American cohort.

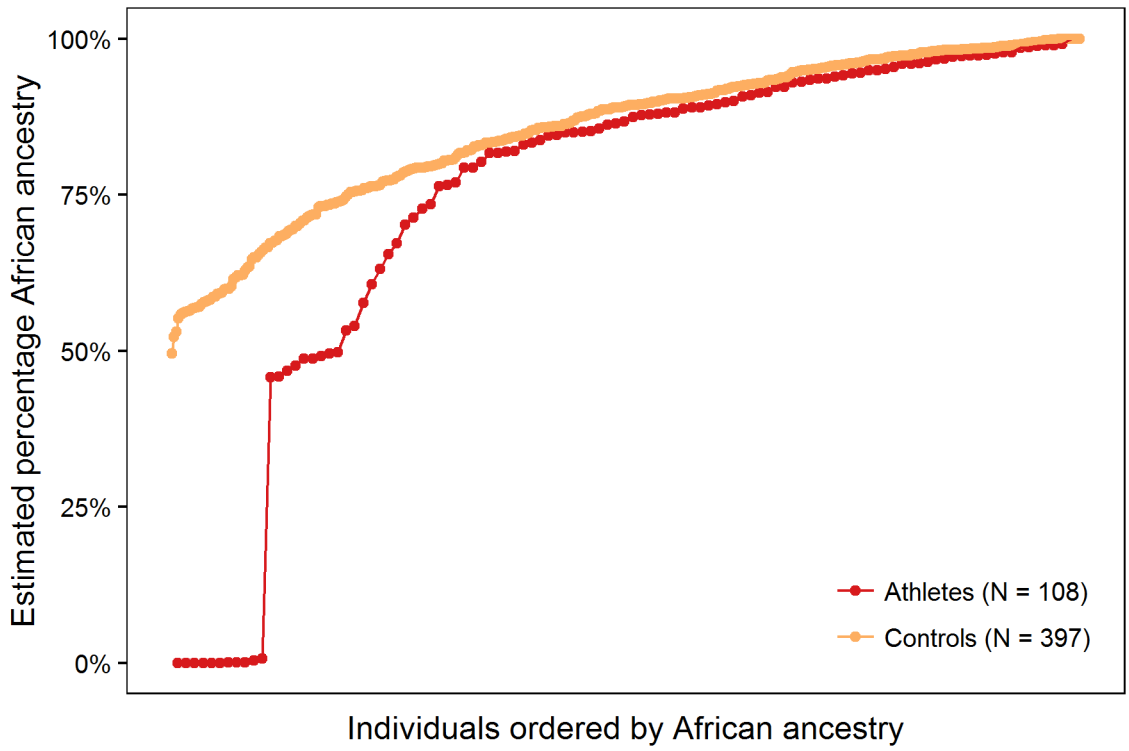


Figure 7.1 Ordered line plot of genome-wide African ancestry for African-Americans including the 11 athletes excluded in Chapter 5. Each point represents an individual. The difference between athletes and controls is exaggerated due to the inclusion of the 11 athletes with less than 5% African ancestry.

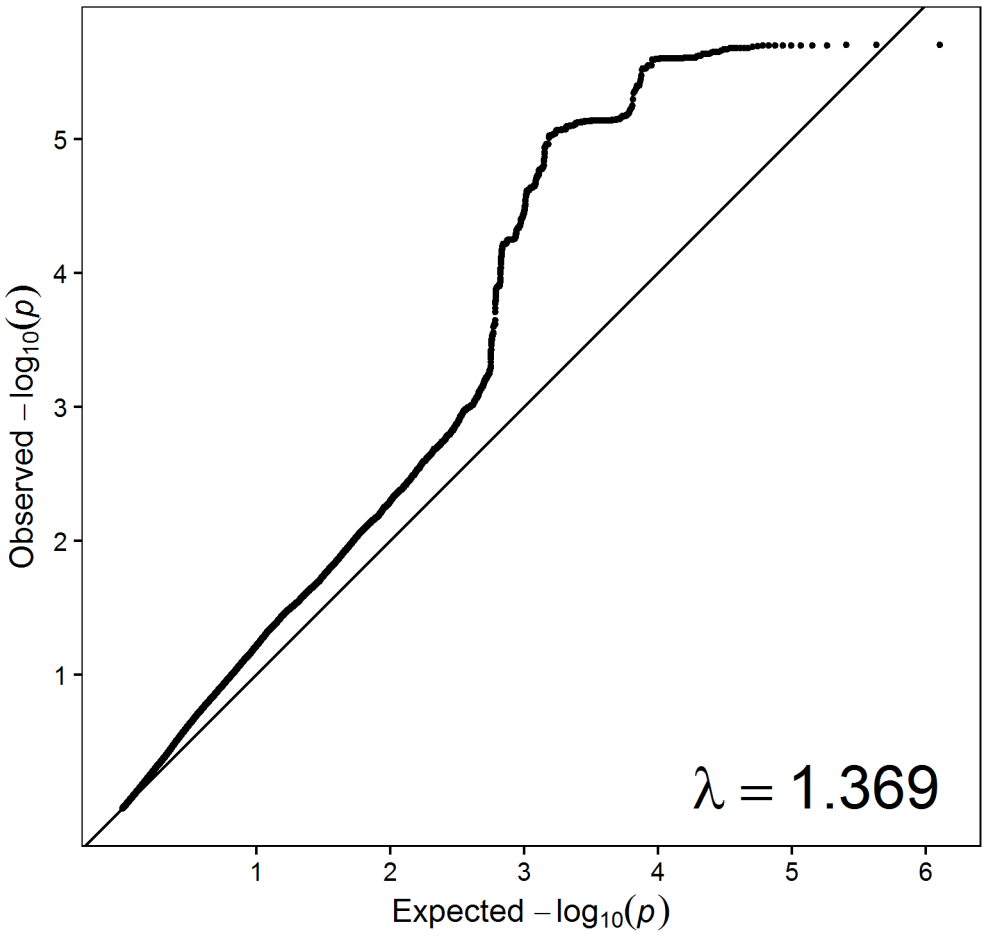


Figure 7.2 Q-Q plot for admixture mapping in African-Americans, completed with the inclusion of 11 athletes with less than 5% African ancestry. Admixture mapping for genome-wide ancestry association testing for local African ancestry. The straight line represents the line of perfect correspondence. The lambda of 1.369 suggests that there is a large amount of population stratification not taken into account. The pattern is largely similar to that shown in Figure 5.4A despite the points resting above the line. This shift also explains why only a single result crossed the line at the extreme end of the right tail. These results suggest that the 11 excluded athletes introduced additional population stratification to the admixture mapping.

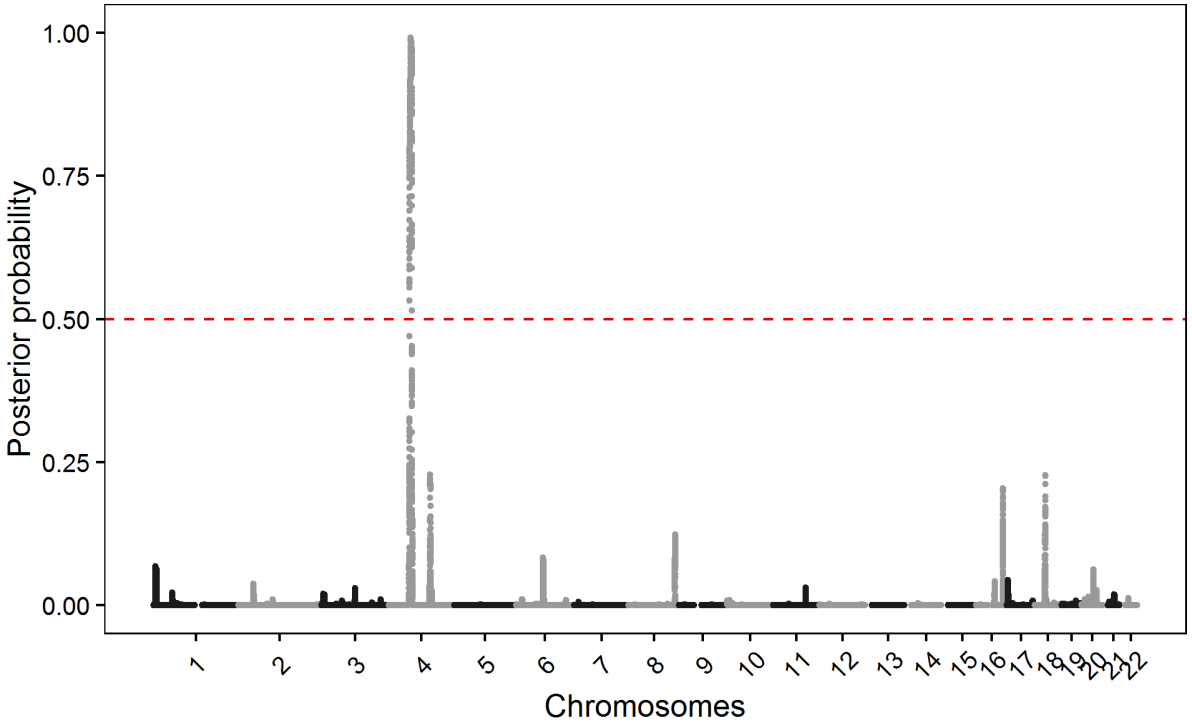


Figure 7.3 Manhattan plot of posterior probabilities from admixture mapping in African-Americans with the 11 excluded athletes reintroduced. Each point represents the posterior probability from a single-locus association test. Candidate admixture mapping peaks were defined as those with a probability of greater than 0.5, represented here as a dashed horizontal line. The peak on chromosome 4 corresponds to the same genomic location found in Figure 5.5, though much more significant. These results are expected. The excluded athletes were predominantly European, meaning that any association between athlete status and European ancestry will be strengthened by their inclusion.

References

1000 Genomes Project Consortium 2010. A map of human genome variation from population-scale sequencing. *Nature*, 467, 1061-1073

1000 Genomes Project Consortium 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56-65

Acosta, M.T., Swanson, J., Stehli, A., Molina, B.S.G., Martinez, A.F., Arcos-Burgos, M., & Muenke, M. 2016. ADGRL3 (LPHN3) variants are associated with a refined phenotype of ADHD in the MTA study. *Molecular Genetics & Genomic Medicine*, 4, 540-547

Ahmetov, I. I. & Fedotovskaya, O. N. 2015, Current Progress in Sports Genomics, *Advances in Clinical Chemistry*, 70, 247-314

Aiken, W.D. 2011. Historical determinants of contemporary attributes of African descendants in the Americas: The androgen receptor holds the key. *Medical hypotheses*, 77, 1121-1124

Alexander, D.H., Novembre, J., & Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655-1664

Allard, M.W., Polanskey, D., Miller, K., Wilson, M.R., Monson, K.L., & Budowle, B. 2005. Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. *Forensic Science International*, 148, 169-179

Alvarez-Iglesias, V., Mosquera-Miguel, A., Cerezo, M., Quintans, B., Zarrabeitia, M.T., Cusco, I., Lareu, M.V., Garcia, O., Perez-Jurado, L., Carracedo, A., & Salas, A. 2009. New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS One*, 4, e5112

Ama, P.F., Lagasse, P., Bouchard, C., & Simoneau, J.A. 1990. Anaerobic performances in black and white subjects. *Medicine and Science in Sports and Exercise*, 22, 508-511

American Society of Hematology, *Statement on screening for sickle cell trait and athletic participation.* 2012. *http://www.hematology.org/Advocacy/Statements/2650.aspx*. Accessed on: 01-09-2016

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., & Howell, N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, 23, 147

Arcos-Burgos, M. & Muenke, M. 2010. Toward a better understanding of ADHD: LPHN3 gene variants and the susceptibility to develop ADHD. *ADHD Attention Deficit and Hyperactivity Disorders*, 2, 139-147

Ardlie, K.G., Lunetta, K.L. & Seielstad, M. 2002. Testing for population subdivision and association in four case-control studies. *The American Journal of Human Genetics*, 71, 304-311

Assimes, T.L., Knowles, J.W., Basu, A., Iribarren, C., Southwick, A., Tang, H., Absher, D., Li, J., Fair, J.M., Rubin, G.D., Sidney, S., Fortmann, S.P., Go, A.S., Hlatky, M.A., Myers, R.M., Risch, N., & Quertermous, T. 2008. Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic ADVANCE study. *Human Molecular Genetics*, 17, 2320-2328

Baharian, S., Barakatt, M., Gignoux, C.R., Shringarpure, S., Errington, J., Blot, W.J., Bustamante, C.D., Kenny, E.E., Williams, S.M., Aldrich, M.C., & Gravel, S. 2016. The Great Migration and African-American Genomic Diversity. *PLoS Genetics*, 12, e1006059

Baker, J.S., McCormick, M.C., & Robergs, R.A. 2010. Interaction among skeletal muscle metabolic energy systems during intense exercise. *Journal of nutrition and metabolism*, 2010, 1-13

