Recent Human History: Inferences from the Y-Chromosome and Mitochondrial DNA

A thesis submitted for the Degree of Doctor of Philosophy

Julia Kirsi Abernethy

University College, London. September 2004.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Disciplines such as palaeoanthropology, archaeology, anthropology, and history have been instrumental in formulating hypotheses relating to human history. Genetics has developed into a powerful tool for human population analysis hence it can complement information derived from other disciplines. To date, however, such studies of genetic history have predominantly focussed on prehistoric events.

The aim of this thesis was to address several questions formulated from written sources and oral tradition relating to the *recent* history of populations in the British Isles and Africa. Y-chromosome markers and sequence information from the mitochondrial genome were employed. The male gene pool of the British Isles was investigated using a thorough sampling strategy, with respect to the impact of historical invaders, revealing geographic structuring within the Isles as a result of differential contact with these invaders. With these data for Britain available, the fidelity of (British) surname inheritance was investigated using the Y-chromosome, revealing evidence for the random and non-random adoption of surnames. The scope in Britain was narrowed to the small, but assumed diverse, metropolitan district of Greater London, to assess levels of Y-chromosome and mitochondrial DNA diversity in relation to the rest of Britain and Europe. Finally, the maternal history of the Lemba from Africa was investigated; oral tradition and Y-chromosome evidence suggests a Semitic component. The evidence presented here precludes a Jewish maternal heritage, but a Middle Eastern component is possible.

This thesis has shown that genetic information can be informative for elucidating the recent history of these populations, therefore confirming the value of including recent events within the scope of genetic history.

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

Appendices

List of Figures

Chapter 2

Chapter 4

Chapter 5

Appendices

Publications Arising From This Thesis

Chapter 2:

Capelli, C., Redhead, N., Abernethy, J. K., Gratrix, F., Wilson J. F., Moen, T., Hervig, T., Richards, M., Stumpf, M. P. H., Underhill, P. A., Bradshaw, P., Shaha, A., Thomas M. G., Bradman, N., Goldstein, D.B. (2003). A Y-Chromosome Census of the British Isles. *Curr. Biol.* **13:** 979-984.

Chapter 3:

Abernethy, J. K., Capelli, C., and Goldstein, D.B. (in preparation). What's In a Name: How do Surnames and Y-Chromosomes Correlate?

Declaration

The work presented in this thesis is all my own, with the following exceptions: Chapter 2: Section 2.2.1 sample design of all non-Channel Islands populations (C Capelli, UCL); Section 2.2.2 and Section 2.2.3 (most non-Channel Islands DNAs were extracted by C Capelli and N Redhead); DNA extraction and genotyping of most non-Channel Island populations (C Capelli and N Redhead); Chapter 4: initial mtDNA HVS1 PCR and sequencing reactions were performed by M-W Burley; Chapter 5/Appendix, Table A.6 mtDNA RFLP assays (A Torroni, Università di Pavia/Università "La Sapienza"). PCR products in Chapter 3 were kindly electrophoresed by A Smith and M-W Burley and those in Chapter 4 in M-W Burley.

Julia Abernethy

Acknowledgements

There are many people to thank who have all helped me over the past four years, both at UCL and my friends and family outside UCL. First of all I have to thank my principle supervisor David Goldstein. I would also like to thank my second supervisor Mark Thomas who has always provided useful discussions about my work.

I am indebted to all of the volunteers from around the world who have kindly contributed the DNA samples used in this thesis. These samples have been invaluable in enabling the work presented here to take place. Several people and organisations have also aided with sample collection, or provided samples, for each of the Chapters: the BBC, F. Falle and W. Galliene (Chapter 2); K. Barnfather, L. Cognetta, A. Causton, J. Causton, S. Farrer, P. Folland, A. MacLeod, T. Sorbie, P.A.A. Speechley, R. Thwaite, and M. Whittock (Chapter 3); The Museum of London, Amy Non and Richard Byrne (Chapter 4); and Mark Thomas and the Centre for Genetic Anthropology (Chapter 5).

I am grateful to Cristian Capelli, Neil Bradman, Martin Richards, and Jim Wilson for providing either unpublished data or data prior to publication. Mari-Wyn Burley, Fiona Gratrix, and Alice Smith ran the sequencing facility used in Chapters 3 and 4. Michael Stumpf provided valuable discussions for Chapter 3, and finally Antonio Torroni and his lab assayed several mtDNA RFLPs for Chapter 5.

Friends and past and present members of the Goldstein and Thomas labs have been great friends, and have helped make my time at UCL as much fun as it could be! They are: Alice, Cecilia, Cristian, Helen, Mari-Wyn, Richard M., Bekah, Holly, Kate, Turi, and the entire CRASH Team. My family, particularly my dad Pete, have all been very understanding when my thesis has had to take priority (i.e. for most of the last four years!). And, Rich: it's difficult to sum up everything in a few words of acknowledgements……You were always there for me and without your love and support I would not have got through the last 12

months, let alone written my thesis with some sanity left. Now we're both done it's time to start enjoying life again!

Chapter 1. Introduction

1.1. Genetics and Human History

Genetics is increasingly becoming an informative tool to elucidate our understanding of the evolution and history of our species. Many of the questions that geneticists seek to answer have arisen and been formulated by colleagues in other disciplines such as palaeontology, archaeology, anthropology, history, and linguistics. Gaps in the fossil and archaeological records (e.g. Lahr and Foley 1998; Klein 1999; McBrearty and Brooks 2000; Gamble *et al.* 2004), uncertainty over the dating and provenance of archaeological remains (e.g. Klein 1999), as well as the lack of written records for most of our species' history, mean that inferences from these sources can be beset by problems. For example, since the discovery of the first recognised "non-modern human" fossil (a Neanderthal from the Neander Valley in Germany in 1856) the collection of fossil hominids has rapidly increased, covering around 5 million years of hominid evolution. Yet, this represents only a handful of fossil remains that are used to track very important events in our history, such as the migration out of Africa to Australia by modern humans (*Homo sapiens*), or the colonisation of Eurasia by *Homo sapiens* (*H. sapiens*), the route and timing of both is still disputed by palaeoanthropologists (Stringer 2000). Other processes such as the Neolithic revolution, which saw the spread of agriculture westwards from the Middle East to Europe and possibly North Africa (Arredi *et al.* 2004), are well characterised by archaeology, but such evidence cannot provide explicit information about the extent to which the Neolithic revolution was simply the movement of culture and ideas (i.e. a cultural diffusion), or if it was accompanied by the spread of people (i.e. a demic diffusion), a topic which is hotly debated (e.g. Cavalli-Sforza *et al.* 1994).

DNA samples, by comparison, can be collected from populations living today and used to infer the genetic history of past populations. The access to such DNA resources is thus "only" limited by the available time and finds of a laboratory and the willingness of DNA donors. Although these are not trivial matters they are issues that can be addressed. In contrast, the scarcity of fossil and archaeological remains is a function of taphonomic and deposition

processes, something which one is powerless over, in much the same way that unavailability of written records covering a particular event or period of interest are outside one's control. The fact that it is possible, and advantageous, to use genetics to study human history is not merely theoretical hyperbole; there are many questions of anthropological interest that have been addressed using genetics. For example the debate surrounding the Out of Africa versus the Multiregional hypothesis of modern human evolution was originally an intractable palaeoanthropological argument, but genetic data has been applied to the question over the last 17 years, starting with the now infamous "mitochondrial eve" paper (Cann *et al.* 1987) and now strongly supports the Out of Africa "side" (Hammer 1995; Underhill *et al.* 2000; Underhill *et al.* 2001; Penny *et al.* 1995; Chen *et al.* 1995; Harpending *et al.* 1998; Ingman *et al.* 2000; Goldstein *et al.* 1995a; Antunez-de-Mayolo *et al.* 2002). Today the Out of Africa theory is now the most widely accepted model of modern human origins (Lahr and Foley 1998; Stringer 2003; but see also Wolpoff 2000).

Genetically derived information cannot be regarded as infallible however. It is not always possible to provide conclusive evidence for or against competing hypotheses, particularly as the present day distribution of genetic variation has been shaped by various processes (Goldstein and Chikhi 2002). These may or may not be possible to differentiate using genetics or may be unknown hence cannot be controlled for. For example, findings based on the Y-chromosome and mitochondrial DNA (mtDNA) have supported opposite sides of the Neolithic revolution argument (Renfrew 2000). This highlights the need for genetic data to be used in conjunction with information from other disciplines (Cavalli-Sforza *et al.* 1994; Di Benedetto *et al.* 2001; Goldstein and Chikhi 2002). As Hurles and Jobling (2001) noted, incongruence between data from difference sources need not be interpreted as a problem, as it may be indicative of intriguing results.

Much of the focus of genetic history has been events that happened many thousands of years ago, such as the early expansions of *H. sapiens* out of Africa prior to 50,000 years ago, which is the oldest date of *H. sapiens* outside Africa (Bowler *et al.* 2003), and subsequent colonisation of the rest of the world

(Stringer 2003). For the field of genetic history, even events such as the supposed Neolithic expansion into Europe around 10,000 years ago (10 kya) from the Middle East (Cavalli-Sforza *et al.* 1994) are relatively recent in comparison, despite being far beyond the range of historical records or oral history. (The term "historical records" will be used here to refer explicitly to written records, other forms of records such as palaeontology and archaeology will not be placed within this group and will be named separately). Questions relating to more recent events that are within the scope of historical records and oral tradition are less frequently tackled. There are some exceptions to this pattern. For example Wilson *et al.* (2001a) addressed (in part) the genetic impact of male Norwegian Vikings and possible Anglo-Saxon influence on Orkney and some other populations in Britain, and Carvajal-Carmona *et al.* (2000) who assessed the influence of Spanish conquistadors on the male and female lineages in Antioquia, Colombia. Both of these events occurred within the last ~1,300 years, and have been documented in written records, oral tradition and folklore.

These two examples also highlight the different types of recent history that can be studied, one of which is intuitively "easier" to investigate. The homelands of Vikings and Anglo-Saxons and the British Isles are geographically very close and the European population as a whole has a relatively recent common origin. In contrast, the Spanish conquistadors and the indigenous population of Colombia are geographically and temporally distinct. It is therefore intuitive that recent events that deal with disparate populations will be easier to study because one expects such populations to be more (genetically) different from each other than those who are separated by smaller geographic distances and who are expected to have more recent common ancestry (Cavalli-Sforza *et al.* 1994). Studies of the sort that address recent history are relatively small in number, particularly those that deal with populations that are temporally and geographically closely related. It is this latter point which motivated the work in this thesis.

It is now useful to briefly highlight some of the ways in which studying ancient and recent historical events differ. Here the adjective "ancient" is used to describe events that happened in the pre-historic period, i.e. before written

records, whilst "recent" describes the historic period, i.e. the period when written records exist. Also note that the time at which historical records are found for the first time in different regions is not the same; the earliest written texts have been found in Mesopotamia in the Middle East, dated to around 5kya, and the earliest Germanic texts to 25 AD (Senner 1991), whilst the oldest texts of languages indigenous to Britain are those written in ogham dated to the 4th century AD (Lehmann 1991), although of course Latin texts from the Roman period can be found (Hall and Conheeney 1998). First, in the context of studying ancient events, it is known that the palaeontological and archaeological records are incomplete (Lahr and Foley 1998; Klein 1999; McBrearty and Brooks 2000; Gamble *et al.* 2004), and as one moves further into the past so these deficiencies increase, yet until the development of writing and written history these are the only sources of evidence that exist from which one can infer the pre-historic past of our species. This has several important implications for the way in which one is able to interpret ancient events. Timescales are inevitably cruder both in terms of the accuracy with which palaeontological and archaeological remains can be dated and in the availability of such remains to date (Klein 1999).

A particularly good example comes from the evolution of modern humans; it is accepted wisdom in most circles that modern humans evolved in Africa during the Pleistocene (Stringer 2003), however until the discovery of *H. sapiens* fossils in Ethiopia dated to around 160kya (Clark *et al.* 2003) the most reliable *H. sapiens* fossils were not from Africa but from Skhul and Qafzeh in Israel, dated to around 115kya (Stringer 2003). Such a dearth of fossils is a feature of African geology in general, compounded by the large area of the continent (Stringer 2003), therefore it is probably going to be difficult to infer with any certainty the exact pattern and timing of modern human evolution.

With such small amounts of data therefore, only large-scale events are likely to be detected; that smaller-scale events took place is undeniable, they simply cannot be identified from the available evidence. Furthermore the fact that ancient events are so intangible means it is inevitable that there is a tendency to compound pre-history into large periods of time and see only the bigger picture, regardless of what evidence is available. In terms of human history however, it

is probably fair to say that most of the large-scale events, such as the migrations out of Africa, did happen in the ancient past. If one assumes that archaic *Homo* populations were completely replaced by *H. sapiens* from an original source in Africa (Stringer 2003) then the presence of humans in all inhabitable parts of the world is the result of these migrations that culminated in the colonisation of the Americas, around 20-15kya (Schurr and Sherry 2004). Notable exceptions to this pattern are the colonisation of Iceland in the 9th century AD by Vikings (Jones 1984), and the European colonialists who explored and settled in many parts of the world.

The historic period is by contrast better documented and defined, both because written records and oral history add to the range of available data sources and because the preservation of archaeological and palaeontological remains improves as one moves towards the present, therefore these sources become more useful. The historical period is short in comparison to prehistory; the first written text dates to only 5kya (Senner 1991), but modern human prehistory extends to at least 160kya (i.e. the oldest date of the most unambiguous *H. sapiens* fossil in Africa; Clark *et al.* 2003). 5kyrs is barely measurable for the pre-historic period. For example one of the confidence intervals for the dates of the Australian sites studied by Bowler *et al.* (2003) mentioned above was 4kyrs, i.e. a period of time akin to the whole of the historic period in the Middle East. Recent history can thus be described in more detail and placed into a better-defined, and narrower, timescale, leading to an increased resolution for such events. For example, British written records that cover the Viking period (primarily the *Anglo-Saxon Chronicles* and *British History,* Richards 1991) pinpoint the time and location of Vikings raids on the British Isles (Richards 1991), making it possible to not only address the impact of Vikings on Britain in general, but on very specific regions. At the same time, it has also seems that Vikings are essentially invisible in the archaeological record (Richards 1991), therefore had the Viking raids been an ancient event it would have been impossible to know that they took place, let alone address where in Britain they landed. However, the prima facie reliability of the written record should not be assumed, a point which is addressed in more detail below. The change in perspective from the relative power to detect and understand the ~160kyrs of

prehistory to the ability to understand the last 5kyrs in much better detail inevitably means that recent events can be seen as smaller events because one has a better understanding and increased refinement of the smaller processes involved. However, to reiterate the converse of the point made above, it is probably also true to say that recent history has been fewer large-scale events than prehistory, with the exceptions noted above.

Some of the questions and issues surrounding the use of genetic history to study recent events will now be discussed, before introducing the populations and questions addressed in this thesis. The following four areas will be covered: (i) why recent history should be studied using genetics; (ii) why genetic history has focussed on ancient events; (iii) is it possible to investigate recent events?; (iv) some specific considerations of studying recent events.

1.2. Why Study Recent History?

Given that the historical period of human history is by definition characterised by written records, as well as improved fossil and archaeological preservation, as discussed above, why should we apply genetic history to studying recent events, given its ample coverage from numerous sources? It is apparent that for ancient events genetics will be useful; as one moves backwards in time the quantity and quality of evidence (be it palaeontological or archaeological), diminishes therefore genetics clearly has a role. The implicit assumption must be that as one encounters events in the written record, and indeed oral histories as one moves even closer to the present, the evidence is detailed enough to not require genetic investigation. However, by using examples drawn from two of the Chapters in this thesis, it is possible to illustrate how this assumption is not always met and written and oral history can be just as incomplete as the fossil or archaeological record. For example the Anglo-Saxon period was an important part of British history, covering 3 centuries from the 5th to 8th centuries AD. The most extensive and reliable written records documenting the migration of Anglo-Saxons to Britain were written by Bede in the 8th century AD, yet Bede's

account (the *Ecclesiastical History of the English Peoples*) was not contemporaneous with the Anglo-Saxons migrations and was based on earlier written sources. Some of these earlier accounts were also written after the event (Welch 1992). For example, Bede's treatment of the early Anglo-Saxon period was primarily drawn from a sermon written by Gildas in the 6th century that was vehemently against what he perceived as the savage lives of indigenous Britons and saw the Anglo-Saxons as a just punishment from God (Welch 1992). The sermon was thus written after the actual event and by a biased author.

The reliability of such historical evidence is therefore open to question. Moreover, the question of whether the Anglo-Saxon period involved the mass-migration of Anglo-Saxons or not is much debated in archaeology (Davies 1999; Graham-Campbell and Batey 1998; Richards *et al.* 2000). This matter has been addressed by archaeologists by inference from the remains of material culture and burial sites in the archaeological record, however the issue has not been resolved. Even more recent history can be equally difficult to understand from historical sources. Surnames in the British Isles are only thought to have been used for a maximum of 1,000 years (Reaney 1997), therefore falling within the scope of historical records, and to some extent oral tradition for their very recent history. But the genealogical study of surnames suffers from scanty written records and the fact that not all events that are relevant to tracing the history of a surname (births, marriages, deaths, infidelity) were always recorded. Anecdotal evidence is also pertinent; from one's own experience, how many generations back is it possible to trace our own surname? Do we even know the maiden names of our grandmothers, or great grandmothers, let alone where they were born or married? Hence, whilst it might be assumed that recent events in our past are adequately documented and do not need further investigation, it is apparent that this is not always the case. Therefore genetic history clearly has a role to play in addressing these questions. Furthermore, some of the issues raised in recent history are also particularly suited to being tackled using genetics.

1.3. Why Has Genetic History Focussed on Ancient Events?

It was previously noted that genetic history has tended to focus on ancient events in human history, yet the range of genetic markers now available have the potential to study recent events, even between closely related populations. This section will address two important reasons why recent events have not been studied in as much detail; these are primarily related to: (i) the history of discovery of suitable genetic markers and (ii) the prevailing questions in related fields at the time that such markers were found and made readily available. First, the ability to detect genetic diversity within and between populations, which any study of genetic history seeks to understand and interpret to make inferences about past processes and events, has greatly increased in recent years, allowing greater power to detect and study recent events. However prior to such technological developments the field was quite different. The existence of genetic similarities and differences between populations has been recognized for almost 100 years since the ABO blood group system was first described by Hirszfeld in 1919 (detailed by Cavalli-Sforza and Feldman 2003). Since this early study many other protein polymorphisms, or classical markers, (~20) were identified (Strachan and Read 1999), allowing the publication of important works such as Mourant's 1954 *The Distribution of Human Blood Groups* and more recently *The History and Geography of Human Genes* (Cavalli-Sforza *et al.* 1994). Although some of the early works on classical markers simply described the distribution of different polymorphisms, later work moved towards using this information to make and test inferences about human evolution and history. The ability to differentiate between populations using these classical markers is limited. For example there are only four different classes of the ABO system: A, B, AB, and O, therefore the entire human population will belong to one of these four groups, allowing little differentiation to be ascertained. Technological advances, such as the development of the polymerase chain reaction (PCR) in 1986 (Mullis *et al.* 1986), coupled with the increased ability to assay larger numbers of samples quickly and cheaply, have allowed the study of more detailed genetic variation than earlier work on protein polymorphisms. Today, we are in the era of studying populations at the DNA level, exemplified

by the publication of the sequence of the human genome (International Human Genome Sequencing Consortium, 2001), the HapMap project (The International HapMap Consortium 2003) and the work of the SNP Consortium (The International SNP Map Working Group 2001). In contrast to protein polymorphisms, polymorphisms at the DNA level are much more diverse. For example by assessing the frequencies of single nucleotide polymorphisms (SNPs) an estimate was made that humans differ from each other at around 1 per 1,250 nucleotides (Reich *et al.* 2003), therefore offering a very high level of resolution (although it is currently unfeasible to sequence the entire nuclear genome for studying genetic history).

Markers that are of relatively low resolution, such as the classical polymorphisms, therefore have less power to differentiate populations as fewer alleles are present to assay. Many of these protein polymorphisms are subject to selection. The effect of selection means that (i) mutation rates may be constrained, and (ii) some of the observed patterns of diversity might be reflecting similarities and differences in selection pressures, rather than the history of migration, gene flow and isolation. Assaying a combination of different protein polymorphisms might negate this problem because it is unlikely that selection will affect the frequency of different polymorphisms in the same direction (but see Fix 1996). Therefore with less ability to differentiate populations it was the large-scale events in human history that tended to be studied and, as argued above. However it should be noted starting 50 years ago, and using classical markers, Cavalli-Sforza and colleagues started using Italian church records to look at the incidences of consanguinity in Italy (Cavalli-Sforza *et al.* 2004). Despite some of the limitations of classical markers however, they have been successfully used to study human evolution and pre-history (see the summary provided by Cavalli-Sforza *et al.*, 1994, for example). More recent work using DNA polymorphisms has modified some of the conclusions of these initial studies, but in many instances the overall conclusions broadly agree (Cavalli-Sforza *et al.* 1994).

The development of more polymorphic markers therefore creates the potential for recent events to be studied, whilst also allowing the ancient events to be

studied in more detail. Two loci in particular, the Y-chromosome and mtDNA, have become particularly popular in all aspects of genetic history (Hurles and Jobling 2001), due to some of their unique properties and range of polymorphic markers that can be easily assayed. These features will be considered briefly in the next section, and covered in more detail in the Introductions to Chapter 2 and Chapter 4. It should also be remembered that although the potential of the Y-chromosome and mtDNA to investigate recent events is great, this potential was not realised for some time because such mutations have to first be identified and assessed for their ability to differentiate populations. For example the first Y-chromosome polymorphism was discovered in 1985 (Casanova *et al.* 1985), yet Y-chromosome studies have only really flourished in the wake of the publication of the worldwide distribution of 166 polymorphisms (Underhill *et al.* 2000). mtDNA has a longer track record of use in genetic history, starting with the publication of Cann *et al.* in 1982 (detailed by Cavalli-Sforza *et al.* 1994), with subsequent years seeing the characterisation of regions and populations in more detail, such as Europe (Torroni *et al.* 1996) and Africa (Salas *et al.* 2002). However the more variable control region only started to be included in analyses after the publication of Graven *et al.* (1995). Furthermore the development of these fields is on-going and new research will undoubtedly reveal additional mutations that will allow the human population to be described in even more detail. Therefore in their early stages as (potentially) useful loci in the study of genetic history, the Y-chromosome and mtDNA did not offer an improvement to the better characterised classical markers.

The second and final point discussed in this section concerns the fact that genetic history does not exist in a vacuum, and in most instances questions that are addressed by genetic history are often raised in other disciplines such as palaeontology and archaeology. Therefore whilst the early classical markers had limited power to detect small scale differences, at the time that DNA-based techniques started to come into use one of the most hotly debated topics in palaeontology and archaeology was the origin of *H. sapiens* and the relative merits of the Out of Africa and Multiregional hypotheses, arguably the largest and most important event in our history. The debate was in part triggered by new fossil and archaeological finds, and the reinterpretation of previous evidence, as

well as the now famous study of mtDNA lineages by Cann *et al.* 1987 (Lahr and Foley 1998). Therefore it was logical that this question was focussed on. As DNA-based methods improved, genetic history studies have tended to return to the same questions to ascertain whether more detailed analyses still support the early conclusions (Cavalli-Sforza *et al.* 1994), as discussed above, or to addresss any flaws in the initial research.

1.4. Is It Possible to Study Recent History?

The preceding sections have shown that our understanding of recent human history could theoretically be elucidated by using genetics. Whilst classical markers have less resolution to detect the small-scale events associated with recent history, the development of highly polymorphic markers, particularly on the Y-chromosome and mtDNA should mean it is possible to apply genetic analyses to recent history. In this section the key points relating to the suitability of the Y-chromosome and mtDNA to studying recent genetic history will be briefly reviewed (fuller reviews can be found in the Introductions to Chapter 2 and Chapter 4) before proceeding to discuss some examples of how these markers have been used to study recent history of closely related populations.

There are several important characteristics of the Y-chromosome and mtDNA that have made mutations on these loci particularly suitable for studying genetic history. Both loci are thought to escape recombination due to their uniparental modes of inheritance (male and female respectively), therefore each can unambiguously trace the male and female histories of humans and represent their genealogy as one single tree, which is impossible with recombining biparental loci (e.g. Nordborg 2000; Cavalli-Sforza and Feldman 2003). This is particularly important given evidence for distinct histories of males and females (Seielstad *et al.* 1998). The range of polymorphic sites that are routinely typed on the Y-chromosome and mtDNA to define lineages are thought to be selectively neutral (but see for example Krausz *et al.* 2001 for the Y-chromosome and Mishmar *et al.* 2003 for mtDNA). Thus the observed

distribution of the lineages defined by these mutations across the world should be a function of drift, hence these lineages should reflect the history of migration, gene flow, and isolation of the populations they are found in, rather than the selective pressures experienced by the populations. Furthermore due to their uniparental inheritance the Y-chromosome and mtDNA have an effective population size that is ¼ that of autosomes (Storz *et al.* 2001) which means these loci experience more drift than the remainder of the genome.

The overall mutation rate of the Y-chromosome is low (or more precisely $\theta=N\mu$, the population mutation parameter which represents the expected level of diversity in population a in terms of the mutation rate and drift, Jobling *et al.* 2003) compared to the rest of the nuclear genome (International SNP Map Working Group, 2001). Thus the initial discovery of polymorphic markers was slow (Hammer and Zegura 2002), although many fast-mutating microsatellites are now known (Kayser *et al.* 2004). Slowly evolving polymorphisms (or Unique Event Polymorphisms, UEPs; [Jobling and Tyler-Smith 1995; Thomas *et al.* 1998]) are used to define stable groups of chromosomes that share a common ancestor (haplogroups or hgs) whilst faster mutating microsatellites are used to define haplotypes within haplogroups to analyse closer evolutionary relationships (de Knijff 2000). In contrast the mitochondrial genome as a whole has a very high mutation rate (Brown *et al.* 1979; Budwole *et al.* 2003). In an analogous strategy to SNPs on the Y-chromosome restriction fragment length polymorphism (RFLP) sites from throughout the mtDNA genome are typically used to assign mtDNA lineages to hgs (Torroni *et al.* 1996; Finnilä *et al.* 2001) because they mutate at a relatively slower rate. The hypervariable regions I and II (HVSI and HVSII), which mutate at particularly high rates (Heyer *et al.* 2001) are used to define haplotypes or subhaplogroups. Note however that the presence of diagnostic HVSI sequence motifs means that hgs can often be defined by sequence information alone (Graven *et al.* 1995).