Balaresque, P., Bowden, G.R., Adams, S.M., Leung, H.Y., King, T.E., Rosser, Z.H., Goodwin, J., Moisan, J.P., Richard, C., Millward, A., Demaine, A.G., Barbujani, G., Previdere, C., Wilson, I.J., Tyler-Smith, C., & Jobling, M.A. 2010. A predominantly neolithic origin for European paternal lineages. *PLoS Biology*, 8, e1000285

Balding, D.J. 2006. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, 7, 781-791

Ballard, J.W.O. & Whitlock, M.C. 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, 13, 729-744

Ballas, S.K. 2011. Defining the phenotypes of sickle cell disease. *Hemoglobin*, 35, 511-519

Bandelt, H.J., Quintana-Murci, L., Salas, A., & Macaulay, V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. *American Journal of Human Genetics*, 71, 1150-1160

Bandelt, H. J., Salas, A., & Bravi, C. Problems in FBI mtDNA Database. *Science*, 305, 1402-1404

Bandelt, H.J., Salas, A., & Lutz-Bonengel, S. 2004b. Artificial recombination in forensic mtDNA population databases. *International Journal of Legal Medicine*, 118, 267-273

Baran, Y., Pasaniuc, B., Sankararaman, S., Torgerson, D.G., Gignoux, C., Eng, C., Rodriguez-Cintron, W., Chapela, R., Ford, J.G., Avila, P.C., Rodriguez-Santana, J., Burchard, E.G., & Halperin, E. 2012. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics*, 28, 1359-1367

Batini, C., Coia, V., Battaggia, C., Rocha, J., Pilkington, M.M., Spedini, G., Comas, D., Destro-Bisol, G., & Calafell, F. 2007. Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa. *Molecular Phylogenetics and Evolution*, 43, 635-644

Batini, C., Lopes, J., Behar, D.M., Calafell, F., Jorde, L.B., van der Veen, L., Quintana-Murci, L., Spedini, G., Destro-Bisol, G., & Comas, D. 2011. Insights into the Demographic History of African Pygmies from Complete Mitochondrial Genomes. *Molecular Biology and Evolution*, 28, 1099-1110

Bean, R.N. 1972. The British Trans-Atlantic slave trade 1650-1775. *The Journal of Economic History,* 409-411

Beckles, H.M. 1989. *White servitude and Black slavery in Barbados, 1627-1715*, 1st ed. Knoxville, TN, University of Tennessee Press.

Bedoya, G., Montoya, P., Garcia, J., Soto, I., Bourgeois, S., Carvajal, L., Labuda, D., Alvarez, V., Ospina, J., Hedrick, P.W., & Ruiz-Linares, A. 2006. Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate. *Proceedings of the National Academy of Sciences*, 103, 7234-7239

Behar, D.M., Rosset, S., Blue-Smith, J., Balanovsky, O., Tzur, S., Comas, D., Mitchell, R.J., Quintana-Murci, L., Tyler-Smith, C., & Wells, R.S. 2007. The genographic project public participation mitochondrial DNA database. *PLoS Genetics*, 3, 1083-1095

Behrendt, S.D., Eltis, D. & Richardson, D. 2001. The Costs of Coercion: African Agency in the Pre-Modern Atlantic World. *The Economic History Review*, 54, 454-476

Beilock, S.L., Rydell, R.J., & McConnell, A.R. 2007. Stereotype threat and working memory: mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, 136, 256

Beleza, S., Gusmao, L., Amorim, A., Carracedo, A., & Salas, A. 2005. The genetic legacy of western Bantu migrations. *Human Genetics*, 117, 366-375

Benn-Torres, J., Bonilla, C., Robbins, C.M., Waterman, L., Moses, T.Y., Hernandez, W., Santos, E.R., Bennett, F., Aiken, W., Tullock, T., Coard, K., Hennis, A., Wu, S., Nemesure, B., Leske, M.C., Freeman, V., Carpten, J., & Kittles, R.A. 2008. Admixture and population stratification in African Caribbean populations. *Annals of Human Genetics*, 72, 90-98

Berlin, I. 2003. *Generations of captivity: a history of African-American slaves* Cambridge, Mass, Belknap Press of Harvard University Press.

Bertorelle, G. & Excoffier, L. 1998. Inferring admixture proportions from molecular data. *Molecular Biology and Evolution*, 15, 1298-1311

Bertram, A., 04-02-2012. Jamaica's Road to the Olympics, *The Jamaican Gleaner*

Bhatia, G., Patterson, N., Pasaniuc, B., Zaitlen, N., Genovese, G., Pollack, S., Mallick, S., Myers, S., Tandon, A., & Spencer, C. 2011. Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. *The American Journal of Human Genetics*, 89, 368-381

Bhatia, G., Tandon, A., Patterson, N., Aldrich, M.C., Ambrosone, C.B., Amos, C., Bandera, E.V., Berndt, S.I., Bernstein, L., Blot, W.J., Bock, C.H., Caporaso, N., Casey, G., Deming, S.L., Diver, W.R., Gapstur, S.M., Gillanders, E.M., Harris, C.C., Henderson, B.E., Ingles, S.A., Isaacs, W., de Jager, P.L., John, E.M., Kittles, R.A., Larkin, E., McNeill, L.H., Millikan, R.C., Murphy, A., Neslund-Dudas, C., Nyante, S., Press, M.F., Rodriguez-Gil, J.L., Rybicki, B.A., Schwartz, A.G., Signorello, L.B., Spitz, M., Strom, S.S., Tucker, M.A., Wiencke, J.K., Witte, J.S., Wu, X., Yamamura, Y., Zanetti, K.A., Zheng, W., Ziegler, R.G., Chanock, S.J., Haiman, C.A., Reich, D., & Price, A.L. 2014. Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. *American Journal of Human Genetics*, 95, 437-444

Bonilla, C., Mason, T., Long, L.R., Ahaghotu, C., Chen, W.D., Zhao, A.Q., Coulibaly, A., Bennett, F., Aiken, W., Tullock, T., Coard, K., Freeman, V., & Kittles, R.A. 2006. E-cadherin Polymorphisms and haplotypes influence risk for prostate cancer. *Prostate*, 66, 546-556

Bosco, C., Tihanyit, J., & Viru, A. 1996. Relationships between field fitness test and basal serum testosterone and cortisol levels in soccer players. *Clinical Physiology*, 16, 317-322

Boucard, A.A., Maxeiner, S., & Sudhof, T.C. 2014. Latrophilins function as heterophilic cell-adhesion molecules by binding to teneurins: regulation by alternative splicing. *Journal of Biological Chemistry*, 289, 387-402

Bouchard, C. 2015. Exercise genomics - a paradigm shift is needed: a commentary. *British Journal of Sports Medicine*, 49, 1492-1496

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. *International Journal of Legal Medicine*, 118, 294-306

Browning, B.L. & Browning, S.R. 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Genetic Epidemiology*, 31, 606

Brucato, N., Cassar, O., Tonasso, L., Tortevoye, P., Migot-Nabias, F., Plancoulaine, S., Guitard, E., Larrouy, G., Gessain, A., & Dugoujon, J.M. 2010. The imprint of the Slave Trade in an African American population: mitochondrial DNA, Y chromosome and HTLV-1 analysis in the Noir Marron of French Guiana. *BMC Evolutionary Biology*, 10, 314

Brutsaert, T.D., Parra, E.J., Shriver, M.D., Gamboa, A., Palacios, J.A., Rivera, M., Rodriguez, I., & Leon-Velarde, F. 2003. Spanish genetic admixture is associated with larger VO2 max decrement from sea level to 4,338 m in Peruvian Quechua. *Journal of Applied Physiology*, 95, 519-528

Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., Williams, S., Froment, A., Bodo, J.M., Wambebe, C., Tishkoff, S.A., & Bustamante, C.D. 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences*, 107, 786-791

Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D., & Mountain, J.L. 2015. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *The American Journal of Human Genetics*, 96, 37-53

Burnard, T. 1996. European Migration to Jamaica, 1655-1780. *The William and Mary Quarterly*, 53, 769-796

Burnard, T. 1998, "The Sexual Life of an Eighteenth-Century Jamaican Slave Overseer," *In Sex and Sexuality in Early America*, M. D. Smith, ed., New York: pp. 163-189.

Burnard, T. 1999. The countrie continues sicklie': white mortality in Jamaica, 1655-1780. *Social History of Medicine*, 12, 45-72

Burnard, T. 2004. Passengers only: The extent and significance of absenteeism in eighteenth century Jamaica. *Atlantic Studies*, 1, 178-195

Burnard, T. 2006. "Rioting in goatish embraces": Marriage and improvement in early British Jamaica. *The History of the Family*, 11, 185-197

Burnard, T. & Morgan, K. 2001. The dynamics of the slave market and slave purchasing patterns in Jamaica, 1655-1788. *The William and Mary Quarterly*, 58, 205-228

Bussmann, G., Alfermann, D., & Hackfort, D. 1994. Drop-out and the female athlete: A study with track-and-field athletes. *Psycho-social issues and interventions in elite sport*, 89-128

Campbell, M.C. & Tishkoff, S.A. 2008. African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics*, 9, 403-433

Cann, R.L., Stoneking, M., & Wilson, A.C. 1987a. Disputed African Origin of Human-Populations - Reply. *Nature*, 329, 111-112

Cann, R.L., Stoneking, M., & Wilson, A.C. 1987b. Mitochondrial-Dna and human evolution. *Nature*, 325, 31-36

Cantagrel, V., Lefeber, D.J., Ng, B.G., Guan, Z., Silhavy, J.L., Bielas, S.L., Lehle, L., Hombauer, H., Adamowicz, M., & Swiezewska, E. 2010. SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder. *Cell*, 142, 203-217

Cardon, L.R. & Palmer, L.J. 2003. Population stratification and spurious allelic association. *Lancet*, 361, 598-604

Cassidy, F.G. 1966. Multiple etymologies in Jamaican creole. *American Speech*, 41, 211-215

Cassidy, F.G. & Le Page, R.B. 2002. *Dictionary of Jamaican English*, 2nd ed. Kingston, Jamaica, University of the West Indies Press.

Castri, L., Tofanelli, S., Garagnani, P., Bini, C., Fosella, X., Pelotti, S., Paoli, G., Pettener, D., & Luiselli, D. 2009. mtDNA Variability in Two Bantu-Speaking Populations (Shona and Hutu) From Eastern Africa: Implications for Peopling and Migration Patterns in Sub-Saharan Africa. *American Journal of Physical Anthropology*, 140, 302-311

Castro, M.G., Terrados, N., Reguero, J.R., Alvarez, V., & Coto, E. 2007. Mitochondrial haplogroup T is negatively associated with the status of elite endurance athlete. *Mitochondrion*, 7, 354-357

Cerezo, M., Achilli, A., Olivieri, A., Perego, U.A., Gomez-Carballa, A., Brisighelli, F., Lancioni, H., Woodward, S.R., Lopez-Soto, M., Carracedo, A., Capelli, C., Torroni, A., & Salas, A. 2012. Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Research*, 22, 821-826

Cerny, V., Hajek, M., Bromova, M., Cmejla, R., Diallo, I., & Brdicka, R. 2006. mtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Human Biology*, 78, 9-27

Cerny, V., Hajek, M., Cmejla, R., Bruzek, J., & Brdicka, R. 2004. mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. *Annals of Human Biology*, 31, 554-569

Cerny, V., Salas, A., Hajek, M., Zaloudkova, M., & Brdicka, R. 2007. A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: A history revealed by the mitochondrial DNA genome. *Annals of Human Genetics*, 71, 433-452

Chakraborty, R., Ferrell, R.E., Stern, M.P., Haffner, S.M., Hazuda, H.P., & Rosenthal, M. 1986. Relationship of Prevalence of Non-Insulin-Dependent Diabetes-Mellitus to Amerindian Admixture in the Mexican-Americans of San-Antonio, Texas. *Genetic Epidemiology*, 3, 435-454

Chakraborty, R. and Weiss, K.M. 1988. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proceedings of the National Academy of Sciences*, 85, 9119-9123

Charchar, F.J., Bloomer, L.D.S., Barnes, T.A., Cowley, M.J., Nelson, C.P., Wang, Y.Z., Denniff, M., Debiec, R., Christofidou, P., Nankervis, S., Dominiczak, A.F., Bani-Mustafa, A., Balmforth, A.J., Hall, A.S., Erdmann, J., Cambien, F., Deloukas, P., Hengstenberg, C., Packard, C., Schunkert, H., Ouwehand, W.H., Ford, I., Goodall, A.H., Jobling, M.A., Samani, N.J., & Tomaszewski, M. 2012. Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. *Lancet*, 379, 915-922

Cheetham, M.E., Boobis, L.H., Brooks, S., & Williams, C. 1986. Human muscle metabolism during sprint running. *Journal of Applied Physiology*, 61, 54-60

Chen, M., Yang, C., Li, C., Hou, L., Chen, X., & Zhao, H. Admixture mapping analysis in the context of GWAS with GAW18 data. *BMC proceedings*, 8, pp. S3