Several key features of the Y-chromosome and mtDNA are particularly useful for studying recent events. Numerous Y-chromosome and mtDNA lineages have been identified, many of which display geographical localisation in their

distribution (see for example Underhill *et al.* 2001 for the Y-chromosome and Forster 2004 for mtDNA). mtDNA does however appear to be less structured by geography than the Y-chromosome however, particularly in Europe (Simoni *et al.* 2000). Now that the worldwide distribution of many of the common lineages is known, for studies of recent history it will be possible to identify those lineages that should be polymorphic in the region being assessed. Some regions are much better characterised than others, therefore the extent to which suitably polymorphic markers can be selected is not consistent for all regions of the world, for example Africa is particularly poorly characterised (Cruciani *et al.* 2002). It is likely that regional-based analyses, built upon the initial global patterns of diversity, will reveal rarer geographically localised lineages (Weale *et al.* 2003; Finnilä *et al.* 2001), as well as describe the geographic distribution of known lineages in better detail (see for example Rootsi *et al.* 2004; Maca-Meyer *et al.* 2003). Furthermore, the presence of fast-mutating regions on both the Y-chromosome and mtDNA allows more recent population history to be studied, given that less genetic differentiation is expected between populations that share a recent common ancestry (Jobling and Tyler-Smith 1995; de Knijff, 2000).

Indeed, the study of recent events based only on Y-linked UEP-defined hgs and mtDNA RFLP-defined hgs, without any haplotype information from either microsatellites or HVS-I sequence information, may not be informative, depending on the amount of geographic structure of the lineages in the region of interest. Moreover, because the hg-defining mutations are typically slowly evolving, it is unlikely that any new (hg-defining) mutations will have arisen in the populations being studied for recent events. As Helgason *et al.* (2000) noted in the context of Europe, recent history has seen the redistribution of hgs, rather than the appearance of new hgs, a statement that can be applied to the recent history of any region. Therefore in studying recent history, information such as the inferred age of the hg will not be informative, a case that was made regarding the analysis of Y-chromosome lineages by Helgason *et al.* (2000) in Iceland (de Knijff 2000). Therefore there is an increased reliance on faster mutating markers, such as Y-linked microsatellites and mtDNA HVSI sequence information.

Nonetheless it is still possible to show that the recent history of populations separated by small geographic distances can be analysed using the Y-chromosome and mtDNA by briefly discussing 4 publications that have dealt with this matter. Using a combination of Y-linked microsatellites and UEPs Weale *et al.* (2002) and Wilson *et al.* (2001a) were able to detect the influence of Anglo-Saxon and (Norwegian) Viking men on a total of 10 British populations using a variety of statistical methods. It was possible to detect parts of Britain that appeared to experience more or less introgression from these European populations, although the patchy coverage of the British Isles by these studies meant that it was not possible to assess the overall pattern for the British Isles, or detect finer-scale patterns. However, this nonetheless showed that there was potentially some geographic structure within the British Isles over relatively small geographic distances, something only previously observed for larger distances in Europe (Rosser *et al.* 2000; Zerjal *et al.* 2001). The work of Sykes and Irven (2000) investigated the more recent history of the Sykes surname in England using Y-linked microsatellites and found that it was possible to identify a Sykes modal haplotype that was absent from two English comparison populations and significantly associated with sampled Sykes men. Although this was an important finding confirming the prediction that the Y-chromosome could be used in surname analysis (Jobling and Tyler-Smith 1995), the significant association between the modal haplotype and Sykes is fortuitous because on the basis of probability it would have been more likely for the Sykes to have a haplotype (the Atlantic Modal Haplotype) that is commonly found in British populations (Wilson *et al.* 2001a; Weale *et al.* 2002). It would thus have been more rigorous to have also typed binary markers to confirm that individuals sharing the same haplotype also belonged to the same hg.

In contrast to the structure seen for British Y-chromosomes in the above studies, recent work on an independent French dataset (Dubut *et al.* 2004) showed that there was not any significant differentiation for mtDNA hgs in 5 regions of France based on the analysis of the frequency of mtDNA hgs. However, analysis of gene (h) and nucleotide (π_n) diversity indicated that it was possible to infer that one of the populations from Brittany had experienced admixture, probably

involving British and/or Irish populations, and AMOVA analysis indicated that the Brittany population was more similar to the British/Irish samples than those from France (Dubut *et al.* 2004). The latter finding of the limited structure (based on hg frequencies) must also be a function of the lack of geographic structure for mtDNA hgs in Europe generally (Simoni *et al.* 2000). This finding also indicates how important it is to use the appropriate genetic markers in the analyses i.e. here faster mutation HVSI sequence information was more informative than the slower mutating RFLP-defined hgs as the events being considered within a relatively recent time frame.

1.5. Issues Associated With Studying Recent Events

Finally, several issues are encountered in the study of recent events that do not pose a problem when investigating events of a deeper timescale. These issues can be considered under the broad heading of sampling strategies. Sampling strategy is an important consideration in any study of genetic history; the type of strategy chosen can greatly affect conclusions drawn from the data (see Hammer *et al.* 2003 for a timely example relating to the Y-chromosome) and is in part dictated by the questions being addressed. As any kind of genetic history relies on collecting samples from modern day populations to represent the past populations being investigated, archaeology and history can be used to infer and identify which modern day populations should be sampled from to best represent the historical populations (Cavalli-Sforza *et al.* 1994). Written records that relate to the population(s) and period of interest are an advantage of studying recent events, even if the fidelity of such records might be questioned, as discussed above, because they allow the potential of more accurate identification of populations. But there are situations where it is not possible to decide upon the most appropriate sample location from written sources or the archaeological record. When studying recent events within a small geographic region small discrepancies in the choice of sample location might be very important. For example, the input of Anglo-Saxons on the British male gene pool has been considered in recent studies (Wilson *et al.* 2001a; Weale *et al.* 2002), as well as

the study presented in Chapter 2. Wilson *et al.* (2001a) hypothesised that samples from Friesland might be a good representative of the Anglo-Saxons population, whilst Frisian samples were explicitly used to represent Anglo-Saxons by Weale *et al.* (2002). However it is not apparent that Frisians are the best choice of Anglo-Saxon source population, and some historians postulate that Schleswig Holstein is more likely to be a better representative (Welch 1992). Indeed, Schleswig Holstein was used in Chapter 2. Such a debate is only possible for recent events when the amount and specificity of evidence allows the question to be of fine resolution and the question about sampling to be about two regions, like Frisia and Schleswig Holstein, that are separated by very small geographic differences. In the absence of existing work to indicate the amount of differentiation between two geographically close populations, such as Frisia and Schleswig Holstein, it is impossible to know the extent to which inappropriate samples would confound any conclusions.

The issue of choosing the correct population from which to sample was recently highlighted by Chikhi *et al.* (2002) in the context of the more distant events of the Neolithic period in Europe, and specifically the extent of gene flow from the Middle East. The method used to analyse the dataset employed a likelihood-based approach (LEA), which explicitly relies on the identification of two parental populations to calculate the degree of input from each of these populations on the hybrid population (Chikhi *et al.* 2001). It is clear that the incorrect assignment of one or both of the parental populations will affect the way in which their influence on the hybrid population is interpreted. Indeed, Chikhi *et al.* (2002) used a sample of Basques to represent indigenous Europeans, a choice that is consistent with various sources of evidence that suggest the Basque population is the best representative of Palaeolithic Europeans (Calafell and Bertanpetit 1994; Comas *et al.* 2000; Cavalli-Sforza *et al.* 1994; Chikhi *et al.* 2002). However it has been shown that the Basques have experienced recent gene flow with their Catalan neighbours (Hurles *et al.* 1999), therefore it is not apparent that Basques are as much a Palaeolithic relict population as often assumed, although they are perhaps the best Palaeolithic representatives currently available. Such problems would be encountered when studying recent events using the same analytical methods.

Once the appropriate sample location has been selected with as much precision as is feasible it is necessary that the samples collected are as representative of the selected location as possible. Again, this is a consideration of any study of genetic history regardless of the timeframe. However the accuracy and precision with which this has to be done is quite different depending on the age of the event being studied. As argued above, events in recent history that are studied using genetics are typically on a smaller scale to those that occurred in the pre-historic period, particularly when the events being studied are within a small geographic area. The main factors to be considered are the effect of gene flow between the populations being studied, which is more likely to be an issue for geographically close populations. Within a wider context any migration to the population being sampled which may introduce genetic types not previously seen in that population needs to be considered. This is a particular problem for large metropolitan districts as these see the largest number of immigrants (Cohen 2004).

The differences between sample design can be illustrated by indicating the sample considerations that need to be borne in mind when sampling for progressively more recent events in human history. The oldest event in the history of *H. sapiens* that can be examined genetically from extant populations is the split between humans and their closest ancestor, *Pan (Chimpanzee).* To study this event samples need to be obtained from humans and chimpanzees. The split between *Pan* and *Homo* was so far in the past (recently dated to 5-6kya; Wildman *et al.* 2003) that there has been sufficient time for genetic divergence to occur between the two genera as a whole. Therefore any differentiation within and between *H. sapiens* and/or *Pan* populations should be small compared to that between *H. sapiens* and *Pan.* Thus, with regard to sampling from humans, whilst it would be prudent to include samples from each of the continents to ensure that maximum human diversity is represented, further consideration is not usually required.

Moving forwards in time from the human/chimp split it is possible to next look at the migration out of Africa of *H. sapiens* to the rest of the world, which is thought to have taken place 100-50kya (Cann 2002). This question has been

studied from many perspectives, such as dating the split between African and non-African populations (for example Goldstein *et al.* 1995a; Ingman *et al.* 2000) and identifying which hgs are associated with the expansions (for example Underhill *et al.* 2001; Forster 2004). To investigate these events samples from each of the continents are required, and some method to ensure that diversity within each continent is captured, by sampling individuals based on ethnic group (Underhill *et al.* 2001) or language affiliation (Ingman *et al.* 2000). For example the "ascertainment set" of Underhill *et al.* (2000) initially employed a total of 53 individuals drawn from representative populations from each of the continents to assess the diversity and geographic distribution of the Y-chromosome lineages they identified. A further 1,009 samples were included to augment the phylogenetic tree.

The genetic study of the Neolithic period in Europe (starting around 10kya) has already been mentioned in several contexts. Most studies attempt to measure in some way the extent to which there has been gene flow from the Middle East. Under the demic diffusion model a Middle Eastern genetic influence is expected to follow a cline from southeast to northwest Europe (Cavalli-Sforza *et al.* 1994), therefore the best sampling strategy to address this question would involve sampling a range of populations from the southeast to northwest of Europe. It is not sufficient to analyse a genetic sample of Europeans of unknown provenance. For example Rosser *et al.* (2000) analysed Y-chromosome variation across Europe by sampling from 47 populations ranging from Iceland in the northwest to North Africa in the southeast. It was possible to detect a significant cline in the distribution of hg J, which appeared to be consistent with a Neolithic demic expansion. Some consideration of the origin of the sample donors would also be beneficial in studying Europe, for example it would be counterproductive to knowingly sample from Middle Eastern communities in western Europe as this might artificially inflate the estimation of Neolithic gene flow from the Middle East. It is within the context of the Neolithic debate that an important example regarding the choice of sample population can be found. Based on an analysis of mtDNA RFLP and sequence information, Richards *et al.* (1996) concluded that the Neolithic had had little impact on the female gene pool of Europe. A Bedouin sample was employed to represent the Middle East,

however subsequent analysis revealed that Bedouins may not be the most representative Middle Eastern population, hence potentially affecting the initial conclusions drawn from this dataset (Richards *et al.* 2000). This also highlights a more general point that the decision to use a particular extant population to represent a past population is based on evidence that is only as good as current knowledge or interpretations. Therefore it is apparent that as one moves towards the present and the questions being addressed in genetic history are typically better refined and involve more closely related populations, increased attention has to be paid to sampling procedures. This ensures that the samples collected are as representative as possible, of the past populations being studied.

The use of ancient DNA collected from skeletal remains of the period and populations of interest might be a more direct way of ensuring that samples from the appropriate time and place are obtained, particularly as one expects the fossil record (Klein, 1999) and DNA preservation to be better for more recent events (Smith *et al.* 2003). However, contamination from a variety of sources (Nicholson *et al.* 2002) will still pose a problem, as will the size of the amplifiable fragment, although the fragment size is expected to increase with better preservation (Smith *et al.* 2003). Additionally it is unlikely that Y-chromosome analysis could occur because the high copy number of mitochondria per cell means that it is usually mtDNA that is amplified in ancient DNA studies. Assuming both of these problems could be dealt with, one would also have difficulty in obtaining large enough sample sizes to have significant statistical power, regardless of the age of the ancient DNA (but see Vernesi *et al.* 2004 who were able to successfully amplify ancient DNA from 80 Etruscan remains). Furthermore one still encounters the problem of identifying which skeletal remains belong to which population, even when remains are associated with potentially informative grave goods, as in Anglo-Saxon London for example (Stevenson 1998).

In conclusion genetic history has tended to focus on the human prehistoric period, particularly questions relating to Out of Africa and the Neolithic expansion in Europe. The reason for such a skew in the focus of genetic history is a result of the history of discovery of informative genetic markers to study

more recent events, and the prevailing interest in modern human origins. However it is now possible, and indeed potentially highly informative to study recent events. Indeed there is a feeling that the main questions and issues arising from the Out of Africa debate has been successfully addressed using available genetic markers (i.e. most researchers now accept that modern human evolution and subsequent population of the continents was an Out of Africa type event rather than through Multiregional evolution (Lahr and Foley 1998; Stringer 2003)), although of course new information may come to light with the development of new markers, and ancient human history may be re-addressed. More recent events such as the Neolithic expansion are still open to debate however (for a brief review see for example Renfrew 2000) Studies of recent history do however face a series of questions relating to sample strategy, which must be considered with care.