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. *American Journal of Human Genetics*, 66, 1362-1383

Cheng, C.Y., Schache, M., Ikram, M.K., Young, T.L., Guggenheim, J.A., Vitart, V., MacGregor, S., Verhoeven, V.J.M., Barathi, V.A., & Liao, J. 2013. Nine loci for ocular axial length identified through genome-wide association studies, including shared loci with refractive error. *The American Journal of Human Genetics*, 93, 264-277

Cheng, Y.C., O'Connell, J.R., Cole, J.W., Stine, O.C., Dueker, N., McArdle, P.F., Sparks, M.J., Shen, J., Laurie, C.C., Nelson, S., Doheny, K.F., Ling, H., Pugh, E.W., Brott, T.G., Brown, R.D., Meschia, J.F., Nalls, M., Rich, S.S., Worrall, B., Anderson, C.D., Biffi, A., Cortellini, L., Furie, K.L., Rost, N.S., Rosand, J., Manolio, T.A., Kittner, S.J., & Mitchell, B.D. 2011. Genome-wide association analysis of ischemic stroke in young adults. *G3-Genes Genomes Genetics*, 1, 505-513

Chiang, C.W., Gajdos, Z.K., Korn, J.M., Kuruvilla, F.G., Butler, J.L., Hackett, R., Guiducci, C., Nguyen, T.T., Wilks, R., Forrester, T. & Haiman, C.A. 2010. Rapid assessment of genetic ancestry in populations of unknown origin by genome-wide genotyping of pooled samples. *PLoS Genetics*, 6, e1000866

Chinnery, P.F., Elliott, H.R., Syed, A., & Rothwell, P.M. 2010. Mitochondrial DNA haplogroups and risk of transient ischaemic attack and ischaemic stroke: a genetic association study. *The Lancet Neurology*, 9, 498-503

Chirico, E.N., Faes, C., Connes, P., Canet-Soulas, E., Martin, C., & Pialoux, V. 2016. Role of exercise-induced oxidative stress in sickle cell trait and disease. *Sports Medicine*, 46, 629-639

Churchhouse, C. & Marchini, J. 2013. Multiway admixture deconvolution using phased or unphased ancestral panels. *Genetic Epidemiology*, 37, 1-12

Coale, A.J. & Demeny, P. 1967. *Methods of estimating basic demographic measures from incomplete data.* New York: United Nations

Coelho, M., Sequeira, F., Luiselli, D., Beleza, S., & Rocha, J. 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evolutionary Biology*, 9, 80

Coh, M., Babic, V., & Mackala, K. 2010. Biomechanical, Neuro-muscular and Methodical Aspects of Running Speed Development. *Journal of Human Kinetics*, 26, 73-81

Coia, V., Destro-Bisol, G., Verginelli, F., Battaggia, C., Boschi, I., Cruciani, F., Spedini, G., Comas, D., & Calafell, F. 2005. Brief communication: mtDNA variation in North Cameroon: Lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. *American Journal of Physical Anthropology*, 128, 678-681

Collins, D. 1971. *Practical Rules for the Management and Medical Treatment of Negro Slaves: In the Sugar Colonies* Freeport, New York, Books for Libraries Press

Collins, M.F. & Buller, J.R. 2003. Social exclusion from high-performance sport are all talented young sports people being given an equal opportunity of reaching the Olympic podium? *Journal of Sport & Social Issues*, 27, 420-442

Comuzzie, A.G., Cole, S.A., Laston, S.L., Voruganti, V.S., Haack, K., Gibbs, R.A., & Butte, N.F. 2012. Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. *PLoS One*, 7, e51954

Consortium on Asthma among African-Ancestry Population in the Americas (CAAPA). http://www.caapaproject.org.2010. Accessed on: 01-09-2016

Consortium, T.Y.C. 2002. A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Research*, 12, 339-348

Cooper, R., Cutler, J., Desvigne-Nickens, P., Fortmann, S.P., Friedman, L., Havlik, R., Hogelin, G., Marler, J., McGovern, P., Morosco, G. & Mosca, L. 2000. Trends and disparities in coronary heart disease, stroke, and other cardiovascular diseases in the United States. *Circulation*, 102, 3137-3147

Cooper, R. & Rotimi, C. 1994. EDITORIAL REVIEW: Hypertension in populations of West African origin: is there a genetic predisposition? *Journal of hypertension*, 12, 215-228

Cooper, R., Rotimi, C., Ataman, S., McGee, D., Osotimehin, B., Kadiri, S., Muna, W., Kingue, S., Fraser, H., & Forrester, T. 1997. The prevalence of hypertension in seven populations of west African origin. *American Journal of Public Health*, 87, 160-168

Corach, D., Lao, O., Bobillo, C., van Der Gaag, K., Zuniga, S., Vermeulen, M., van Duijn, K., Goedbloed, M., Vallone, P.M., Parson, W., de Knijff, P., & Kayser, M. 2010. Inferring Continental Ancestry of Argentineans from Autosomal, Y-Chromosomal and Mitochondrial DNA. *Annals of Human Genetics*, 74, 65-76

Costa, A.M., Breitenfeld, L., Silva, A.J., Pereira, A., Izquierdo, M., & Marques, M.C. 2012. Genetic Inheritance Effects on Endurance and Muscle Strength An Update. *Sports Medicine*, 42, 449-458

Costill, D.L., Daniels, J., Evans, W., Fink, W., Krahenbuhl, G., & Saltin, B. 1976. Skeletal-Muscle Enzymes and Fiber Composition in Male and Female Track Athletes. *Journal of Applied Physiology*, 40, 149-154

Craton, M. 1978. Hobbesian or Panglossian? The two extremes of slave conditions in the British Caribbean, 1783 to 1834. *The William and Mary Quarterly: A Magazine of Early American History,* 324-356

Creanza, N. & Feldman, M.W. 2016. Worldwide genetic and cultural change in human evolution. *Current Opinion in Genetics & Development*, 41, 85-92

Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Colomb, E.B., Zaharova, B., Lavinha, J., Vona, G., Aman, R., Cali, F., Akar, N., Richards, M., Torroni, A., Novelletto, A., & Scozzari, R. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *American Journal of Human Genetics*, 74, 1014-1022

Cruciani, F., Santolamazza, P., Shen, P.D., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., Wallace, D.C., Oefner, P.J., Torroni, A., Cavalli-Sforza, L.L., Scozzari, R., & Underhill, P.A. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *American Journal of Human Genetics*, 70, 1197-1214

Curtin, P. 1969. *The Atlantic Slave Trade: A Census* Madison, WI; Milwaukee, WI; London, UK, The University of Wisconsin Press.

Curtin, P.D. 1992. The slavery hypothesis for hypertension among African Americans: the historical evidence. *American Journal of Public Health*, 82, 1681-1686

Damon, A. 1966. Negro-white differences in pulmonary function (vital capacity, timed vital capacity, and expiratory flow rate). *Human Biology*, 38, 381-393

Daley, M.A., Felix, G., & Biewener, A.A. 2007. Running stability is enhanced by a proximo-distal gradient in joint neuromechanical control. *Journal of Experimental Biology*, 210, 383-394

Darlu, P. & Tassy, P. 1987. Disputed African Origin of Human-Populations. *Nature*, 329, 111

Davids, K. & Baker, J. 2007. Genes, environment and sport performance - Why the nature-nurture dualism is no longer relevant. *Sports Medicine*, 37, 961-980

de Filippo, C., Barbieri, C., Whitten, M., Mpoloka, S.W., Gunnarsdottir, E.D., Bostoen, K., Nyambe, T., Beyer, K., Schreiber, H., de Knijff, P., Luiselli, D., Stoneking, M., & Pakendorf, B. 2011. Y-Chromosomal Variation in Sub-Saharan Africa: Insights Into the History of Niger-Congo Groups. *Molecular Biology and Evolution*, 28, 1255-1269

de Moor, M.H.M., Spector, T.D., Cherkas, L.F., Falchi, M., Hottenga, J.J., Boomsma, D.I., & de Geus, E.J.C. 2007. Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Research and Human Genetics*, 10, 812-820

Deason, M.L., Salas, A., Newman, S.P., Macaulay, V.A., st A Morrison, E.Y. & Pitsiladis, Y.P. 2012. Interdisciplinary approach to the demography of Jamaica. *BMC evolutionary biology*, 12, 24

Delaney, J.T., Jeff, J.M., Brown, N.J., Pretorius, M., Okafor, H.E., Darbar, D., Roden, D.M., & Crawford, D.C. 2012. Characterization of Genome-Wide Association-Identified Variants for Atrial Fibrillation in African Americans. *PLoS One*, 7, e32338

DeNavas-Walt, C., Proctor, B.D., & Smith, J.C. 2012. *Income, Poverty, and Health Insurance Coverage in the United States: 2011* Washington, DC, U.S. Goverment Printing Office

Destro-Bisol, G., Coia, V., Boschi, I., Verginelli, F., Caglia, A., Pascali, V., Spedini, G., & Calafell, F. 2004. The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *American Naturalist*, 163, 212-226

di Prampero, P.E., Fusi, S., Sepulcri, L., Morin, J.B., Belli, A., & Antonutto, G. 2005. Sprint running: a new energetic approach. *Journal of Experimental Biology*, 208, 2809-2816

Diamond, J. 1991. The saltshaker's curse. *Natural History*, 10, 20-26

Diegoli, T.M., Irwin, J.A., Just, R.S., Saunier, J.L., O'Callaghan, J.E., & Parsons, T.J. 2009. Mitochondrial control region sequences from an African American population sample. *Forensic Science International-Genetics*, 4, E45-E52

DiMauro, S. & Schon, E.A. 2003. Mechanisms of disease: Mitochondrial respiratory-chain diseases. *New England Journal of Medicine*, 348, 2656-2668

Dosil, J. 2006, "The Psychology of Athletics," Wiley Online Library, pp. 265-284.

Druzhevskaya, A.M., Ahmetov, I.I., Astratenkova, I.V., & Rogozkin, V.A. 2008. Association of the ACTN3 R577X polymorphism with power athlete status in Russians. *European Journal of Applied Physiology*, 103, 631-634

Dubner, S.J. 2005. Toward a unified theory of Black America. *New York Times Magazine*, 20

Duffield, R., Dawson, B., & Goodman, C. 2004. Energy system contribution to 100-m and 200-m track running events. *Journal of Science and Medicine in Sport*, 7, 302-313

Dunn, R.S. 2000. *Sugar and slaves: the rise of the planter class in the English West Indies, 1624-1713* Chapel Hill, Published for the Omohundro Institute of Early American History and Culture, Williamsburg, Virginia, by the University of North Carolina Press

Dupanloup, I. & Bertorelle, G. 2001. Inferring admixture proportions from molecular data: Extension to any number of parental populations. *Molecular Biology and Evolution*, 18, 672-675

Durand, E.Y., Do, C.B., Mountain, J.L. & Macpherson, J.M. 2014. Ancestry composition: a novel, efficient pipeline for ancestry deconvolution. *bioRxiv*, 010512

Eaves, L.J., Martin, N.G. & Heath, A.C. 1990. Religious affiliation in twins and their parents: Testing a model of cultural inheritance. *Behavior genetics*, 20, 1-22

Eblen, J.E. 1974. New estimates of the vital rates of the United States black population during the nineteenth century. *Demography*, 11, 301-319

Egerton, D.R. 1996. Markets without a market revolution: Southern planters and capitalism. *Journal of the Early Republic*, 16, 207-221

Elbein, S.C., Das, S.K., Hallman, D.M., Hanis, C.L., & Hasstedt, S.J. 2009. Genome-Wide Linkage and Admixture Mapping of Type 2 Diabetes in African American Families From the American Diabetes Association GENNID (Genetics of NIDDM) Study Cohort. *Diabetes*, 58, 268-274

Elferink-Gemser, M.T., Jordet, G., Coelho-E-Silva, M., & Visscher, C. 2011. The marvels of elite sports: how to get there? *British Journal of Sports Medicine*, 45, 683-684

Eltis, D., 1983. Free and coerced transatlantic migrations: Some comparisons. The American Historical Review, 251-280

Eltis, D. Estimates Database. http://www.slavevoyages.org/ Voyages: The Trans-Atlantic Slave Trade Database. 2009. Accessed on: 12-10-2015

Eltis, D. Construction of the trans-Atlantic slave trade database: sources and methods. http://www.slavevoyages.org . 2010. Accessed on: 12-5-2016

Eltis, D. & Richardson, D. The Transatlantic Slave Trade Database. http://www.slavevoyages.org . 2010. Yale University Press

Ely, B., Wilson, J.L., Jackson, F.L.C., & Jackson, B.A. 2006. African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. *BMC Biology*, 4, 34

Emery, L.S., Magnaye, K.M., Bigham, A.W., Akey, J.M., & Bamshad, M.J. 2015. Estimates of continental ancestry vary widely among individuals with the same mtDNA haplogroup. *The American Journal of Human Genetics*, 96, 183-193

Emmer, P. 2000. 'A spirit of independence' or lack of education for the market? Freedmen and Asian indentured labourers in the post-emancipation Caribbean, 1834-1917. *Slavery & Abolition*, 21, 150-168

Emmer, P.C. 1995. Indentured Labor, Caribbean Sugar - Chinese and Indian Migrants to the British-West-Indies, 1838-1918. *Albion*, 27, 184-185

Engerman, S.L. 1976. Some economic and demographic comparisons of slavery in United-States and British-West-Indies. *Economic History Review*, 29, 258-275

English, C.A., Merson, S., & Keer, J.T. 2006. Use of elemental analysis to determine comparative performance of established DNA quantification methods. *Analytical Chemistry*, 78, 4630-4633

Ensembl. Ensembl Human MapView release 43. 2014. Accessed on: 5-6-2014

Entine, J. 2000. Breaking the taboo (Why Black athletes dominate sports). *Index on Censorship*, 29, 62-64

Ericsson, K.A. 2007. Deliberate practice and the modifiability of body and mind: Toward a science of the structure and acquisition of expert and elite performance. *International Journal of Sport Psychology*, 38, 4-34

Ericsson, K.A., Nandagopal, K., & Roring, R.W. 2009. Toward a Science of Exceptional Achievement Attaining Superior Performance through Deliberate Practice. *Longevity, Regeneration, and Optimal Health: Integrating Eastern and Westen Perspectives*, 1172, 199-217

Ewis, A.A., Lee, J., Naroda, T., Sano, T., Kagawa, S., Iwamoto, T., Shinka, T., Shinohara, Y., Ishikawa, M., Baba, Y., & Nakahori, Y. 2006. Prostate cancer incidence varies among males from different Y-chromosome lineages. *Prostate Cancer and Prostatic Diseases*, 9, 303-309

Excoffier, L. & Langaney, A. 1989. Origin and Differentiation of Human Mitochondrial-Dna. *American Journal of Human Genetics*, 44, 73-85

Excoffier, L. & Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564-567

Eyerman, R. 2001. *Cultural trauma: Slavery and the formation of African American identity* Cambridge University Press

Eynon, N., Duarte, J.A., Oliveira, J., Sagiv, M., Yamin, C., Meckel, Y., Sagiv, M., & Goldhammer, E. 2009. ACTN3 R577X Polymorphism and Israeli Top-level Athletes. *International Journal of Sports Medicine*, 30, 695-698

Falush, D., Stephens, M., & Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567-1587

Farley, R. 1965. The demographic rates and social institutions of the nineteenth-century negro population: A stable population analysis. *Demography*, 2, 386-398

Farley, R. 1968. The urbanization of negroes in the United States. *Journal of Social History*, 1, 241-258

Fearon, J.D. 2003. Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8, 195-222

Flores, C., Ma, S.F., Pino-Yanes, M., Wade, M.S., Perez-Mendez, L., Kittles, R.A., Wang, D., Papaiahgari, S., Ford, J.G., & Kumar, R. 2012. African ancestry is associated with asthma risk in African Americans. *PLoS One*, 7, e26807

Frazer, K.A., Murray, S.S., Schork, N.J., & Topol, E.J. 2009. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*, 10, 241-251

Freedman, M.L., Haiman, C.A., Patterson, N., McDonald, G.J., Tandon, A., Waliszewska, A., Penney, K., Steen, R.G., Ardlie, K., & John, E.M. 2006. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proceedings of the National Academy of Sciences*, 103, 14068-14073

Friedrich, D.C., Santos, S.E.B., Ribeiro-dos-Santos, A.K.C., & Hutz, M.H. 2012. Several different lactase persistence associated alleles and high diversity of the lactase gene in the admixed Brazilian population. *PLoS One*, 7, e46520

Fu, X.N. 2008. Interracial marriage and family socio-economic well-being: Equal status exchange or caste status exchange? *Social Science Journal*, 45, 132-155

Galenson, D.W. 1984. The Rise and Fall of Indentured Servitude in the Americas - An Economic-Analysis. *Journal of Economic History*, 44, 1-26

Genovese, E.D. 1989. *The Political Economy of Slavery: Studies in the Economy and Society of the Slave South,* Wesleyan University Press

Gething, P.W., Patil, A.P., Smith, D.L., Guerra, C.A., Elyazar, I.R.F., Johnston, G.L., Tatem, A.J., & Hay, S.I. 2011. A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malaria journal*, 10, 378

Giacomoni, P.U., Mammone, T., & Teri, M. 2009. Gender-linked differences in human skin. *Journal of dermatological science*, 55, 144-149

Gibson, G., 2012. Rare and common variants: twenty arguments. *Nature Reviews Genetics*, 13, 135-145

Ginalski, K., Rychlewski, L., Baker, D. and Grishin, N.V. 2004. Protein structure prediction for the male-specific region of the human Y chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 2305-2310

Gómez-Carballa, A., Ignacio-Veiga, A., Álvarez-Iglesias, V., Pastoriza-Mourelle, A., Ruíz, Y., Pineda, L., Carracedo, Á. and Salas, A. 2012. A melting pot of multicontinental mtDNA lineages in admixed Venezuelans. *American journal of physical anthropology*, 147, 78-87

Gonzalez, A.M., Cabrera, V.M., Larruga, J.M., Tounkara, A., Noumsi, G., Thomas, B.N., & Moulds, J.M. 2006. Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other western Africa populations. *Annals of Human Genetics*, 70, 631-657

Grandjean, A.C. 1997. Diets of elite athletes: Has the discipline of sports nutrition made an impact? *Journal of Nutrition*, 127, S874-S877

Gray, M.W. 1989. Origin and evolution of mitochondrial DNA. *Annual review of cell biology*, 5, 25-50

Guan, Y.T. 2014. Detecting Structure of Haplotypes and Local Ancestry. *Genetics*, 196, 625-642

Guo, G., Fu, Y., Lee, H., Cai, T., Mullan Harris, K., & Li, Y. 2013. Genetic Bio-Ancestry and Social Construction of Racial Classification in Social Surveys in the Contemporary United States. *Demography*, 51, 141-172

Guo, Y., He, J., Zhao, S.L., Wu, H., Zhong, X., Sheng, Q.H., Samuels, D.C., Shyr, Y., & Long, J.R. 2014. Illumina human exome genotyping array clustering and quality control. *Nature Protocols*, 9, 2643-2662

Haines, R., McDonald, J., & Shlomowitz, R. 2001. Mortality and voyage length in the middle passage revisited. *Explorations in Economic History*, 38, 503-533

Hall, D. 1971. *Five of the Leewards, 1834-1870 the major problems of the post-emancipation period in Antigua, Barbuda, Montserrat, Nevis and St. Kitts* St. Laurence, Barbados, Caribbean Universities Press

Hammer, M.F. & Zegura, S.L. 1996. The role of the Y chromosome in human evolutionary studies. *Evolutionary Anthropology: Issues, News, and Reviews*, 5, 116-134

Han, Y.Y., Pei, Y.F., Liu, Y.Z., Zhang, L., Wu, S.Y., Tian, Q., Chen, X.D., Shen, H., Zhu, X.Z., Papasian, C.J., & Deng, H.W. 2012. Bivariate genome-wide association study suggests fatty acid desaturase genes and cadherin DCHS2 for variation of both compressive strength index and appendicular lean mass in males. *Bone*, 51, 1000-1007

Hanchard, N.A., Hambleton, I., Harding, R.M., & McKenzie, C.A. 2006. Predicted declines in sickle allele frequency in Jamaica using empirical data. *American Journal of Hematology*, 81, 817-823

Hancock, A.M., Witonsky, D.B., Alkorta-Aranburu, G., Beall, C.M., Gebremedhin, A., Sukernik, R., Utermann, G., Pritchard, J.K., Coop, G., & di Rienzo, A. 2011. Adaptations to climate-mediated selective pressures in humans. *PLoS Genetics*, 7, e1001375

Hanis, C.L., Chakraborty, R., Ferrell, R.E., & Schull, W.J. 1986. Individual Admixture Estimates - Disease Associations and Individual Risk of Diabetes and Gallbladder-Disease Among Mexican-Americans in Starr County, Texas. *American Journal of Physical Anthropology*, 70, 433-441

Hankinson, J.L., Kinsley, K.B., & Wagner, G.R. 1996. Comparison of spirometric reference values for Caucasian and African American blue-collar workers. *Journal of occupational and environmental medicine*, 38, 137-143

Harich, N., Costa, M.D., Fernandes, V., Kandil, M., Pereira, J.B., Silva, N.M., & Pereira, L. 2010. The trans-Saharan slave trade-clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC Evolutionary Biology*, 10, 1

Harpending, H.C. 1994. Signature of Ancient Population-Growth in A Low-Resolution Mitochondrial-DNA Mismatch Distribution. *Human Biology*, 66, 591-600

Harris, K.M., Haas, T.S., Eichner, E.R., & Maron, B.J. 2012. Sickle cell trait associated with sudden death in competitive athletes. *The American journal of cardiology*, 110, 1185-1188

Hassell, K.L. 2010. Population estimates of sickle cell disease in the US. *American Journal of Preventative Medicine*, 38, S512-S521

Hautier, C.A., Wouassi, D., Arsac, L.M., Bitanga, E., Thiriet, P., & Lacour, J.R. 1994. Relationships between postcompetition blood lactate concentration and average running velocity over 100-m and 200-m races. *European Journal of Applied Physiology and Occupational Physiology*, 68, 508-513

Heard, E. & Martienssen, R.A. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell*, 157, 95-109

Helgason, A., Nicholson, G., Stefansson, K., & Donnelly, P. 2003. A reassessment of genetic diversity in Icelanders: Strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Annals of Human Genetics*, 67, 281-297

Herbst, K.L. & Bhasin, S. 2004. Testosterone action on skeletal muscle. *Current Opinion in Clinical Nutrition & Metabolic Care*, 7, 271-277

Higman, B.W. 1976. *Slave population and economy in Jamaica, 1807-1834* Cambridge, UK, Cambridge University Press

Hilling, D. 1969. The Evolution of the Major Ports of West Africa. *The Geographical Journal*, 135, 365-378

Hinch, A.G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C.D., Chen, G.K., Wang, K., Buxbaum, S.G., & Akylbekova, E.L. 2011. The landscape of recombination in African Americans. *Nature*, 476, 170-175

Hoggart, C.J., Shriver, M.D., Kittles, R.A., Clayton, D.G., & McKeigue, P.M. 2004. Design and analysis of admixture mapping studies. *The American Journal of Human Genetics*, 74, 965-978

Hollings, S.C., Mallett, C.J., & Hume, P.A. 2014. The transition from elite junior track-and-field athlete to successful senior athlete: why some do, why others don't. *International Journal of Sports Science & Coaching*, 9, 457-471

Hosick, M.B. 13-04-2010. Protocol decided for sickle cell testing. *The NCAA News*

Howie, B.N., Donnelly, P., & Marchini, J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5, e1000529

Hughey, M.W. 2014. Survival of the Fastest? *Contexts*, 13, 56-58

Illumina Inc. Illumina GenCall data analysis software. 2005. Accessed on: 24-7-2016.

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

Ingram, C.J.E., Raga, T.O., Tarekegn, A., Browning, S.L., Elamin, M.F., Bekele, E., Thomas, M.G., Weale, M.E., Bradman, N., & Swallow, D.M. 2009. Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. *Journal of molecular evolution*, 69, 579-588

International Association of Athletics Federations. All time top peformance lists. 2016a.

International Association of Athletics Federations. IAAF World Championships Archive of Past Events. 2016b. Accessed on: 27-8-2016.

International Olympic Committee. Olympic Athletics - Rio 2016 Summer Olympics. 2016. Accessed on: 14-9-2016.