1.6. Aims of this Thesis

The work in this thesis addresses the recent history of several populations, who are found in close geographical proximity to each other, using a combination of Y-chromosome and mtDNA techniques. The aims of each of the chapters are briefly summarised here.

The first work presented in this thesis (Chapter 2) studied the British male population using Y-chromosome microsatellites and binary markers:

- To assess the relative influence of invading populations (Anglo-Saxons and Vikings) and indigenous Europeans on the British male gene pool. Anglo-Saxons and Vikings have had a known cultural and linguistic influence on the British Isles but the extent to which they contributed Y-chromosomes to the gene pool is unknown. A novel sample strategy was employed to comprehensively sample men from across Britain in higher resolution than prior studies to ascertain any small regional differences in Y-chromosome frequencies.

Chapters 3 and 4 drew on this comprehensive picture of Y-chromosome diversity in Britain to investigate even more recent events in the male British population:

- Chapter 3 investigated the fidelity of surname inheritance of men with British surnames using Y-chromosome microsatellites and binary markers. As surnames are inherited patrilinearly, mimicking Y-chromosome inheritance, the Y-chromosome is ideal to study surname history.
- Chapter 4 examined Y-chromosome diversity in London, a large metropolitan district, to assess whether it reflected the known history of immigration to the city, or if the Y-chromosome diversity was similar to that in the rest of Britain. London mtDNA diversity was also assayed, to examine any differences in male and female history, given the utility of such comparisons in other investigations.

Having studied the British population and the influence of documented migrations from a predominantly male perspective it was interesting to assess a very different population:

- Chapter 5 studied the maternal history of the Lemba, an African population who, based on oral tradition and evidence from the Y-chromosome, are hypothesised to have migrated from the Middle East and have potential Jewish ancestry.

Chapter 2. A Y-Chromosome Census of the British Isles

Some of the results presented in this Chapter appear as Capelli, C. Redhead, N., Abernethy, J. K., Gratrix, F., Wilson J. F., Moen, T., Hervig, T., Richards, M., Stumpf., Underhill, P. A., Bradshaw, P., Shaha, A., Thomas M. G., Bradman, N., Goldstein, D.B. (2003). A Y-Chromosome Census of the British Isles. *Curr. Biol.* **13:** 979-984.

2.1. Introduction

British history has been punctuated by periods of cultural change often associated with the migration of peoples from the continent. The extent to which these events correlate with population replacement has been subject to much debate (Davies 1999; Graham-Campbell and Batey 1998; Richards 2000) in both the archaeological and historical literature. Before the processual school of archaeology emerged in the 1960s, archaeological evidence of cultural change was usually interpreted as evidence of mass immigration. European archaeological thought in particular has tended to associate the distribution of cultural traits with the movement of ethnic groups (Burmeister 2000). However processual archaeology effectively turned the migrationist view around and argued that cultural change could happen through trade or the effect of a small ruling elite, neither of which would have a dramatic influence on the gene pool. Indeed, it argued that only positive physical evidence for migration should be used to conclude that migration occurred (Welch 1992). Recently, archaeological thought seems to be coming full circle and models of migration to explain cultural change are being reconsidered (Burmeister 2000). Indeed in the archaeological context, migration *per se* does not seem to be well understood or researched (Burmeister 2000) and has been used as a lazy way of explaining cultural change (Anthony 2000).

As previously discussed in the Introduction (Chapter 1), genetics is an important tool to be exploited in studying the recent history of human populations, augmenting and informing questions raised by other disciplines. The history of migration to the British Isles is a question that is particularly suited to genetic investigation, especially as the questions are reasonably well defined, as will be shown in the following paragraphs. This description of the history of migration to the British Isles will be followed by a discussion outlining the choice of markers used in this Chapter and a summary of current knowledge of relevant genetic diversity.

2.1.1. The History of Migration to the British Isles

The first (potential) mass invading force to be felt in Britain was the Roman invasion of the 1st century AD and subsequent rule until 410AD when most of the Roman army was withdrawn to campaign in Gaul and Spain (Welch 1992). However the heterogeneity of the Roman Army means it is impossible to identify a single source population, therefore they will not be considered further. Instead the focus turns to Anglo-Saxons (and Jutes) and Norwegian and Danish Vikings. From the 5th to 8th centuries AD Angles, Saxons and Jutes invaded (according to historical sources) and settled primarily in southern and central Britain eventually forming several kingdoms: South Saxon Kingdom (Sussex), West Saxon Kingdom (Wessex), and the East Saxon Kingdom (Essex). Anglo-Saxons are thought to have come from northern Germany and southern Scandinavia (Welch 1992); specifically Schleswig-Holstein in northern Germany has been identified as the source of Angles, whilst Saxons and Jutes came from Jutland (Davies 1999). It is the relative merits of a supposed mass migration of Anglo-Saxons that has been particularly hotly debated by archaeologists and historians. A full treatment of the various arguments for and against migration is beyond the scope of this thesis, however the salient point here is that it is predominantly the southern and eastern parts of England that were affected by Anglo-Saxons, whether by elite dominance of the local peoples, or mass migration (Welch 1992).

Norwegian and Danish Viking raids were part of British life for 300 years from the end of the 8th century (Richards 1991). Norwegians and Danes tended to follow distinct sea routes when travelling away from their native lands towards Britain: Norwegians moved along the northern and western coasts of Britain and the Danes moved southwards mainly concentrating on England (Richards 1991; Davies 1999). The western route included stops on the islands of Shetland, Orkney, the Hebrides, the Isle of Man and Angelsey, as well as in other parts of the United Kingdom such as Ulster, South Wales, Lundy and Cornwall (Davies 1999; Hill 1981). The Isle of Man is a particularly interesting case; from the little documentary evidence that exists it does not seem to be part of any known

kingdom before the Viking age and is believed to have been inhabited by Celtic speaking peoples. Two differing theories about the effect of Norsemen on the isle reflect general arguments for and against the role of mass migration: the first argument proposes a mass immigration of Vikings who replaced the existing population at all levels of society and the alternative posits only a replacement of the ruling elite (Richards 1991). Whichever scenario is closer to the truth it is certainly the case that Vikings had a big influence on the Isle of Man that has persisted to the present day in the form of the Tynwald, or Manx Parliament which meets every summer to announce laws passed in the previous year (the term Tynwald derives from the Scandinavian word *Thing* meaning assembly [Richards 1991]).

Viking raids on England have been classified into 3 phases, only the second of which is thought to have seen permanent colonisation (Richards 1991) and the establishment of the Danelaw, which divided England into those areas that were inside and outside the limits of Danish influence (Wormald 1991). The line demarking the two areas ran from north of London to Chester (Wormald 1991) South and West Saxon were outside the Danelaw, East Anglia, Danish Mercia, and York inside (Davies 1999). In Ireland a particular focus of Viking activity was in Dublin which was founded by Norwegian Vikings around 840 and grew to form a base from which to attack Britain (Davies 1999).

The Norman Conquest of 1066 can be considered "the last wave of the Northmen [Vikings]" (Davies 1999, p247) as much as the first influx of the French. Normandy was founded by Danish Vikings (although the later Duchy of Normandy also contained Roman, Carolingian, and Frankish elements [Rowley 1997]), who under the leadership of Rollo, settled around the estuary of the Seine. The Norman period is generally considered as affecting and involving the aristocratic, ecclesiastical and mercantile classes rather than the majority of the English people who were under the rule of these classes (Rowley 1997). For example English manors were passed into Norman hands and the whole ruling class of the Church and State spoke French (Davies 1999). Indeed some archaeologists argue that without the documentary evidence for a Norman Conquest archaeology alone could not show this event (Rowley 1997).

Compared to the earlier periods of Anglo-Saxon and Viking influence the Norman era was short and is considered to come to an end in 1154, less than a hundred years after the initial conquest, with the death of King Stephen (Rowley 1997).

The Channel Islands are often treated as somewhat of a footnote to British history, perhaps because of their location (Figure 2.1) close to the French mainland, their historically strong political links with France (see below) or their desire to retain political distance from the British mainland. For example, Jersey is not part of the United Kingdom, despite being in the British Isles (http://www.bbc.co.uk/jersey/about_jersey/general_info/about_government.shtml; 30th March 2004), and Sark is under the control of the Seigneur of Sark, effectively the (only) Lord of the Manor, making Sark a feudal society, indeed one of the few remaining in Europe (http://islandlife.org/history.htm; 30th March 2004). Despite their somewhat lowly status in the context of British history, they have a complex and intriguing history that deserves closer attention. The Channel Islands are located in the English Channel with the closest landmass being France (Figure 2.1) and are British Dependencies, although all retain a degree of autonomy. There are four main islands (in order of size): Jersey, Guernsey, Alderney, and Sark, all of which are densely populated (Lempriere 1976), the other two much smaller islands are Herm and Jethou. As DNA samples were only collected from Jersey and Guernsey (see Materials and Methods, section 2.2.1, below) these two islands will be described. The earliest evidence for settlement on the islands is 186-127kya during an interglacial period when the site of Cotte St. Brelade on Jersey was occupied (Klein 1999); there is not evidence for comparable occupation of Guernsey (Bender 1986). More consistent occupation started during the Neolithic around 4,000 BC, during which time the islands were part of an extensive alliance network with the British and French mainlands (Bender 1986). The geographical location and history of Jersey and Guernsey are distinct (Briggs 1995), it has been suggested that during the Neolithic Jersey and Guernsey were differentially influenced by peoples from the French mainland and Iberia respectively due to their locations (Hawkes 1937); Jersey is much closer to France, whilst Guernsey lies further west into the Atlantic.

Figure 2.1. Map Showing the Channel Islands in Relation to Britain and France.

Local Channel Island historians believe that Danish Vikings raided and settled on the Islands during the middle of the 9th century as a natural extension of the raids that took place on the nearby French coast (Nicolle 1935; F Falle personal communication), although the mainstream opinion contradicts this (Graham-Campbell, personal communication). Indeed it is impossible to define with accuracy the exact nature and timing of any raids because of the lack of written records on the Islands before 1066 (Stevenson 1986). It is certainly possible that the Channel Islands experienced Danish Viking raids because of the southern sea route they are thought to have taken (Davies 1999). Similarly, although there is not any suggestion of Anglo-Saxon influence on the Channel Islands, it is feasible that they made some contact with the Islands. Strong ties between the islands and Normans on the other hand is indisputable; they were part of the Duchy of Normandy from the 10th to the 13th centuries and Norman influence seems to have been so strong as to obliterate previous laws and customs (Stevenson 1986), many Norman rules and laws still exist today and until recently the local dialect was based on Norman French (http://islandlife.org/history.htm; 30th March 2004).

World War II saw the occupation of the Channel Islands by German Forces (1940-1945). Around half of the inhabitants of Guernsey chose to evacuate the island, whilst approximately only one fifth of Jersey evacuated (Briggs 1995) leaving the majority of Jersey locals on the island (Myhill 1964). Large numbers of German troops were involved in the occupation, for example by the end of 1941 there were 11,500 troops on Jersey (Briggs 1995) and towards the end of occupation the number of troops on Guernsey was roughly equal to the number of local civilians (around 21,000) (http://www.thisisguernsey.com/code/showarticle.pl?ArticleID=000023; 30th March 2004). Slave labour from across Europe (Spain, Poland, Russia, Ukraine) was brought in during the war to build fortifications at the western frontier of German rule (Briggs 1995). As is bound to happen during occupation, some of the occupied are more congenial towards their occupiers than others, which may explain why the illegitimacy rate in Guernsey increased from 5.4% prior to World War II to 21.8% by 1944. After World War II both islands saw an influx of immigrants from Britain

(http://www.bbc.co.uk/jersey/about_jersey/history/history_germanoccupation2.shtml; 30th March 2004), but from the 1970s onwards regulations on Jersey have been tightened and immigration kept to a minimum. Today it is primarily wealthy people that migrate to Jersey, in particular, because the locally set taxes make the island a tax haven. There is also a small but growing number of mainly Spanish and Portuguese immigrants employed in menial labour (http://www.bbc.co.uk/legacies/immig_emig/channel_islands/jersey/article_2.shtml; 30th March 2004).

2.1.2. The Y-Chromosome – An Overview

The previous paragraphs have described historical and archaeological evidence for the impact of Anglo-Saxons and Norwegian and Danish Vikings on the British Isles. This section will discuss the choice of genetic locus used to study the British Isles in this Chapter, and review current knowledge of their distribution across Britain and relevant regions of Europe. The Y-chromosome was chosen to study the genetic history of Britain as (i) the Y-chromosome has been shown to display higher levels of geographic structuring within Europe than mtDNA (see for example Rosser *et al.* 2000 compared to Simoni *et al.* 2000), hence had the potential to differentiate better between geographically close populations, such as those of Britain and northern Europe; (ii) migration events are normally associated with men (Burmeister 2000), although female migration associated with patrilocality, has been reported (Seielstad *et al.* 1998). Moreover earlier Y-chromosome work on 3 British populations (Orkney, Ireland and Wales) revealed differentiation between these populations and the ability to detect a hypothesised Norwegian Viking influence (Wilson *et al.* 2001a). European Y-chromosome history is well documented; the last four to five years has seen the publication of a huge range of studies, several of which have taken a continent-wide perspective, such as Casalotti *et al.* (1999), Malaspina *et al.* (2000), Roewer *et al.* (2000), Rosser *et al.* (2000), Semino *et al.* (2000), Wells *et al.* (2001), Cruciani *et al.* (2004). Such work allows one to make inferences about the distribution of Y-chromosome types in Britain. However the small

number of (British) samples involved means that only a partial description of British Y-chromosomes is possible. Three studies have focussed on British populations (Wilson *et al.* 2001a; Hill *et al.* 2000; and Weale *et al.* 2002), however a study to examine Y-chromosome diversity across the entirety of the British Isles has not yet been completed. The following paragraphs will first review the Y-chromosome and the markers used in studies of genetic history before turning to describe the relevant findings of the studies of European and British Y-chromosomes mentioned above.