Jackson, B.A., Wilson, J.L., Kirbah, S., Sidney, S.S., Rosenberger, J., Bassie, L., Alie, J.A.D., McLean, D.C., Garvey, W.T., & Ely, B. 2005. Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *American Journal of Physical Anthropology*, 128, 156-163

Jackson, F.L. 1991. An evolutionary perspective on salt, hypertension, and human genetic variability. *Hypertension*, 17, I129

Jackson, F.L.C. 2008a. Ancestral links of Chesapeake Bay region African Americans to specific Bight of Bonny (West Africa) microethnic groups and increased frequency of aggressive breast cancer in both regions. *American Journal of Human Biology*, 20, 165-173

Jackson, F.L.C. 2008b. Ethnogenetic layering (EL): an alternative to the traditional race model in human variation and health disparity studies. *Annals of Human Biology*, 35, 121-144

Jamaica Olympic Association. Jamaica Olympic Association. 20-8-2016. Accessed on: 27-8-2016

James, G. D. & Baker, P. T. 1995, "Human population biology and blood pressure: evolutionary and ecological considerations and interpretations of population studies," 2nd ed. J. H. Laragh & B. M. Brenner, eds., New York, New York: Raven Press, pp. 115-126

Jin, W.F., Li, R., Zhou, Y., & Xu, S.H. 2014. Distribution of ancestral chromosomal segments in admixed genomes and its implications for inferring population history and admixture mapping. *European Journal of Human Genetics*, 22, 930-937

Jin, W.F., Xu, S.H., Wang, H.F., Yu, Y.G., Shen, Y.P., Wu, B.L., & Jin, L. 2012. Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Research*, 22, 519-527

Jobling, M.A. & Tyler-Smith, C. 2003. The human Y chromosome: An evolutionary marker comes of age. *Nature Reviews Genetics*, 4, 598-612

Johnson, D.C., Shrestha, S., Wiener, H.W., Makowsky, R., Kurundkar, A., Wilson, C.M., & Aissani, B. 2013. Mitochondrial DNA diversity in the African American population. *Mitochondrial DNA,* 26, 445-451

Johnson, D.M. & Campbell, R.R. 1981. *Black migration in America: a social demographic history* Duke University Press

Kahn, L. M. 2000, "A level playing field? Sports and discrimination," *In The Economics of Sports*, Kern & W.S., eds., Kalamazoo, MI: WE Upjohn Institute, pp. 115-130

Kane, M. 1971. An assessment of black is best. *Sports Illustrated*, 34, 72-83

Kao, W.H., Klag, M.J., Meoni, L.A., Reich, D., Berthier-Schaad, Y., Li, M., Coresh, J., Patterson, N., Tandon, A., Powe, N.R., Fink, N.E., Sadler, J.H., Weir, M.R., Abboud, H.E., Adler, S.G., Divers, J., Iyengar, S.K., Freedman, B.I., Kimmel, P.L., Knowler, W.C., Kohn, O.F., Kramp, K., Leehey, D.J., Nicholas, S.B., Pahl, M.V., Schelling, J.R., Sedor, J.R., Thornley-Brown, D., Winkler, C.A., Smith, M.W., & Parekh, R.S. 2008. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nature Genetics*, 40, 1185-1192

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., & Hammer, M.F. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research*, 18, 830-838

Kark, J.A., Labotka, R.J., Gardner, J.W., & Ward, F.T. 2010. Prevention of exercise-related death unexplained by preexisting disease (EDU) associated with sickle cell trait (SCT) without hemoglobin (Hb) screening or Hb specific management. *Blood*, 116, 945

Karlsson, E.K., Kwiatkowski, D.P., & Sabeti, P.C. 2014. Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15, 379-393

Keer, J.T. & Birch, L. 2008. *Essentials of nucleic acid analysis: A robust approach* Royal Society of Chemistry

Keinan, A., Mullikin, J.C., Patterson, N., & Reich, D. 2009. Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nature Genetics*, 41, 66-70

Kenney, M.C., Chwa, M., Atilano, S.R., Falatoonzadeh, P., Ramirez, C., Malik, D., Tarek, M., del Carpio, J.C., Nesburn, A.B., Boyer, D.S. & Kuppermann, B.D. 2014. Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: Implications for population susceptibility to diseases. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1842, 208-219

Key, N.S. & Derebail, V.K. 2010. Sickle-cell trait: novel clinical significance. *ASH Education Program Book*, 2010, 418-422

Khanna, N. 2010. "If you're half black, you're just black": Reflected appraisals and the persistence of the one-drop rule. *The Sociological Quarterly*, 51, 96-121

Kidd, L.R., Coulibaly, A., Templeton, T.M., Chen, W., Long, L.O., Mason, T., Bonilla, C., Akereyeni, F., Freeman, V., Isaacs, W., Ahaghotu, C., & Kittles, R.A. 2006. Germline BCL-2 sequence variants and inherited predisposition to prostate cancer. *Prostate Cancer and Prostatic Diseases*, 9, 284-292

Kiely, J. & Collins, D.J. 2016. Uniqueness of human running coordination: The integration of modern and ancient evolutionary innovations. *Frontiers in Psychology*, 7, 262

Kiple, K.F. & Higgins, B.T. 1989. Mortality caused by dehydration during the middle passage. *Social Science History*, 13, 420-437

Kiple, K.F. & King, V.H. 2003. *Another dimension to the black diaspora: diet, disease and racism* Cambridge University Press

Kirkegaard, E.O.W., Wang, M., & Fuerst, J. 2016. Biogeographic ancestry and socioeconomic outcomes in the Americas: a Meta-analysis. *bioRxiv,* 055681

Kittles, R.A., Young, D., Weinrich, S., Hudson, J., Argyropoulos, G., Ukoli, F., Adams-Campbell, L., & Dunston, G.M. 2001. Extent of linkage disequilibrium between the androgen receptor gene CAG and GGC repeats in human populations: implications for prostate cancer risk. *Human Genetics*, 109, 253-261

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. *American Journal of Human Genetics*, 75, 752-770

Klein, H. 1999. *The Atlantic slave trade*. Cambridge, UK, Cambridge University Press

Knight, A., Underhill, P.A., Mortensen, H.M., Zhivotovsky, L.A., Lin, A.A., Henn, B.M., Louis, D., Ruhlen, M., & Mountain, J.L. 2003. African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Current Biology*, 13, 464-473

Kofler, B., Mueller, E.E., Eder, W., Stanger, O., Maier, R., Weger, M., Haas, A., Winker, R., Schmut, O., Paulweber, B., & Iglseder, B. 2009. Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study. *BMC medical genetics*, 10, 35

Krings, M., Salem, A.H., Bauer, K., Geisert, H., Malek, A.K., Chaix, L., Simon, C., Welsby, D., di Rienzo, A., Utermann, G., Sajantila, A., Paabo, S., & Stoneking, M. 1999. mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration. *American Journal of Human Genetics*, 64, 1166-1176

Kulikoff, A. 1986. *Tobacco and slaves: The development of southern cultures in the Chesapeake, 1680-1800,* UNC Press Books

Kumar, R., Tsai, H.J., Hong, X.M., Liu, X., Wang, G.Y., Pearson, C., Ortiz, K., Fu, M., Pongracic, J.A., Bauchner, H., & Wang, X.B. 2011. Race, Ancestry, and Development of Food-Allergen Sensitization in Early Childhood. *Pediatrics*, 128, E821-E829

Kuroda-Kawaguchi, T.K., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S., & Page, D.C. 2001. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genetics*, 29, 279-286

Kuroki, Y., Iwamoto, T., Lee, J.W., 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. *Journal of Human Genetics*, 44, 289-292

Lao, O., Vallone, P.M., Coble, M.D., Diegoli, T.M., van Oven, M., van der Gaag, K.J., Pijpe, J., de Knijff, P., & Kayser, M. 2010. Evaluating Self-declared Ancestry of US Americans With Autosomal, Y-chromosomal and Mitochondrial DNA. *Human Mutation*, 31, E1875-E1893

Law, R. 1991. *The slave coast of West Africa 1550-1750: the impact of the Atlantic slave trade on an African society,* Oxford University Press, USA

Le Gallais, D., Prefaur, C., Dulat, C., Macabies, J., & Lonsdorfer, J. 1991. Sickle cell trait in Ivory Coast athletic champions, 1956-1989. *International Journal of Sports Medicine*, 12, 509-510

Lee, C.H., Olson, P., & Evans, R.M. 2003. Minireview: lipid metabolism, metabolic diseases, and peroxisome proliferator-activated receptors. *Endocrinology*, 144, 2201-2207

Lee, I.M., Shiroma, E.J., Lobelo, F., Puska, P., Blair, S.N., & Katzmarzyk, P.T. 2012a. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*, 380, 219-229

Lee, Y.C., Kuo, H.C., Chang, J.S., Chang, L.Y., Huang, L.M., Chen, M.R., Liang, C.D., Chi, H., Huang, F.Y., & Lee, M.L. 2012b. Two new susceptibility loci for Kawasaki disease identified through genome-wide association analysis. *Nature Genetics*, 44, 522-525

Lewis, M.P. 2009. *Ethnologue: Languages of the World*, 16 ed. Dallas, Texas, SIL International

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L. & Myers, R.M. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319, 1100-1104

Lind, J.M., Hutcheson-Dilks, H.B., Williams, S.M., Moore, J.H., Essex, M., Ruiz-Pesini, E., Wallace, D.C., Tishkoff, S.A., O'Brien, S.J., & Smith, M.W. 2007. Elevated male European and female African contributions to the genomes of African American individuals. *Human Genetics*, 120, 713-722

Lind, P.M., Riserus, U., Salihovic, S., van Bavel, B., & Lind, L. 2013. An environmental wide association study (EWAS) approach to the metabolic syndrome. *Environment international*, 55, 1-8

Lombardo, M.P. & Deaner, R.O. 2014. You can't teach speed: sprinters falsify the deliberate practice model of expertise. *PeerJ*, 2, e445

Long, J.C. 1991. The Genetic-Structure of Admixed Populations. *Genetics*, 127, 417-428

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., & Foster, B. 2013. The genotype-tissue expression (GTEx) project. *Nature Genetics*, 45, 580-585

Lopez, C., Saravia, C., Gomez, A., Hoebeke, J., & Patarroyo, M.A. 2010. Mechanisms of genetically-based resistance to malaria. *Gene*, 467, 1-12

Lovejoy, P.E. 1983. *Transformations in slavery: A history of slavery in Africa*, Cambridge University Press

Lovejoy, P.E. 1989. The impact of the Atlantic slave trade on Africa: a review of the literature. *The Journal of African History*, 30, 365-394

Lovejoy, P.E. 1994. Background to rebellion: The origins of Muslim slaves in Bahia. *Slavery & Abolition*, 15, 151-180

Lynch, M., Ackerman, M.S., Gout, J.F., Long, H., Sung, W., Thomas, W.K. & Foster, P.L. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, 17, 704-714

Ma, F., Yang, Y., Li, X., Zhou, F., Gao, C., Li, M. & Gao, L. 2013. The association of sport performance with ACE and ACTN3 genetic polymorphisms: a systematic review and meta-analysis. *PLoS One*, 8, e54685

MacArthur, D.G., Seto, J.T., Raftery, J.M., Quinlan, K.G., Huttley, G.A., Hook, J.W., Lemckert, F.A., Kee, A.J., Edwards, M.R., Berman, Y. & Hardeman, E.C. 2007. Loss of ACTN3 gene function alters mouse muscle metabolism and shows evidence of positive selection in humans. *Nature Genetics*, 39, 1261-1265

Macnamara, B.N., Moreau, D., & Hambrick, D.Z. 2016. The relationship between deliberate practice and performance in sports: A meta-analysis. *Perspectives on Psychological Science*, 11, 333-350

Maji, S., Krithika, S., & Vasulu, T.S. 2009. Phylogeographic distribution of mitochondrial DNA macrohaplogroup M in India. *Journal of Genetics*, 88, 127-139

Majumdar, A.S. & Robergs, R.A. 2011a. The Science of Speed: Determinants of Performance in the 100m Sprint. *International Journal of Sports Science & Coaching*, 6, 479-493

Majumdar, A.S. & Robergs, R.A. 2011b. The Science of Speed: Determinants of Performance in the 100m Sprint A Response to Commentary. *International Journal of Sports Science & Coaching*, 6, 499-500

Manning, P. 1983. Contours of Slavery and Social-Change in Africa. *American Historical Review*, 88, 835-857

Marcuello, A., Martínez-Redondo, D., Dahmani, Y., Casajús, J.A., Ruiz-Pesini, E., Montoya, J., López-Pérez, M.J. & Díez-Sánchez, C. 2009. Human mitochondrial variants influence on oxygen consumption. *Mitochondrion*, 9, 27-30

Margulis, L., 1981. *Symbiosis in cell evolution: Life and its environment on the early earth*. San Francisco: Freeman

Marlin, L., Etienne-Julan, M., Le Gallais, D., & Hue, O. 2005. Sickle cell trait in French West Indian elite sprint athletes. *International Journal of Sports Medicine*, 26, 622-625

Marshall, D. 1982. Migration As An Agent of Change in Caribbean Island Ecosystems. *International Social Science Journal*, 34, 451-467

Martinez, A.F., Muenke, M., & Arcos-Burgos, M. 2011. From the black widow spider to human behavior: Latrophilins, a relatively unknown class of G protein–coupled receptors, are implicated in psychiatric disorders. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 156, 1-10

Mateu, E., Comas, D., Calafell, F., Perez-Lezaun, A., Abade, A., & Bertranpetit, J. 1997. A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea. *Annals of Human Genetics*, 61, 507-518

Mathias, R.A., Grant, A.V., Rafaels, N., Hand, T., Gao, L., Vergara, C., Tsai, Y.J., Yang, M., Campbell, M., & Foster, C. 2010. A genome-wide association study on African-ancestry populations for asthma. *Journal of Allergy and Clinical Immunology*, 125, 336-346

Mathias, R.A., Taub, M.A., Gignoux, C.R., Fu, W., Musharoff, S., O'Connor, T.D., Vergara, C., Torgerson, D.G., Pino-Yanes, M., Shringarpure, S.S. & Huang, L. 2016. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nature Communications, 7, 12522

Maughan, R.J., Gleeson, M. & Greenhaff, P.L. 1997. *Biochemistry of exercise and training*. Oxford University Press, USA

McGowan, W. 1990. African resistance to the Atlantic slave trade in West Africa. *Slavery & Abolition*, 11, 5-29

McKeigue, P.M. 1998. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *The American Journal of Human Genetics*, 63, 241-251

McKeigue, P.M., Carpenter, J.R., Parra, E.J., & Shriver, M.D. 2000. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Annals of Human Genetics*, 64, 171-186

McLean, D.C., Spruill, I., Argyropoulos, G., Page, G.P., Shriver, M.D., & Garvey, W.T. 2005. Mitochondrial DNA (mtDNA) haplotypes reveal maternal population genetic affinities of Sea Island Gullah-speaking African Americans. *American Journal of Physical Anthropology*, 127, 427-438

McLean, D.C., Spruill, I., Gevao, S., Morrison, E.Y., Bernard, O.S., Argyropoulos, G., & Garvey, W.T. 2003. Three novel mtDNA restriction site polymorphisms allow exploration of population affinities of African Americans. *Human Biology*, 75, 147-161

Medland, S.E., Loesch, D.Z., Mdzewski, B., Zhu, G., Montgomery, G.W. & Martin, N.G. 2007. Linkage analysis of a model quantitative trait in humans: Finger ridge count shows significant multivariate linkage to 5q14.1. *PLoS Genetics*, 3, e165

Meeusen, R., Watson, P., Hasegawa, H., Roelands, B., & Piacentini, M.F. 2006. Central fatigue. *Sports Medicine*, 36, 881-909

Mehta, P.H. & Beer, J. 2010. Neural mechanisms of the testosterone–aggression relation: The role of orbitofrontal cortex. *Journal of Cognitive Neuroscience*, 22, 2357-2368

Mendizabal, I., Sandoval, K., Berniell-Lee, G., Calafell, F., Salas, A., Martínez-Fuentes, A. and Comas, D. 2008. Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC evolutionary biology*, 8, 213

Mero, A., Komi, P.V., & Gregor, R.J. 1992. Biomechanics of sprint running. *Sports Medicine*, 13, 376-392

Mikami, E., Fuku, N., Murakami, H., Tsuchie, H., Takahashi, H., Ohiwa, N., Tanaka, H., Pitsiladis, Y.P., Higuchi, M., & Miyachi, M. 2014. ACTN3 R577X genotype is associated with sprinting in elite Japanese athletes. *International Journal of Sports Medicine*, 35, 172-177

Mikami, E., Fuku, N., Takahashi, H., Ohiwa, N., Scott, R.A., Pitsiladis, Y.P., Higuchi, M., Kawahara, T., & Tanaka, M. 2010. Mitochondrial haplogroups associated with elite Japanese athlete status. *British Journal of Sports Medicine*, 45, 1179-1183

Mintz, S.W. & Price, R. 1976. *An anthropological approach to the Afro-American past: a Caribbean perspective,* Philadelphia, PA, Institute for the Study of Human Issues

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. *Proceedings of the National Academy of Sciences*, 100, 171-176

Mishra, P. & Chan, D.C. 2014. Mitochondrial dynamics and inheritance during cell division, development and disease. *Nature Reviews Molecular Cell Biolgy*, 15, 634-646

Monson, K. L., Miller, K. W. P., Wilson, M. R., Dizinno, J. A., & Budowle, B. The mtDNA population database: an integrated software and database resource for forensic comparison. *Forensic Science Communications* 2. 2002. Accessed on: 01-10-2009.

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

Moore, J.W. 2003. The Modern World-System as environmental history? Ecology and the rise of capitalism. *Theory and Society*, 32, 307-377

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. *Human Genetics*, 115, 492-497

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J.L., Byrnes, J.K., Gignoux, C.R., Ortiz-Tello, P.A., Martínez, R.J., Hedges, D.J., Morris, R.W. & Eng, C. 2013. Reconstructing the population genetic history of the Caribbean. *PLoS Genetics*, 9, e1003925

Morera, B. 2002. Análisis del Polimorfismo del ADNmt en la Población General de Costa Rica: un Asunto Pendiente. *Revista Latinoamericana de Derecho Médico y Medicina Legal*, 7, 21-34

Morgan, K. 2006. Slave women and reproduction in Jamaica, c.1776-1834. *History*, 91, 231-253

Morgan, K. 2007. *Slavery and the British Empire: from Africa to America*, First ed. New York, United States, Oxford University Press

Morrison, E.Y. & Cooper, P.D. 2006. Some bio-medical mechanisms in athletic prowess. *West Indian Medical Journal*, 55, 205-209

Mozos, I., 2015. Mechanisms linking red blood cell disorders and cardiovascular diseases. *BioMed research international*, 2015, 682054

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. *Japanese Journal of Physiology*, 52, 247-256

National Center for Health Statistics 2013, *Special feature on emergency care*, National Center for Health Statistics (US).

Nelson, D.A., Deuster, P.A., Carter, R., Hill, O.T., Wolcott, V.L., & Kurina, L.M. 2016. Sickle Cell Trait, Rhabdomyolysis, and Mortality among US Army Soldiers. *New England Journal of Medicine*, 375, 435-442

Ni, X., Yang, X., Guo, W., Yuan, K., Zhou, Y., Ma, Z., & Xu, S. 2016. Length Distribution of Ancestral Tracks under a General Admixture Model and Its Applications in Population History Inference. *Scientific reports*, 6

Niemi, A.K. & Majamaa, K. 2005. Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *European Journal of Human Genetics*, 13, 965-969

Nogales-Dagea, G., Pinos, T., Ruiz, J.R., Marzo, P.F., Fiuza-Luces, C., Lopez-Gallardo, E., Ruiz-Pesini, E., Martin, M.A., Arenas, J., Moran, M., Andreu, A.L., & Lucia, A. 2011. Are mitochondrial haplogroups associated with elite athletic status? A study on a Spanish cohort. *Mitochondrion*, 11, 905-908

Novembre, J. & Ramachandran, S. 2011. Perspectives on Human Population Structure at the Cusp of the Sequencing Era. *Annual Review of Genomics and Human Genetics,* 12, 245-274

Nuñez, C., Baeta, M., Sosa, C., Casalod, Y., Ge, J., Budowle, B. and Martínez-Jarreta, B. 2010. Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers. *American journal of physical anthropology*, 143, 591-600

Nunn, N. 2008. The long-term effects of Africa's slave trades. *Quarterly Journal of Economics*, 123, 139-176

O'Malley, G.E. 2009. Beyond the Middle Passage: Slave migration from the Caribbean to North America, 1619-1807. *The William and Mary Quarterly*, 66, 125-172

O'Sullivan, M.L., de, W.J., Savas, J.N., Comoletti, D., Otto-Hitt, S., Yates, J.R., III, & Ghosh, A. 2012. FLRT proteins are endogenous latrophilin ligands and regulate excitatory synapse development. *Neuron*, 73, 903-910

Ong, B.A., Li, J., McDonough, J.M., Wei, Z., Kim, C., Chiavacci, R., Mentch, F., Caboot, J.B., Spergel, J., Allen, J.L., Sleiman, P.M.A., & Hakonarson, H. 2013. Gene Network Analysis in a Pediatric Cohort Identifies Novel Lung Function Genes. *PLoS One*, 8, e72899

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

Orsini, C.A., Setlow, B., DeJesus, M., Galaviz, S., Loesch, K., Ioerger, T., & Wallis, D. 2016. Behavioral and transcriptomic profiling of mice null for *LPHN3*, a gene implicated in ADHD and addiction. *Molecular Genetics & Genomic Medicine*, 4, 322-343

Papadimitriou, I.D., Papadopoulos, C., Kouvatsi, A., & Triantaphyllidis, C. 2008. The *ACTN3* gene in elite Greek track and field athletes. *International Journal of Sports Medicine*, 29, 352-355

Parra, E.J., Kittles, R.A., Argyropoulos, G., Pfaff, C.L., Hiester, K., Bonilla, C., Sylvester, N., Parrish-Gause, D., Garvey, W.T., Jin, L., McKeigue, P.M., Kamboh, M.I., Ferrell, R.E., Pollitzer, W.S., & Shriver, M.D. 2001. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *American Journal of Physical Anthropology*, 114, 18-29

Parra, E.J., Marcini, A., Akey, J., Martinson, J., Batzer, M.A., Cooper, R., Forrester, T., Allison, D.B., Deka, R., Ferrell, R.E., & Shriver, M.D. 1998. Estimating African American admixture proportions by use of population-specific alleles. *American Journal of Human Genetics*, 63, 1839-1851

Pasaniuc, B., Zaitlen, N., Lettre, G., Chen, G.K., Tandon, A., Kao, W.H.L., Ruczinski, I., Fornage, M., Siscovick, D.S., & Zhu, X. 2011. Enhanced statistical tests for GWAS in admixed populations: assessment using African Americans from CARe and a Breast Cancer Consortium. *PLoS Genetics*, 7, e1001371

Patel, C.J., Bhattacharya, J., & Butte, A.J. 2010. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One*, 5, e10746

Patterson, H.O.L. 1966. Slavery, acculturation, and social change: The Jamaican case. *British Journal of Sociology*, 17, 151-164

Pego, C.M., Hill, R.F., Solomon, G.W., Chisholm, R.M., & Ivey, S.E. 1995. Tobacco, culture, and health among American Indians: A historical review. *American Indian Culture and Research Journal*, 19, 143-164

Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J., & Amorim, A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Annals of Human Genetics*, 65, 439-458

Periyasamy, K.P., Lakshmi, P.T.V., Dwibedi, C.K. & Annamalai, A. 2009. A database for human Y chromosome protein data. *Bioinformation*, 4, 184-186

Piganeau, G. & Eyre-Walker, A. 2004. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity*, 92, 282-288

Pilegaard, H., Domino, K., Noland, T., Juel, C., Hellsten, Y., Halestrap, A.P. & Bangsbo, J. 1999. Effect of high-intensity exercise training on lactate/H+ transport capacity in human skeletal muscle. *American Journal of Physiology-Endocrinology And Metabolism*, 276, E255-E261

Pinto, F., Gonzalez, A.M., Hernandez, M., Larruga, J.M., & Cabrera, V.M. 1996. Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Annals of Human Genetics*, 60, 321-330

Pitsiladis, Y., Davis, A., & Johnson, D. 2011. The Science of Speed: Determinants of Performance in the 100m Sprint A Commentary. *International Journal of Sports Science & Coaching*, 6, 495-498

Pitsiladis, Y.P., Tanaka, M., Eynon, N., Bouchard, C., North, K.N., Williams, A.G., Collins, M., Moran, C.N., Britton, S.L., Fuku, N., Ashley, E.A. 2016. Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance. *Physiological genomics*, 48, 183-190

Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, A., & Comas, D. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Human Genetics*, 115, 439-447

Plummer, M., Best, N., Cowles, K., & Vines, K. 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7-11

Posnansky, M. 1973. Aspects of early West African trade. *World Archaeology*, 5, 149-162

Poston, W.S.C., Pavlik, V.N., Hyman, D.J., Ogbonnaya, K., Hanis, C.L., Haddock, C.K., Hyder, M.L., & Foreyt, J.P. 2001. Genetic bottlenecks, perceived racism, and hypertension risk among African Americans and first-generation African immigrants. *Journal of human hypertension*, 15, 341-351

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., & Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38, 904-909

Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D., & Myers, S. 2009. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genetics*, 5, e1000519

Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D.L., Rotter, J.I., Torres, E., Taylor, K.D., Goldstein, D.B., & Reich, D. 2008. Long-range LD can confound genome scans in admixed populations. *American Journal of Human Genetics*, 83, 132-135

Pritchard, J.K., Stephens, M., & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., & Sham, P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559-575

Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K. and Santachiara-Benerecetti, A.S., 1999. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nature Genetics*, 23, 437-441

Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., Sica, L., Mouguiama-Daouda, P., Comas, D., Tzur, S., Balanovsky, O., Kidd, K.K., Kidd, J.R., van der Veen, L., Hombert, J.M., Gessain, A., Verdu, P., Froment, A., Bahuchet, S., Heyer, E., Dausset, J., Salas, A., & Behar, D.M. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proceedings of the National Academy of Sciences*, 105, 1596-1601

R Core Development Team 2008. *R: A language and environment for statistical computing (Vienna: R Foundation for Statistical Computing)*

Raj, T., Kuchroo, M., Replogle, J.M., Raychaudhuri, S., Stranger, B.E., & De Jager, P.L. 2013. Common Risk Alleles for Inflammatory Diseases Are Targets of Recent Positive Selection. *American Journal of Human Genetics*, 92, 517-529

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. *Annals of Human Genetics*, 62, 531-550

Rank, M.R. 2009. Measuring the Economic Racial Divide Across the Course of American Lives. *Race and Social Problems*, 1, 57-66

Rankinen, T., Pérusse, L., Borecki, I., Chagnon, Y.C., Gagnon, J., Leon, A.S., Skinner, J.S., Wilmore, J.H., Rao, D.C. & Bouchard, C. 2000. The Na+-K+-ATPase α2 gene and trainability of cardiorespiratory endurance: the HERITAGE Family Study. *Journal of Applied Physiology*, 88, 346-351

Rankinen, T., Roth, S.M., Bray, M.S., Loos, R., Perusse, L., Wolfarth, B., Hagberg, J.M., & Bouchard, C. 2010. Advances in Exercise, Fitness, and Performance Genomics. *Medicine and Science in Sports and Exercise*, 42, 835-846

Rappold, G.A. 1993. The Pseudoautosomal Regions of the Human Sex-Chromosomes. *Human Genetics*, 92, 315-324

Rastogi, S. & United States Bureau of the Census. 2011. *The black population: 2010* US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

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

Rebbeck, T.R., Devesa, S.S., Chang, B.L., Bunker, C.H., Cheng, I., Cooney, K., Eeles, R., Fernandez, P., Giri, V.N., & Gueye, S.M. 2013. Global patterns of prostate cancer incidence, aggressiveness, and mortality in men of African descent. *Prostate cancer*, 2013, 560857

Reed, F.A. & Tishkoff, S.A. 2006. African human diversity, origins and migrations. *Current Opinion in Genetics & Development*, 16, 597-605

Reiner, A.P., Ziv, E., Lind, D.L., Nievergelt, C.M., Schork, N.J., Cummings, S.R., Phong, A., Burchard, E.G., Harris, T.B., Psaty, B.M. & Kwok, P.Y. 2005. Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *The American Journal of Human Genetics*, 76, 463-477

Revolution Analytics & Weston, S. 2014. doParallel: Foreach parallel adaptor for the parallel package. *R package version* 1.0.8

Richardson, B. C. 1989, "Caribbean Migrations, 1838-1985," *In The Modern Caribbean*, F. W. Knight & C. A. Palmer, eds., Chapel Hill, North Carolina, USA: UNC Press Books, pp. 203-228.