The Y-chromosome is one of the 2 mammalian sex chromosomes, X and Y; males have an X and Y-chromosome and females have 2 X-chromosomes. Through the presence of the SRY (sex determining region on the Y-chromosome) gene, thus its gene product, the Y-chromosome confers maleness by initiating a cascade of events, which induce differentiation of the bipotential gonads into testes (Brook and Marshall 1996). The Y-chromosome is paternally inherited, and for ~95% of its 60Mb length it escapes recombination in the region termed the male-specific Y, or MSY (Skaletsky *et al.* 2003), which is where the markers used in genetic history are located (Jobling and Tyler-Smith 2003). Here the term Y-chromosome is used to also mean the MSY, unless otherwise stated. The remaining 5% of the chromosome, the pseudoautosomal region (PAR) does recombine with the X-chromosome (e.g. Skaletsky *et al.* 2003). The MSY is also known as the non-recombining Y (NRY) in much of the literature due to its lack of reciprocal recombination, however the apparent finding of gene conversion (non-reciprocal recombination) in this region has prompted use of the term MSY (Skaletsky *et al.* 2003), which is used here. For many years the MSY was thought to be functionally redundant, however many more Y-linked markers have now been identified, some of which are implicated in conditions such as gonadal sex reversal, Turner syndrome, graft rejection and spermatogenic failure (Skaletsky *et al.* 2003, and references therein), some of which are illustrated in Figure 2.2.

A physical map of the Y-chromosome was published in 1992 (Foote *et al.* 1992), building upon a deletion map of the Y-chromosome (Vollrath *et al.* 1992) published earlier in the same year. These studies started the characterisation of

a

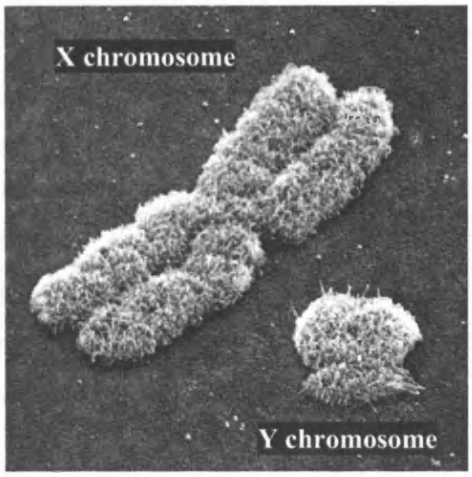
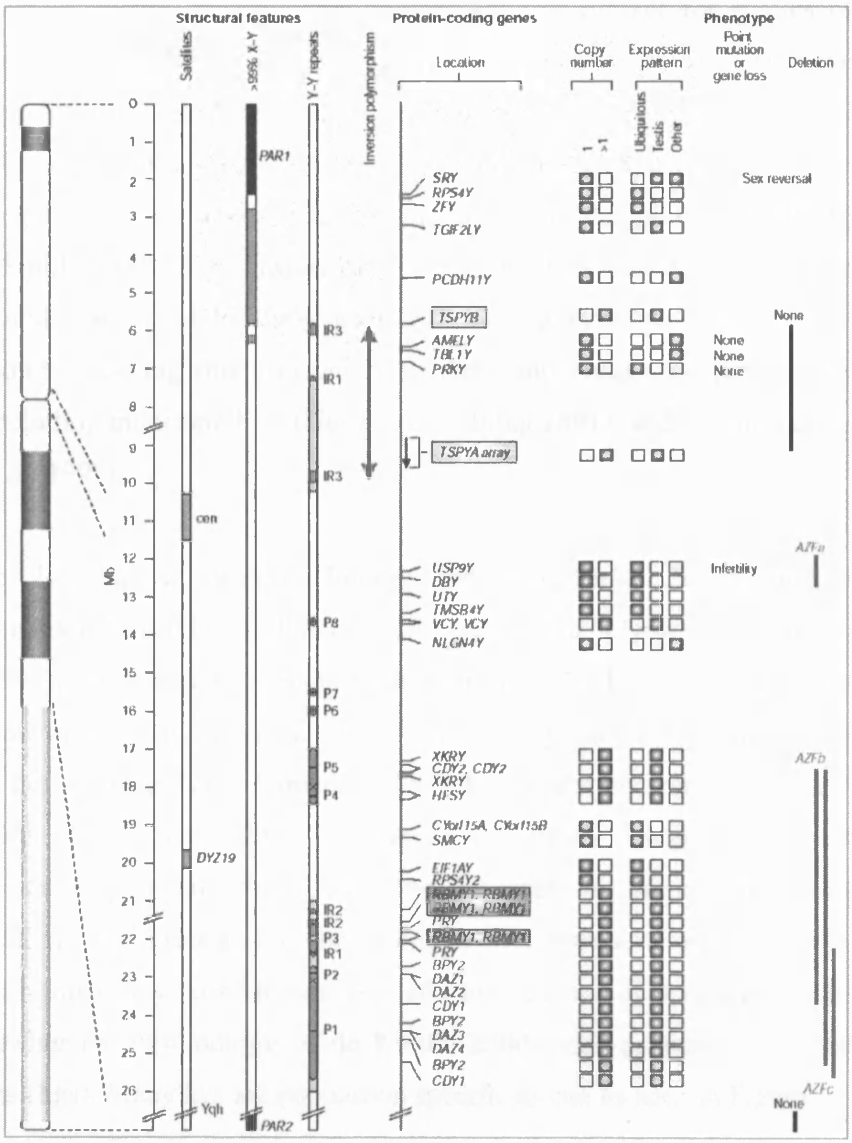


Figure 2.2. The Human Y-Chromosome.

a) Image of the human Y-chromosome in comparison to the X-chromosome magnified ~10,000 times. The small size of the Y chromosome can be seen. *Figure taken from Willard 2003.*

b) Schematic representation of the human Y chromosome. The pseudo autosomal regions (PAR1 and PAR2) can be seen, the remaining length of the Y chromosome is male specific (the MSY). Also indicated are known protein-coding genes and associated phenotypes. *Figure taken from Jobling and Tyler-Smith (2003).*

b



the Y-chromosome and highlighted its sequence complexity, which eventually enabled subsequent researchers to identify markers suitable for studies of human genetic history. For example the presence of different types of sequences means that the Y-chromosome can be targeted to identify polymorphisms with different mutation rates. This property has made the Y-chromosome particularly amenable in studies of population history, as will be seen below. Recently, the almost complete sequence of the MSY has been published (Skaletsky *et al.* 2003). This is an important step towards an even greater understanding of the Y-chromosome's structure and function and will undoubtedly lead to extremely high resolution analysis of human populations. Therefore the picture today is now very different and the number of useful markers has increased steadily. The first Y-linked polymorphism ($p12f_2$) was identified almost 20 years ago (Casanova *et al.* 1985). Although the potential of this marker for studies of genetic history was acknowledged at the time (Casanova *et al.* 1985) the identification of other markers was slow (Hurles and Jobling 2001). Now however the situation is quite different and the Y-chromosome has been considered the "best characterised haplotypic system in the genome" (Jobling and Tyler-Smith 2000). Two classes of Y-linked markers that are assumed to evolve neutrally are used to study male genetic history either singly, or in tandem: slowly evolving (mainly) binary markers and rapidly evolving DNA stretches including microsatellites (Hurles and Jobling 2001), and a minisatellite (Jobling *et al.* 1998).

More than 200 binary markers (Jobling and Tyler-Smith 2003), several insertions and deletions (Underhill *et al.* 2000) and an Alu polymorphism (YAP, Hammer 1994), are now known on the Y-chromosome. The low mutation rate of these markers means that they are assumed to have only mutated once in human evolution, hence have been termed Unique Event Polymorphisms (UEPs) (Jobling and Tyler-Smith 1995, Thomas *et al.* 1998). Furthermore their mutational stability means that UEPs can be used to define groups of genealogically related groups of Y-chromosomes who share a common ancestor more recently than they do with members of other hgs, as illustrated in Figure 2.3a. Following the terminology of de Knijff (2000) such groups are termed haplogroups (hgs). Many hgs are population specific as can be seen in Figure

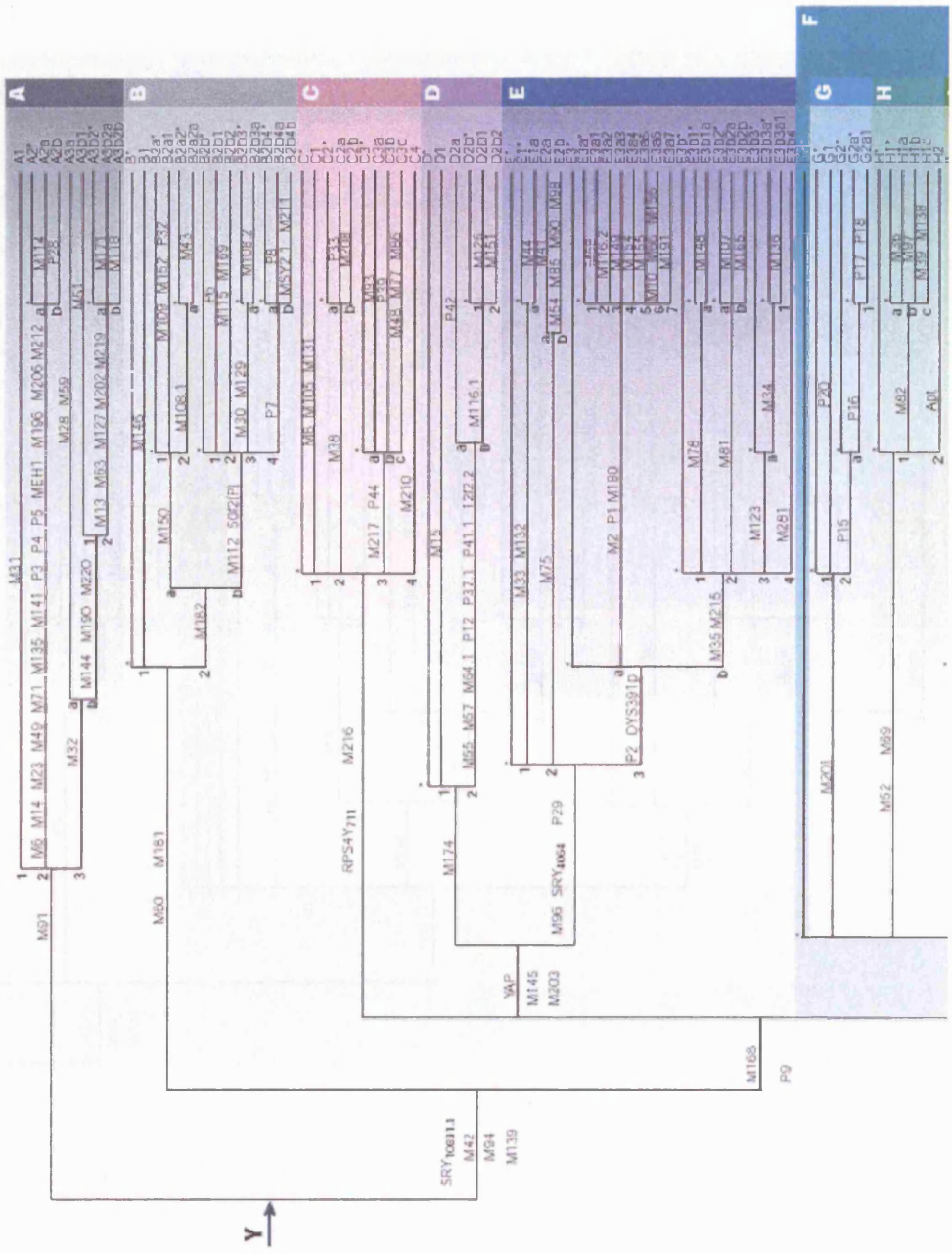


Figure 2.3a The YCC (2002) Tree of the Most Parsimonious Relationship of 153 Haplogroups. Legend on following page

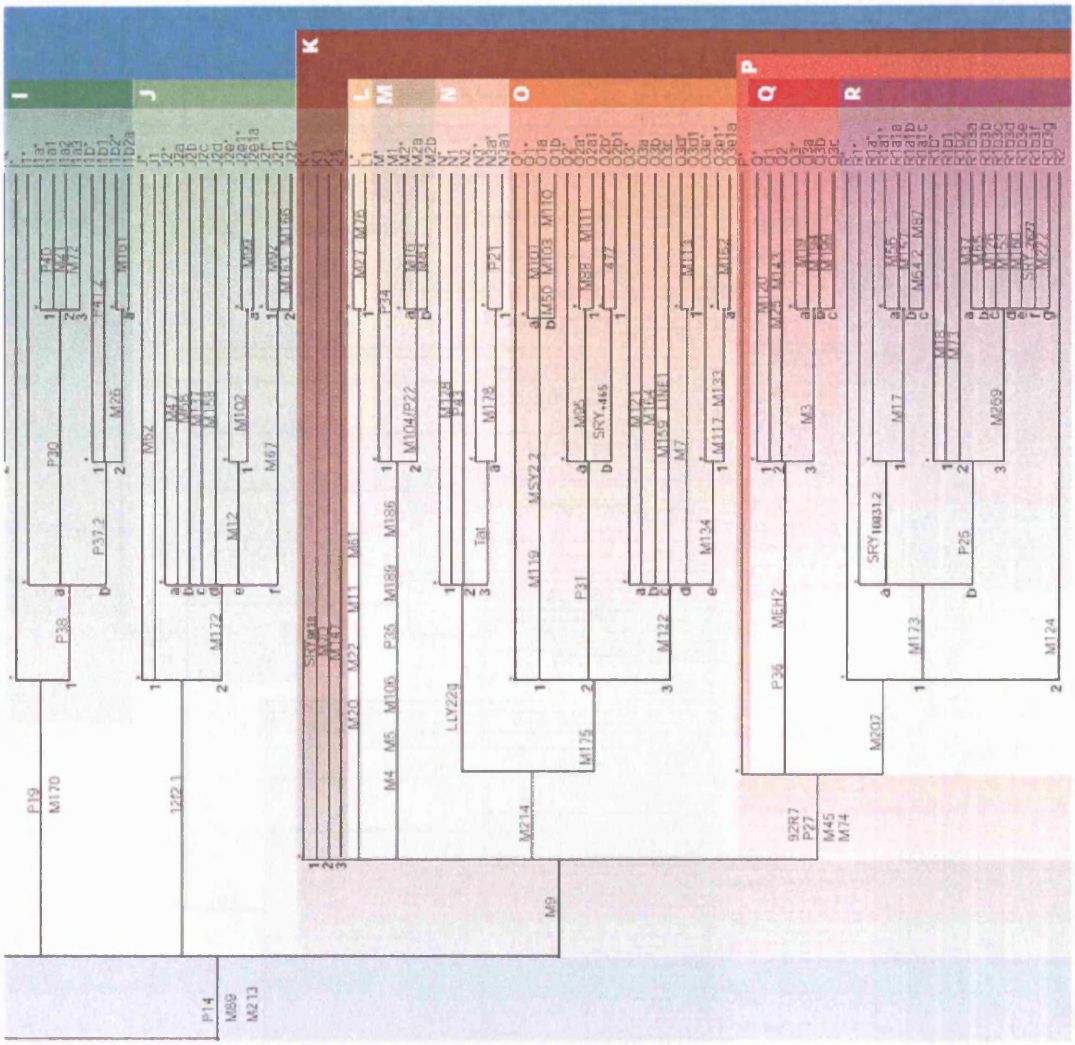


Figure 2.3a. Continued. The root of the tree is shown with an arrow at the top left of the diagram. The entire tree is called "Y", and major clades are labelled with capital letters. Hg names are indicated at the tips of the branches, and defining mutations are written along the lengths of appropriate branches, the order in which phylogenetically equivalent markers are placed is arbitrary. Branch lengths are not proportional to the number of mutations or age of mutations. *Figure taken from Jobling and Tyler-Smith (2003).*

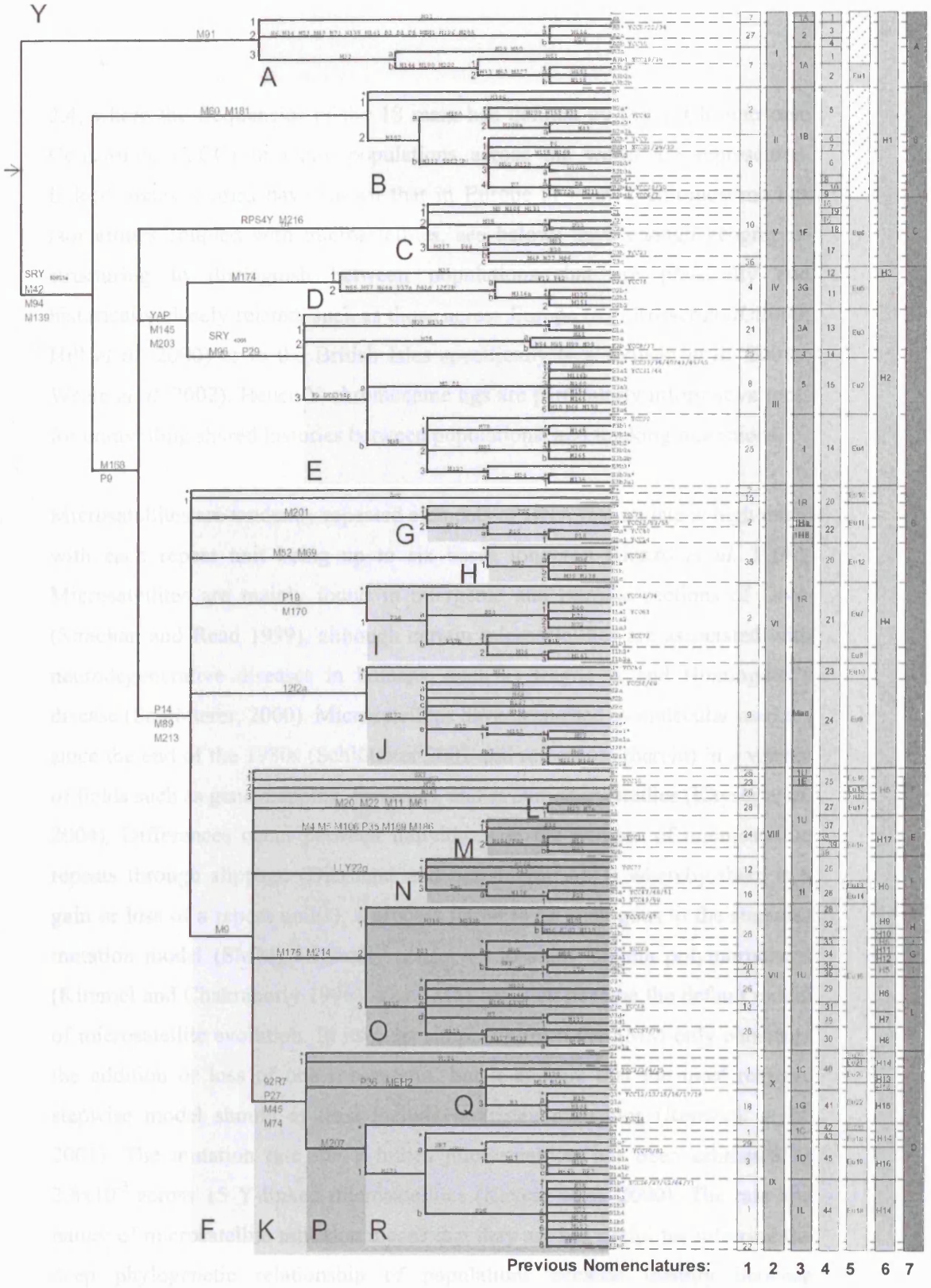


Figure 2.3b. Previous Nomenclatures for Y-Chromosome Hgs Illustrated by Using the YCC Tree. Previous nomenclatures have been taken from the following publications: (1) Jobling and Tyler-Smith (2000) and Kaladjeva *et al.* (2001); (2) Underhill *et al.* (2000); (3) Hammer *et al.* (2001); (4) Karafet *et al.* (2001); (5) Semino *et al.* (2000); (6) Su *et al.* (1999); (7) Capelli *et al.* (2001). *Figure modified from YCC (2002).*

2.4, where the frequencies of the 18 main hgs defined by the Y Chromosome Consortium (YCC) in many populations across the world are represented. Indeed, many studies have shown that in Europe at least Y-chromosome hgs (sometimes coupled with microsatellites, see below) have enough geographic structuring to distinguish between populations that are physically and historically closely related, such as those across Europe (e.g. Rosser *et al.* 2000; Hill *et al.* 2000) or in the British Isles specifically (e.g. Wilson *et al.* 2001a; Weale *et al.* 2002). Hence Y-chromosome hgs are particularly informative tools for unravelling shared histories between populations, and tracking migrations.

Microsatellites are tandemly repeated elements of DNA that mutate at high rates, with each repeat unit being up to six bases long (Di Rienzo *et al.* 1998). Microsatellites are mainly found in intergenic and intronic sections of DNA (Strachan and Read 1999), although certain microsatellites are associated with neurodegenerative diseases in humans, such as fragile X and Huntingdon's disease (Schlötterer, 2000). Microsatellites have been used as molecular markers since the end of the 1980s (Schlötterer 2001 and references therein) in a variety of fields such as gene mapping, forensics, and evolutionary studies (Kayser *et al.* 2004). Differences occur between individuals in the number of microsatellite repeats through slippage (Dieringer and Schlötterer 2003) whereby there is a gain or loss of a repeat unit(s), a process found to be analogous to the stepwise mutation model (SMM) originally conceived to study protein polymorphisms (Kimmel and Chakraborty 1996). The SMM has now become the default model of microsatellite evolution. In its most simplistic form the SMM only considers the addition or loss of one repeat unit, but it appears that the most realistic stepwise model should at least include multistep mutations (Renwick *et al.* 2001). The mutation rate for Y-linked microsatellites has been estimated as 2.8×10^{-3} across 15 Y-linked microsatellites (Kayser *et al.* 2000). The rate and nature of microsatellite mutation means that they are not useful for inferring the deep phylogenetic relationship of populations because identity between individuals in the repeat count may be due to state rather than descent. Microsatellites are thus used to study more recent evolutionary relationships (Hammer and Zegura 2002 for example). However Gusmão *et al.* (2003) did find evidence for geographic structuring of human populations in Europe using

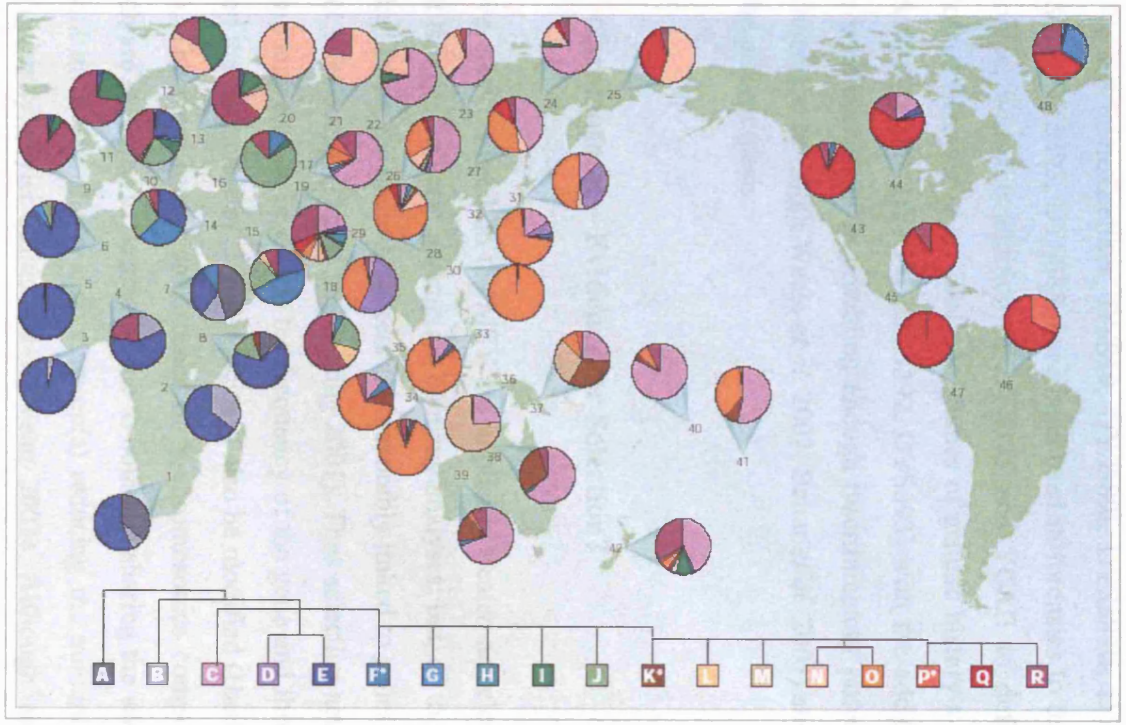


Figure 2.4. The Geographic Distribution of the Main Y-Chromosome Hgs. Each colour in the pie charts represents the frequency of the YCC hgs in each population. The YCC tree is summarised at the bottom of the Figure, and can be found in full in Figure 2.3. Numbers next to each pie chart refer to references the populations studied, for brevity these references are not listed here and can be found in Jobling and Tyler-Smith 2003 (Figure 2). *Figure modified from Jobling and Tyler-Smith (2003).*

only microsatellites. Nonetheless UEPs are still preferentially used to define hgs. For consistency, the allelic states defined by microsatellites are termed haplotypes (de Knijff, 2000).

A recent study to identify polymorphic Y-linked microsatellites has increased the number of known microsatellites with a repeat unit ≥3 to 139 (Kayser *et al.* 2004). Seven "core" microsatellites, DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, have been used in forensics to define a minimal haplotype, with the addition of DYS385 and YCAII to define an extended haplotype (Roewer *et al.* 2001). In studies of genetic history a subset of these (DYS19, DYS390, DYS391, DYS392, DYS393) with the addition of DYS388 have been successful in providing enough discriminatory power (see for example Thomas *et al.* 2000; Weale *et al.* 2002; Behar *et al.* 2003) and have been used in the present thesis.