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. *Medicine and Science in Sports and Exercise*, 30, 687-690

Robergs, R.A., Ghiasvand, F., & Parker, D. 2004. Biochemistry of exercise-induced metabolic acidosis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 287, R502-R516

Roberts, D.F. & Hiorns, R.W. 1965. Methods of Analysis of Genetic Composition of A Hybrid Population. *Human Biology*, 38-43

Robinson, P. 2008. *Jamaican Athletics: A Model for the World*. BlackAmber

Rogers, A.S., Futterman, D.K., Moscicki, A.B., Wilson, C.M., Ellenberg, J., & Vermund, S.H. 1998. The REACH project of the adolescent medicine HIV/AIDS research network: Design, methods, and selected characteristics of participants. *Journal of Adolescent Health*, 22, 300-311

Romero, R., Kuivaniemi, H., Tromp, G., & Olson, J.M. 2002. The design, execution, and interpretation of genetic association studies to decipher complex diseases. *American Journal of Obstetrics and Gynecology*, 187, 1299-1312

Rosa, A., Brehm, A., Kivisild, T., Metspalu, E., & Villems, R. 2004. MtDNA profile of West Africa Guineans: Towards a better understanding of the Senegambia region. *Annals of Human Genetics*, 68, 340-352

Ross, A., Leveritt, M., & Riek, S. 2001. Neural influences on sprint running - Training adaptations and acute responses. *Sports Medicine*, 31, 409-425

Rowell, D.P. 2001. Teleconnections between the tropical Pacific and the Sahel. *Quarterly Journal of the Royal Meteorological Society*, 127, 1683-1706

Roy, J.L.P., Hunter, G.R., Fernandez, J.R., McCarthy, J.P., Larson-Meyer, D.E., Blaudeau, T.E., & Newcomer, B.R. 2006. Cardiovascular factors explain genetic background differences in VO(2)max. *American Journal of Human Biology*, 18, 454-460

Salas, A., Bandelt, H.J., Macaulay, V., & Richards, M.B. 2007. Phylogeographic investigations: The role of trees in forensic genetics. *Forensic Science International*, 168, 1-13

Salas, A., Carracedo, A., Macaulay, V., Richards, M., & Bandelt, H.J. 2005a. A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. *Biochemical and Biophysical Research Communications*, 335, 891-899

Salas, A., Carracedo, A., Richards, M., & Macaulay, V. 2005b. Charting the ancestry of African Americans. *American Journal of Human Genetics*, 77, 676-680

Salas, A., Lovo-Gomez, J., Alvarez-Iglesias, V., Cerezo, M., Lareu, M.V., Macaulay, V., Richards, M.B., & Carracedo, A. 2009. Mitochondrial Echoes of First Settlement and Genetic Continuity in El Salvador. *PLoS One*, 4, e6882

Salas, A., Richards, M., de la Fe, T., Lareu, M.V., Sobrino, B., Sanchez-Diz, P., Macaulay, V., & Carracedo, A. 2002. The making of the African mtDNA landscape. *American Journal of Human Genetics*, 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. *American Journal of Human Genetics*, 74, 454-465

Samuels, D.C., Carothers, A.D., Horton, R., & Chinnery, P.F. 2006. The power to detect disease associations with mitochondrial DNA haplogroups. *The American Journal of Human Genetics*, 78, 713-720

Sandoval, K., Buentello-Malo, L., Penaloza-Espinosa, R., Avelino, H., Salas, A., Calafell, F. and Comas, D. 2009. Linguistic and maternal genetic diversity are not correlated in Native Mexicans. *Human genetics*, 126, 521

Satoh, M. & Kuroiwa, T. 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Experimental Cell Research*, 196, 137-140

Satten, G.A., Flanders, W.D., & Yang, Q.H. 2001. Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model. *American Journal of Human Genetics*, 68, 466-477

Scheet, P. & Stephens, M. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, 78, 629-644

Schiaffino, S. & Reggiani, C. 2011. Fiber types in mammalian skeletal muscles. *Physiological reviews*, 91, 1447-1531

Schiller, J.S., Lucas, J.W., Ward, B.W., & Peregoy, J.A. 2012. Summary health statistics for US adults: National Health Interview Survey, 2010. *Vital and Health Statistics.Series 10, Data from The National Health Survey*, 1-207

Schneider, S. & Excoffier, L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: application to human mitochondrial DNA. *Genetics*, 152, 1079-1089

Schwartz, M. & Vissing, J. 2002. Paternal inheritance of mtDNA in a patient with mitochondrial myopathy. *European Journal of Human Genetics*, 10, 239

Scott, R.A., Fuku, N., Onywera, V.O., Boit, M., Wilson, R.H., Tanaka, M., Goodwin, H., & Pitsiladis, Y.P. 2009. Mitochondrial haplogroups associated with elite Kenyan athlete status. *Medicine and Science in Sports and Exercise*, 41, 123-128

Scott, R.A., Irving, R., Irwin, L., Morrison, E., Charlton, V., Austin, K., Tladi, D., Deason, M., Headley, S.A., Kolkhorst, F.W., Yang, N., North, K., & Pitsiladis, Y.P. 2010. ACTN3 and ACE Genotypes in Elite Jamaican and US Sprinters. *Medicine and Science in Sports and Exercise*, 42, 107-112

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. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 140, 497-503

Sears, E.S. 2001. *Running Through the Ages*, McFarland

Seldin, M.F., Pasaniuc, B., & Price, A.L. 2011. New approaches to disease mapping in admixed populations. *Nature Reviews Genetics*, 12, 523-528

Serjeant, G.R. 2013. The natural history of sickle cell disease. *Cold Spring Harbor perspectives in medicine*, 3, a011783

Serpa, J.A., Teeter, L.D., Musser, J.M., & Graviss, E.A. 2009. Tuberculosis Disparity between US-born Blacks and Whites, Houston, Texas, USA. *Emerging Infectious Diseases*, 15, 899-904

Sewell, W. G., 1968. The ordeal of free labor in the British West Indies, *New York Times*, 2nd ed., new impression.

Sheridan, R.B. 1961. The West India Sugar Crisis and British Slave Emancipation, 1830-1833. *The Journal of Economic History*, 21, 539-551

Sheridan, R.B. 1985. *Doctors and slaves: A medical and demographic history of slavery in the British West Indies, 1680-1834*. Cambridge, UK, Cambridge University Press

Shih, M., Ambady, N., Richeson, J.A., Fujita, K., & Gray, H.M. 2002. Stereotype performance boosts: the impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology*, 83, 638

Shin, W.S., Tanaka, M., Suzuki, J., Hemmi, C., & Toyo-oka, T. 2000. A novel homoplasmic mutation in mtDNA with a single evolutionary origin as a risk factor for cardiomyopathy. *American Journal of Human Genetics*, 67, 1617-1620

Shriner, D. 2013. Overview of admixture mapping. *Current Protocols in Human Genetics,* 1-23

Shriner, D., Adeyemo, A., & Rotimi, C.N. 2011. Joint ancestry and association testing in admixed individuals. *PLoS Computional Biology*, 7, e1002325

Simms, T.M., Rodriguez, C.E., Rodriguez, R., & Herrera, R.J. 2010. The Genetic Structure of Populations from Haiti and Jamaica Reflect Divergent Demographic Histories. *American Journal of Physical Anthropology*, 142, 49-66

Simms, T.M., Wright, M.R., Hernandez, M., Perez, O.A., Ramirez, E.C., Martinez, E., & Herrera, R.J. 2012. Y-chromosomal diversity in Haiti and Jamaica: Contrasting levels of sex-biased gene flow. *American Journal of Physical Anthropology*, 148, 618-631

Simms, T.M., Wright, M.R., Martinez, E., Regueiro, M., McCartney, Q., & Herrera, R.J. 2013. Y-STR diversity and sex-biased gene flow among Caribbean populations. *Gene*, 516, 82-92

Sinha, M., Larkin, E.K., Elston, R.C., & Redline, S. 2006. Self-reported race and genetic admixture. *New England Journal of Medicine*, 354, 421-422

Slatkin, M. 2000. Allele age and a test for selection on rare alleles. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 355, 1663-1668

Sluiter, E. 1997. New light on the" 20. and Odd Negroes" arriving in Virginia, August 1619. *The William and Mary Quarterly*, 54, 395-398

Smith, A.A., Toone, R., Peacock, O., Drawer, S., Stokes, K.A., & Cook, C.J. 2013. Dihydrotestosterone is elevated following sprint exercise in healthy young men. *Journal of Applied Physiology*, 114, 1435-1440

Smith, M.W., Patterson, N., Lautenberger, J.A., Truelove, A.L., McDonald, G.J., Waliszewska, A., Kessing, B.D., Malasky, M.J., Scafe, C., Le, E., de Jager, P.L., Mignault, A.A., Yi, Z., de The, G., Essex, M., Sankale, J.L., Moore, J.H., Poku, K., Phair, J.P., Goedert, J.J., Vlahov, D., Williams, S.M., Tishkoff, S.A., Winkler, C.A., de la Vega, F.M., Woodage, T., Sninsky, J.J., Hafler, D.A., Altshuler, D.M., Gilbert, D.A., O'Brien, S.J., & Reich, D. 2004. A high-density admixture map for disease gene discovery in African Americans. *American Journal of Human Genetics*, 74, 1001-1013

Smith, S.M. 2015. Seasoning and Abolition: Humoural medicine in the eighteenth-century British Atlantic. *Slavery & Abolition*, 36, 684-703

Sohi, A.S. & Yusuff, K.B. 1987. The socioeconomic status of elite nigerian athletes in perspective of social stratification and mobility. *International Review for the Sociology of Sport*, 22, 295-303

Spiriev, B. IAAF scoring tables of athletics. International Association of Athlteics Federations . 2011. Monaco

Spuhler, J.N. 1988. Evolution of Mitochondrial-Dna in Monkeys, Apes, and Humans. *Yearbook of Physical Anthropology*, 31, 15-48

Steckel, R.H. & Jensen, R.A. 1986. New evidence on the causes of slave and crew mortality in the Atlantic slave trade. *The Journal of Economic History*, 46, 57-77

Stefflova, K., Dulik, M.C., Barnholtz-Sloan, J.S., Pai, A.A., Walker, A.H., & Rebbeck, T.R. 2011. Dissecting the Within-Africa Ancestry of Populations of African Descent in the Americas. *PLoS One*, 6, e14495

Stone, J., Lynch, C.I., Sjomeling, M., & Darley, J.M. 1999. Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, 77, 1213-1227

Tadman, M. 2000. The demographic cost of sugar: debates on slave societies and natural increase in the Americas. *The American Historical Review*, 105, 1534-1575

Tajima, F. 1983. Evolutionary Relationship of DNA-Sequences in Finite Populations. *Genetics*, 105, 437-460

Tajima, F. 1989. Statistical-Method for Testing the Neutral Mutation Hypothesis by Dna Polymorphism. *Genetics*, 123, 585-595

Tajima, F. 1993, "Measurement of DNA polymorphism," *In Mechanisms of Molecular Evolution. Introduction of Molecular Paleopopulation Biology*, N. Takahata & A. G. Clark, eds., Tokyo, Sunderland: MA: Japan Scientific Societies Press, Sinauer Associates, Inc., 37-59

Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E.G., & isch, N.J. 2007. Recent genetic selection in the ancestral admixture of Puerto Ricans. *The American Journal of Human Genetics*, 81, 626-633

Tang, H., Peng, J., Wang, P., & Risch, N.J. 2005. Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology*, 28, 289-301

Tanner, H.H. & Reiff, J.L. 1995. *The settling of North America: the atlas of the great migrations into North America from the Ice Age to present*. Macmillan.