2.1.3. The Y-Chromosome – Evidence for Selection?

As noted above, markers used on the MSY are implicitly treated as selectively neutral, which is indeed a basic assumption of many analyses; but, as the MSY is a single linked locus, the neutral markers are inexorably linked to genes which may be subject to selection (Hurles and Jobling 2001). Thus selection acting on a Y-linked gene may elevate or reduce the frequency of the gene and through a selective sweep the frequency of all other loci will also be modified (Hurles and Jobling 2001). The more recent coalescence of Y-chromosomes compared to mtDNA (which are expected to be similar due to both loci sharing the same N_e) is one argument in favour of a selective sweep(s) reducing the amount of Y-chromosome diversity (Tyler-Smith and McVean 2003). Although as these authors and others point out, alternative variables may also explain the discrepancy. These include differences in male and female reproductive behaviour (Tyler-Smith and McVean 2003), and heterogeneity of mutation rate between different sites in the mitochondrial genome (Cavalli-Sforza and Feldman 2003), which may be compounded by the substantial variation in

published estimates of the mutation rate (see for example Sigurðardóttir *et al.* 2000) in the mitochondrial genome. Several studies have looked at evidence for an association between specific diseases, known or hypothesised to be Y-linked, and the Y-chromosome lineages used in population studies. For example Quintana-Murci *et al.* (2003) tested for association between testicular cancer and Y-chromosome lineages in British men and found no evidence for an association. In contrast Krausz *et al.* (2001) found that reduced sperm count was associated with a Y-chromosome lineage KR*(xP,R1a1) in Danish men, which has the potential to affect frequencies of this lineage.

The possibility that the Y-linked markers used in population studies are not selectively neutral must now be entertained in the light of new evidence for selection (Krausz *et al.* 2001). Indeed as increasing numbers of genes are identified on the Y-chromosome (Skaletsky *et al.* 2003) the chance that selection is acting on some of these, hence affecting the frequencies of the assumed neutral alleles must also increase. Work on this subject is in its infancy however, and published research to date does not overwhelmingly show that disease phenotypes are associated with Y-chromosome lineages (Quintana-Murci *et al.* 2003). Therefore this thesis treats the markers described, and their resultant geographic distribution, as neutral, whilst acknowledging the fact that future work may indicate a different picture.

2.1.4. Y-Chromosome Nomenclature

The recent publication by the YCC (YCC, 2002) proposed a new nomenclature system for the Y-chromosome, which has been adopted by most laboratories and is used in the present work. As this nomenclature is still new to the field, and many of the studies that are cited in thesis were published prior to the new nomenclature, a brief summary of the suggestions is provided here before proceeding to describe Y-chromosome diversity in Britain. The YCC typed a global sample of 74 individuals for 234 polymorphic Y-linked UEPs resulting in 153 different hgs. Note that the apparent discrepancy between the greater

number of haplogroups than individuals sampled is explained by the fact that many haplogroups are hierarchical as illustrated in Figure 2.3a. Using several primate species as outgroups, the most parsimonious tree was constructed, which is illustrated in Figure 2.3a. Major clades were assigned to groups A-R and nested subclades classified by an alternating number/letter system, which is used throughout this thesis unless otherwise specified. For example, chromosomes sharing the M60/M181 derived state are placed into B, chromosomes that are further derived at M182 are called B2, chromosomes derived at M60/M181, M182, and M150 are called B2a and so on. Chromosomes that are only derived at M60/M180 and no other internal markers within hg B are labelled as B*. The YCC suggested that this latter group of lineages, signified by an asterix, be termed a paragroup rather than a hg because they are not defined by further derived markers. A system was also devised by the YCC whereby the hg name also signifies which UEPs have and have not been typed by using the "x" notation. To take the example of hg B again; if chromosomes known to be derived at M60/M181 were then typed for all internal markers, underived chromosomes would be classified B* and the remaining chromosomes according to the most terminal marker they were found to be derived at. However, if the chromosomes were only typed for the internal marker M182, underived chromosomes could be either B* or B1, and derived chromosomes B2a, or B2b. Hence, with the latter example where only M182 was typed, underived chromosomes would be called B*(xB2) to signify that the marker defining the lineage B1 was not typed, and derived chromosomes called B2.

Prior to the introduction of the YCC nomenclature navigating between the various hg naming systems was an arduous task as there was not a uniform system used across laboratories (Gusmão 2003). This is clearly illustrated in Figure 2.3b, where 7 nomenclatures previously used in the literature are shown. Such differences in the choice of nomenclature also reflect to some extent, heterogeneity in the choice of UEPs markers between different laboratories, which also complicates cross-study comparisons. Table 2.1 shows the frequencies of three common hgs in British populations taken from several published studies. As can be seen several of the studies have employed different

Table 2.1. Frequencies of the Major Y-Chromosome Hgs in British Populations

	Frequencies		
Population	*P Lineages [Markers Used]*	*Hg2 and Equivalents [Markers Used]*	*R Lineages [Markers Used]*
	[P*(xR1a)]	[Y*(xDE,K)]	[R1a]
Scottish[a]	0.77	0.13	0.06
	[P*(xR1a,R1b8)]	[BR*(xB2b,CE,F1,H,JK)]	[R1a]
Western Scottish[b]	0.72	0.19	0.07
Scottish[b]	0.79	0.12	0.07
Cornish[b]	0.82	0.18	0
	[P*(xR1a1)]	[BR*(DE,JR)]	[R1a1]
Llangefni[c]	0.89	0.04	0.01
Abergele[c]	0.56	0.06	0
Ashbourne[c]	0.65	0.22	0.04
Soutwell[c]	0.64	0.19	0.06
Bourne[c]	0.67	0.33	0
Fakenham[c]	0.57	0.42	0
North Walsham[c]	0.56	0.31	0.04
	[M173]	[M170, M89]	[M17, M87]
Orkney[d]	0.65	0.08	0.27
Britain (Essex)[d]	0.72	0.24	0
	[P*(xR1a1)]	[BR*(xDE,LR)]	[R1a1]
Orkney[e]	0.66	0.14	0.2
Ireland[e]	0.85	0.09	0
Wales[e]	0.89	0.06	0.01

Note. Hg designations refer to the YCC 2002 terminology found in Figure 2.3, except hg2 (see section 2.1.5). Row totals do not total 1 as only the frequencies of the 3 commonest hgs are shown, several other hgs not shown here are also present.

[a] Helgason *et al.* (2000)

[b] Rosser *et al.* (2000)

[c] Weale *et al.* (2002)

[d] Wells *et al.* (2001)

[e] Wilson *et al.* (2001)

markers; the frequencies in this table are discussed in more detail below. Whilst the introduction of the YCC nomenclature is unlikely to promote uniformity in marker choice between laboratories, it at least means that a uniform nomenclature exists.

2.1.5. Y-Chromosome Diversity in the British Isles

Table 2.1 summarises the 3 hgs that comprise most of the Y-chromosomes observed in the British male population; heterogeneity in the choice of markers means that whilst hg frequencies from different studies are not identical there is enough overlap to allow comparison. Hgs defined by the broadly equivalent markers 92r7 and M173 are the most frequent Y-chromosome hgs found in Britain; depending on which internal nodes were typed (typically M17 and SRY_{1083b}) these lineages are assigned to hgs as shown in Table 2.1. In Ireland Wilson *et al.* (2001a) found that 0.89 of the samples belonged to P*(xR1a1), whilst Hill *et al.* (2000) found that the slightly lower resolution hg P reached a high of 0.983 in the Irish sample from Connaught. Scottish and Welsh populations have similarly high frequencies of these lineages (Table 2.1). The British sample of Wells *et al.* (2001), obtained from Castle Hedingham in Essex, shows a high frequency of P*(xR1a1). There is some indication of an east-west cline in frequencies of these lineages in England, with the highest frequencies in the west. For example, in Cornwall the frequency of P*(xR1a,R1b8) is 0.82, whilst in East Anglia it is 0.56 (Rosser *et al.* 2000). A similar cline can be seen in the transect of Wales and England studied by Weale *et al.* (2002), where the frequency of P*(xR1a) clearly decreases from west to east (with the exception of Abergele). This cline conforms to the general trend seen in the rest of Eurasia (Wells *et al.* 2001; Rosser *et al.* 2000; Semino *et al.* 2000), although it should be noted that frequencies in Scandinavia are typically lower than for more southerly populations on a similar easting (Rosser *et al.* 2000). Thus in the Scandinavian samples from Norway (Rosser *et al.* 2000), Wilson *et al.* 2001a), northern Sweden and Gotland (Rosser *et al.* 2000) the frequencies range from 0.29-0.17. Frequencies in western continental populations are slightly higher than in

Scandinavia, but still lower than in Britain and Ireland: 0.50 in Denmark, 0.40 in Germany (Rosser *et al.* 2001), and, 0.56 in Friesland (Wilson *et al.* 2001a). Concordance between 92r7 and M173 markers is suggested by the fact that the frequency of R1*(xR1a1) in the Norwegian sample of Passarino *et al.* (2002) is very similar to the frequency of P*(xR1a,R1b8) in the Norwegian sample of Rosser *et al.* (2000) and P*(xR1a1) in the Norwegian sample of Wilson *et al.* (2001a) (0.278, 0.29, 0.26 respectively). PR lineages (excluding R1a and R1a1) are also extremely common in Basques, possibly as a result of drift, with frequency estimates ranging from 0.73-0.90 (Rosser *et al.* 2000; Semino *et al.* 2000; Wilson *et al.* 2001a).

On the basis of 6 YSTR1 microsatellites (DYS19, 388, 390, 391, 392, 393) Wilson *et al.* (2001a) identified a modal haplotype in P*(xR1a1) chromosomes which was termed the Atlantic Modal Haplotype (AMH) because of its high frequency in Atlantic fringe European populations. Due to the high mutation rate of microsatellites and the instability associated with their mutation, so-called one-step neighbours of the modal haplotype were also included within the AMH cluster, consequently termed AMH+1 (the "+1" nomenclature means that one-step neighbours have been included in the modal cluster). One-step neighbours are defined as haplotypes that differ by one mutational step away from the modal haplotype. The frequency of AMH+1 haplotypes differed markedly between several European and Middle Eastern populations (Orkney, Ireland, Wales, Norway, Friesland, Basques, Syria and Turkey), being most common in the Basque and British populations, and was thus a useful marker to differentiate populations (Wilson *et al.* 2001a) particularly given the high frequency of P*(xR1a1) chromosomes. The frequency of AMH+1 chromosomes from an east-west transect of Britain can be inferred from Table 3 of Weale *et al.* (2002), where the modal cluster is found to be fairly constant at a frequency of ~0.30-0.45 apart from the sample from Llangefni (north coast of Wales) where AMH+1 chromosomes are particularly common (0.675).

Semino *et al.* (2000) concluded that R1*(xR1a1) is an ancient lineage associated with the Upper Palaeolithic population that entered Europe around 40kya (Semino *et al.* 2000) migrating from east to west from a homeland in Central

Asia where the ancestral lineage to R1*(xR1a1), defined by the mutation M45, is found (Underhill *et al.* 2001; Wells *et al.* 2001; Zerjal *et al.* 2002). However the Last Glacial Maximum (LGM) from 20-13kya would have caused any populations that had reached into the north and western-most parts of Europe populations to retreat into southerly refugia. The distribution of R1*(xR1a1) today, therefore, reflects expansion from these refugia, hypothesised to be in Iberia (Semino *et al.* 2000). The high frequencies of PR(excluding R1a and R1a1) in Basques supports this argument as Basques reside in the Iberian peninsula today and are assumed to be an isolated population (Calafell and Bertanpetit 1994; Comas *et al.* 2000; Cavalli-Sforza *et al.* 1994; Chikhi *et al.* 2002) who have not experienced substantial gene flow for millennia (but see Hurles *et al.* 1999), hence are considered to be good representatives of Palaeolithic Europeans. However as previously discussed (Introduction), some of the assumptions underlying such analyses may or may not be valid.

A series of overlapping and broadly equivalent hgs comprise the next most frequent Y-chromosome lineages seen in British populations (Table 2.1), which has been termed hg2 in the terminology of Jobling and Tyler-Smith (2000). As Table 2.1 and Figure 2.3a show, in most studies this hg is polyphyletic, containing representatives from at least 4 of the major clades of the phylogeny (B, F, G and I). Geographic resolution of this hg is therefore low. However, based on existing knowledge of the distribution of Y-chromosome diversity on a world-wide basis it is possible to infer that most of the hg2 chromosomes found in European populations either belong to hg F*, G or I, and given that F* and G individuals are very rare, one can hypothesise that most hg2 Europeans indeed belong to hg I. Note that the markers (M170) employed by Semino *et al.* (2000) and Wells *et al.* (2001) define a single non-polyphyletic hg (hgI) that is equivalent to hg2, but the fact that hgI is not polyphyletic means is can be considered to be of higher resolution than hg2. Due to the range of criteria used to classify this collection of Y-chromosomes (Figure 2.3b) the term hg2 is used here to generically refer to all of the broadly equivalent definitions used, the marker name is used when more resolution is required. As Table 2.1 shows, frequency differences can be seen for hg2 in Britain with the highest frequencies in the east of England, such as Fakenham (0.43) and Bourne (0.33), although

frequencies are never as high as for 92r7 and M173 derived lineages. Ireland and Welsh populations have the lowest frequencies.

Frequencies of hg2 in northern and north western continental Europe are similar to those seen in England: 0.33 and 0.44 in Norway (Rosser *et al.* 2000 and Wilson *et al.* 2001a respectively); 0.48 in northern Sweden, 0.59 in Gotland, 0.32 in Denmark, 0.20 in Germany (Rosser *et al.* 2000), and 0.29 in Friesland (Wilson *et al.* 2001a). The slightly higher resolution of the studies of Semino *et al.* (2000) and Passarino *et al.* (2002) yield generally similar results, Germany 0.375 (Semino *et al.* 2000), and 0.408 for Norway (Passarino *et al.* 2002). An additional downstream marker of M170 (M26) shows an extremely interesting distribution being found at high frequency in Sardinia (35.1%) and only elsewhere in Spanish and French Basques (at 4.4% and 9.1% respectively) which is explained as local differentiation of the M170 lineage. The I lineage is restricted to Europe and hypothesised to have originated in the European Palaeolithic around 22,000 years ago in descendents of men arriving from the Middle East (Semino *et al.* 2000).