Therrell, B.L. & Hannon, W.H. 2006. National evaluation of US newborn screening system components. *Mental retardation and developmental disabilities research reviews*, 12, 236-245

Thiessen, D. & Gregg, B. 1980. Human assortative mating and genetic equilibrium: An evolutionary perspective. *Ethology and Sociobiology*, 1, 111-140

Thomas, M.G., Barnes, I., Weale, M.E., Jones, A.L., Forster, P., Bradman, N., & Pramstaller, P.P. 2008. New genetic evidence supports isolation and drift in the Ladin communities of the South Tyrolean Alps but not an ancient origin in the Middle East. *European Journal of Human Genetics*, 16, 124-134

Thomas, M.G., Weale, M.E., Jones, A.L., Richards, M., Smith, A., Redhead, N., Torroni, A., Scozzari, R., Gratrix, F., Tarekegn, A., Wilson, J.F., Capelli, C., Bradman, N., & Goldstein, D.B. 2002. Founding mothers of Jewish communities: Geographically separated Jewish groups were independently founded by very few female ancestors. *American Journal of Human Genetics*, 70, 1411-1420

Thompson, A.A. 2013. Sickle cell trait testing and athletic participation: a solution in search of a problem? *ASH Education Program Book*, 2013, 632-637

Thornton, J. 1992. *Africa and Africans in the Making of the Atlantic World, 1400-1680.* Cambridge, UK, Cambridge University Press

Tian, C., Hinds, D.A., Shigeta, R., Kittles, R., Ballinger, D.G., & Seldin, M.F. 2006. A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. *American Journal of Human Genetics*, 79, 640-649

Tilford, C.A., Kuroda-Kawaguchi, T., Skaletsky, H., Rozen, S., Brown, L.G., Rosenberg, M., McPherson, J.D., Wylie, K., Sekhon, M., Kucaba, T.A., Waterston, R.H., & Page, D.C. 2001. A physical map of the human Y chromosome. *Nature*, 409, 943-945

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., Ibrahim, M., Juma, A.T., Kotze, M.J., Lema, G., Moore, J.H., Mortensen, H., Nyambo, T.B., Omar, S.A., Powell, K., Pretorius, G.S., Smith, M.W., Thera, M.A., Wambebe, C., Weber, J.L., & Williams, S.M. 2009. The Genetic Structure and History of Africans and African Americans. *Science*, 324, 1035-1044

Tofanelli, S., Bertoncini, S., Castri, L., Luiselli, D., Calafell, F., Donati, G., & Paoli, G. 2009. On the Origins and Admixture of Malagasy: New Evidence from High-Resolution Analyses of Paternal and Maternal Lineages. *Molecular Biology and Evolution*, 26, 2109-2124

Torres, J.B., Doura, M.B., Keita, S.O.Y., & Kittles, R.A. 2012. Y Chromosome Lineages in Men of West African Descent. *PLoS One*, 7, e29687

Torres, J.B., Kittles, R.A., & Stone, A.C. 2007. Mitochondrial and Y chromosome diversity in the English-speaking Caribbean. *Annals of Human Genetics*, 71, 782-790

Torres, J.B., Stone, A.C., & Kittles, R. 2013. An Anthropological Genetic Perspective on Creolization in the Anglophone Caribbean. *American Journal of Physical Anthropology*, 151, 135-143

Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M., & Wallace, D.C. 1993. Asian Affinities and Continental Radiation of the 4 Founding Native-American Mtdnas. *American Journal of Human Genetics*, 53, 563-590

Touhami, I., Fattoum, S., Bibi, A., Siala, H., Messaoud, T., Koubaa, D., Mankai, R., Bartagi, Z., & Le Gallais, D. 2010. The epidemiology of abnormal hemoglobins in Mediterranean high-level athletes. *European Journal of Applied Physiology*, 108, 1075-1081

Tucker, R. & Collins, M. 2012. What makes champions? A review of the relative contribution of genes and training to sporting success. *British Journal of Sports Medicine*, 46, 555-561

Tucker, R., Santos-Concejero, J., & Collins, M. 2013. The genetic basis for elite running performance. *British Journal of Sports Medicine*, 47, 545-549

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Lahr, M.M., Foley, R.A., Oefner, P.J., & Cavalli-Sforza, L.L. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of Human Genetics*, 65, 43-62

United States Bureau of the Census 1975. *Historical statistics of the United States, colonial times to 1970* US Department of Commerce, Bureau of the Census.

USA Track and Field. Team USA at the IAAF World Championships. http://www.usatf.org/statistics/TeamUSA/IAAFWorldChampionships.aspx. 2016a. Accessed on: 01-09-2016

USA Track and Field. Team USA at the Olympic Games - Medal Winners. http://www.usatf.org/statistics/TeamUSA/OlympicGames_kentico/medalsWinners.aspx. 2016b. Accessed on: 01-09-2016

Vaeyens, R., Gullich, A., Warr, C.R., & Philippaerts, R. 2009. Talent identification and promotion programmes of Olympic athletes. *Journal of Sports Sciences*, 27, 1367-1380

Van Oven, M. & Kayser, M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30, E386-E394

Van Praagh, E. & Dore, E. 2002. Short-term muscle power during growth and maturation. *Sports Medicine*, 32, 701-728

Van Someren, K.A. 2006. The physiology of anaerobic endurance training. *The Physiology of Training. UK: Churchill Livingstone*, 85-115

Vaughan, A.T. 1972. Blacks in Virginia: a note on the first decade. *The William and Mary Quarterly*, 29, 469-478

Veeramah, K.R., Connell, B.A., Pour, N.A., Powell, A., Plaster, C.A., Zeitlyn, D., Mendell, N.R., Weale, M.E., Bradman, N., & Thomas, M.G. 2010. Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evolutionary Biology*, 10, 1

Vergara, C., Caraballo, L., Mercado, D., Jimenez, S., Rojas, W., Rafaels, N., Hand, T., Campbell, M., Tsai, Y.J., Gao, L., Duque, C., Lopez, S., Bedoya, G., Ruiz-Linares, A., & Barnes, K.C. 2009. African ancestry is associated with risk of asthma and high total serum IgE in a population from the Caribbean Coast of Colombia. *Human Genetics*, 125, 565-579

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., & Wilson, A.C. 1991. African Populations and the Evolution of Human Mitochondrial-Dna. *Science*, 253, 1503-1507

Vincent, L., Feasson, L., Oyono-Enguelle, S., Banimbek, V., Monchanin, G., Dohbobga, M., Wouassi, D., Martin, C., Gozal, D., & Geyssant, A. 2010. Skeletal muscle structural and energetic characteristics in subjects with sickle cell trait, α-thalassemia, or dual hemoglobinopathy. *Journal of Applied Physiology*, 109, 728-734

Visscher, P.M., Yang, J.A., & Goddard, M.E. 2010. A Commentary on 'Common SNPs Explain a Large Proportion of the Heritability for Human Height' by Yang et al. (2010). *Twin Research and Human Genetics*, 13, 517-524

Wallace, D.C., Brown, M.D., & Lott, M.T. 1999. Mitochondrial DNA variation in human evolution and disease. *Gene*, 238, 211-230

Wallace, D.C. 2010. Mitochondrial DNA mutations in disease and aging. *Environmental and molecular mutagenesis*, 51, 440-450

Wallace, D.C. 2015. Mitochondrial DNA variation in human radiation and disease. *Cell*, 163, 33-38

Wallis, D., Hill, D.S., Mendez, I.A., Abbott, L.C., Finnell, R.H., Wellman, P.J., & Setlow, B. 2012. Initial characterization of mice null for Lphn3, a gene implicated in ADHD and addiction. *Brain Research*, 1463, 85-92

Walsh, L.S. 1989. Plantation management in the Chesapeake, 1620-1820. *The Journal of Economic History*, 49, 393-406

Wang, G. 2013. *Genetic studies of elite athlete status.* Ph.D. thesis, University of Glasgow

Wang, G., Mikami, E., Chiu, L.L., de Perini, A., Deason, M.L., Fuku, N., Miyachi, M., Kaneoka, K., Murakami, H., Tanaka, M., Hsieh, L.L., Hsieh, S.S., Caporossi, D., Pigozzi, F., Hilley, A., Lee, R., Galloway, S.D.R., Gulbin, J., Rogozkin, V.A., Ahmetov, I.I., Yang, N., North, K.N., Ploutarhos, S., Montgomery, H.E., Bailey, M.E.S., & Pitsiladis, Y.P. 2013. Association analysis of ACE and ACTN3 in elite caucasian and East Asian swimmers. *Medicine and Science in Sports and Exercise*, 45, 892-900

Wang, S., Ray, N., Rojas, W., Parra, M.V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., Hurtado, A.M. and Camrena, B. 2008. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genetics*, 4, e1000037

Watson, E., Forster, P., Richards, M., & Bandelt, H.J. 1997. Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics*, 61, 691-704

Weder, A.B. & Schork, N.J. 1994. Adaptation, allometry, and hypertension. *Hypertension*, 24, 145-156

Weiss, K.M. & Long, J.C. 2009. Non-Darwinian estimation: My ancestors, my genes' ancestors. *Genome Research*, 19, 703-710

White, D.J., Bryant, D., & Gemmell, N.J. 2013. How Good Are Indirect Tests at Detecting Recombination in Human mtDNA? *G3-Genes Genomes Genetics*, 3, 1095-1104

Whitehead, N. 1999, "Native society and the European occupation of the Caribbean islands and coastal *Tierra Firme*, 1492-1650," *In General History of the Caribbean*, P. C. Emmer, ed., London: UNESCO Publishing, pp. 180-200

Wickham, H. 2009. *ggplot2: elegant graphics for data analysis*. Springer Publishing Company, Incorporated

Wilks, R., Younger, N., Tulloch-Reid, M., McFarlane, S., & Francis, D. 2008. Jamaica health and lifestyle survey 2007-8. *Kingston: Tropical Medicine Research Institute, University of the West Indies, Mona*

Williams, E. 1994. *Capitalism and slavery*. UNC Press Books

Williamson, J. 1995. *New People: Miscegenation and Mulattoes in the United States* Louisiana State University Press

Wilson, T.W. & Grim, C.E. 1991. Biohistory of slavery and blood pressure differences in blacks today. A hypothesis. *Hypertension*, 17, I122

Winkler, C.A., Nelson, G.W., & Smith, M.W. 2010. Admixture Mapping Comes of Age. *Annual Review of Genomics and Human Genetics, Vol 11*, 11, 65-89

Wood, P.H. 1996. *Black majority: Negroes in colonial South Carolina from 1670 through the Stono Rebellion*. WW Norton & Company

Worldbank 2011. *Migration and remittances factbook*, 2nd ed ed. Washington D.C, World Bank

Wray, N. and Visscher, P. 2008. Estimating trait heritability. *Nature Education*, 1, 29

Xie, P., Kranzler, H.R., Yang, C., Zhao, H., Farrer, L.A., & Gelernter, J. 2013. Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. *Biological Psychiatry*, 74, 656-663

Yang, J.A., Benyamin, B., Mcevoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., & Visscher, P.M. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42, 565-U131

Yang, J.J., Burchard, E.G., Choudhry, S., Johnson, C.C., Ownby, D.R., Favro, D., Chen, J., Akana, M., Ha, C., & Kwok, P.Y. 2008. Differences in allergic sensitization by self-reported race and genetic ancestry. *Journal of Allergy and Clinical Immunology*, 122, 820-827

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. *American Journal of Human Genetics*, 73, 627-631

Yao, Y.G., Salas, A., Logan, I., & Bandelt, H.J. 2009. mtDNA Data Mining in GenBank Needs Surveying. *American Journal of Human Genetics*, 85, 929-933

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T.L. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13,

Yehuda, R., Daskalakis, N.P., Bierer, L.M., Bader, H.N., Klengel, T., Holsboer, F., & Binder, E.B. 2015. Holocaust exposure induced intergenerational effects on FKBP5 methylation. *Biological Psychiatry*, 80, 372-380

Zakharia, F., Basu, A., Absher, D., Assimes, T.L., Go, A.S., Hlatky, M.A., Iribarren, C., Knowles, J.W., Li, J., Narasimhan, B., Sidney, S., Southwick, A., Myers, R.M., Quertermous, T., Risch, N., & Tang, H. 2009. Characterizing the admixed African ancestry of African Americans. *Genome Biology*, 10, R141.1-R141.11

Zhu, X., Luke, A., Cooper, R.S., Quertermous, T., Hanis, C., Mosley, T., Gu, C.C., Tang, H., Rao, D.C., Risch, N., & Weder, A. 2005. Admixture mapping for hypertension loci with genome-scan markers. *Nature Genetics*, 37, 177-181