As noted above R1a and R1a1 lineages are internal to 92r7 and M173, being defined by the derived state at SRY_{10831b} or M17 respectively. These lineages are essentially equivalent however as only small numbers of individuals (2 in Armenia (Weale *et al.* 2001) and 1 in Belarus (Behar *et al.* 2003)) have presently been found to belong to R1a. Thus R1a1 is primarily considered here. R1a1 is the 3rd commonest seen in most British populations, although the frequencies are typically much lower than the previous 2 hgs described. As Weale *et al.* (2002) show, R1a1 chromosomes are only observed in 3 of the English populations (Ashbourne, Southwell, and North Walsham), where it is always at low frequency (0.037-0.057). A similar picture emerges for Wales, where R1a1 is either absent (Weale *et al.* 2002) or again at low frequencies (~0.01, Wilson *et al.* 2001a, Weale *et al.* 2002). Irish Y-chromosomes are also characterised by an absence of R1a1 lineages (Wilson *et al.* 2001a) or low frequencies of R1a lineages (0.01; Rosser *et al.* 2000). In Scotland, however, R1a lineages are slightly more common, comprising 0.07 of the Y-chromosomes found in the "Western Scottish" and "Scottish" samples of Rosser *et al.* 2000. Orcadian Y-

chromosomes exhibit even higher frequencies of R1a1 lineages where they comprise 0.20-0.25 of sampled Y-chromosomes (Wilson *et al.* 2001a; Wells *et al.* 2001). R1a/R1a1 lineages are absent in Basques (Wilson *et al.* 2001a; Rosser *et al.* 2000).

Within Europe R1a and R1a1 hgs are distributed across Europe and Central Asia at varying frequencies, being most common in northern and central Europe and central Asia (Rosser *et al.* 2000; Wilson *et al.* 2001a; Passarino *et al.* 2002; Semino *et al.* 2000; Wells *et al.* 2001; Zerjal *et al.* 2001; Zerjal *et al.* 2002). Thus moderately high frequencies of R1a1 have been observed in Norway (0.236 and 0.26; Wilson *et al.* 2001a, Passarino *et al.* 2002, respectively) and in Scandinavia and Germany for R1a (0.31 in Norway, 0.19 in northern Sweden, and 0.16 in Gotland, 0.30 in Germany; Rosser *et al.* 2000). Central European populations have much higher frequencies of R1a1, such as Hungarians and Poles (0.60 and 0.54 respectively, Semino *et al.* 2000). The central Asian populations of the Kyrgyz and Tajiks both had high frequencies of R1a1 (~0.64; Zerjal *et al.* 2002) as did the Russian/Tashkent sample (0.47) of Wells *et al.* (2001).

R1a1 is thought to have its origins in Central Asia and spread to the rest of Eurasia through the development of Nomadic pastoralism in the Steppes and the domestication of the horse around 5kya (Wells *et al.* 2001; Zerjal *et al.* 2002). Due to the high frequency of R1a1 and particularly its modal haplotype 3.65+1 (Wilson *et al.* 2001a) in Norway, its presence in Britain has been interpreted as a signature of Norwegian Viking genetic influence (Wilson *et al.* 2001a) mediated by Viking invasions from the 8th to 10th centuries AD (Davies 1999). Given that the distribution of R1a1 is clearly not restricted to Norway it is feasible that the occurrence of this hg in Britain is the result of migrations from Central Europe or Central Asia, rather than, or in addition to, migrations from Norway. However the documented historical links Britain has with Vikings (Davies 1999), rather than with Central European and Central Asian populations, means that the interpretation of the data by Wilson *et al.* (2001a) is more plausible. This example highlights an important problem with historical inference from genetic data: in the absence of a historical, archaeological, or

palaeoanthropological context within which to place genetic information one has to make assumptions about a population's history based on the extant distribution of genetic diversity. These assumptions may well be false. Hence, wherever possible an interdisciplinary approach should be pursued.

There is also a range of haplogroups found at lower frequency in western Europe that one might expect to sporadically see in a large enough sample of the British male population. These hgs will now be briefly discussed based on the order in which they appear on the YCC tree. Hg E is defined by the derived state at SRY_{4064}, M96, or P29. Although the group as a whole is primarily found in Africa (Cruciani *et al.* 2004), the hg E3b defined by the additional derived state at M35 is thought to have an East African origin and is found in African, Middle Eastern and Mediterranean populations (Semino *et al.* 2002; Cruciani *et al.* 2004; Semino *et al.* 2004). Hg J is defined by the derived state at 12f2; some laboratories additionally use the M172 marker which is internal to 12f2 and defines the hg J2. J and J2 hgs reach high frequencies in the Middle East (Hammer *et al.* 2000; Rosser *et al.* 2000; Semino *et al.* 2000; Quintana-Murci *et al.* 2001) and Central Asia (Zerjal *et al.* 2002) as well as Jewish populations (Hammer *et al.* 2000). Weale *et al.* (2002) found J chromosomes at low frequency (0.013-0.057) in 4 of the 7 Welsh and English populations they studied, and Wilson *et al.* (2001a) found hg J chromosomes in 0.01 of the Welsh sample. The cline in frequencies of hg J in Europe, from high frequencies in the east to lower frequencies in the west, has been interpreted as a signature of the Neolithic expansion of farmers from the Middle East (Underhill *et al.* 2001). The final hg considered in detail here is N3, defined by the Tat mutation. This hg is particularly common in Finno-Ugric speakers, such as the Saami, Finns, and Mari, where it is found at frequencies ranging from ~0.30-0.70 (Rosser *et al.* 2000; Semino *et al.* 2000), as well as eastern European populations such as Lithuanians (0.47; Rosser *et al.* 2000). Many other rarer hgs might be observed in large enough samples of British and European populations, representing sporadic gene flow from Asia, the Americas and Africa.

In summary PR and hg2 lineages describe most of the Y-chromosome diversity seen in Britain as well as much of western Europe, with some rarer hgs present

at much lower frequencies. However, the picture of Y-chromosome diversity in Britain is skewed towards coverage of Scottish populations and is somewhat patchy as it is composed of data from several studies. The data that do exist, however, point to some very intriguing differences in the frequencies of the 3 main hgs (Table 2.1), possibly reflecting disparate histories of contact with other European populations, which could be addressed from a Y-chromosome perspective using a more rigorous sampling strategy. Indeed, given that frequency differences can be found over small geographical distances within the same country is intriguing and supports the finding that Y-chromosome diversity in Europe is primarily a function of geography (Rosser *et al.* 2000; Zerjal *et al.* 2001), although such conclusions are typically made for populations spread over much larger geographical areas.

2.1.6. Aims of this Chapter

The above studies thus provide an incomplete picture of the Y-chromosome diversity in the British Isles. Therefore the aim of the work described in this Chapter was to systematically sample from populations across the British Isles to comprehensively study their Y-chromosome diversity. The UEP markers employed were selected to assay hgs known to be present in the relevant populations, based on the work described above.

2.2. Materials and Methods

2.2.1. Sample Collection

Male DNA samples were systematically collected from small urban locations in Britain. Locations were primarily chosen by placing a 3x5 grid over the British Isles and selecting small towns (defined as a population size of 5-20,000 individuals) near to the intersection of grid points (Figure 2.5). If the gird point

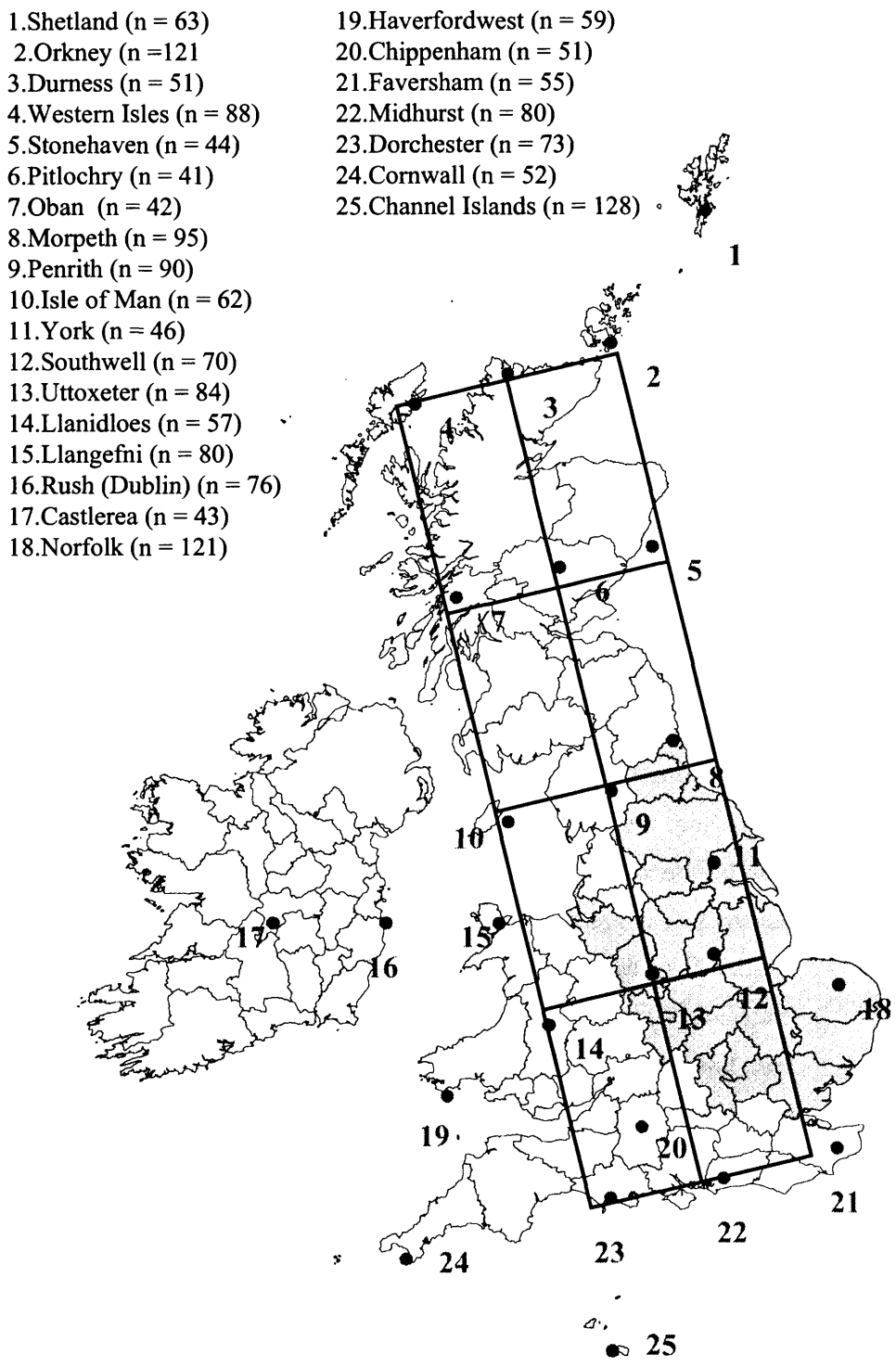


Figure 2.5. British Isles Sampling Locations and Sample Sizes, Indicating the Danelaw. Small urban locations were primarily selected using the 3x5 grid with additional sites chosen to maximise coverage of Britain. *Figure modified from Capelli et al. (2003).* The Danelaw is also indicated by the dark grey shading, and regions outside the Danelaw are in light grey (modified from Davies (1999)

fell in the sea the nearest location on the coast was used (Morpeth and Stonehaven). Additional sampling locations were included to cover areas not found near to the intersection of grid points (Shetland, York, Norfolk, Haverfordwest, Llangefni, Chippenham, Cornwall, Castlerea and Rush and Channel Islands, see below). Individuals that were sampled had to be able to trace their paternal grandfather's birthplace to within a 20 mile radius of the sampling location (except Midhurst where the radius was 40 miles). Volunteers gave appropriate informed consent and were over the age of 18. Including the Channel Islands, a total of 1,772 Y-chromosomes were sampled from 25 British locations. For comparative purposes samples were collected from the towns of Bergen and Trondheim (west Norway), Copenhagen (Denmark) and Schleswig-Holstein (North Germany) to represent Norwegian Vikings, Danish Vikings, and Anglo-Saxons. Bergen and Trondheim were both identified as probable origins of Norwegian Viking raids (Graham-Campbell, personal communication), and today these towns are relatively small, (237,430 and 154, 351 inhabitants respectively on January 1st 2004, source: Statistics Norway [http://www.ssb.no; 30th March 2004]) hence assumed to be unaffected by recent population movements. Schleswig-Holstein is a region in North Germany identified as a likely source for Angles and Saxons (Graham-Campbell, personal communication). It was not feasible to sample from small towns in Denmark, hence samples had to be collected from Copenhagen. Previously collected Basque samples (Bosch et 1999; Bosch *et al.* 2001) were also analysed and used as representatives of the indigenous (or so-called Palaeolithic) Y-chromosome gene pool of Europe (Bosch *et al.* 2001; Wilson *et al.* 2001a; but see Hurles *et al.* 1999). In total 433 European Y-chromosomes were included.

DNA samples were collected from 172 male inhabitants from the larger two Channel Islands, Jersey (n=118) and Guernsey (n=54). Samples were not collected from Alderney and Sark. For Alderney the concern was that recent influxes of people from the British mainland (F Falle, personal communication) would obscure past patterns of population history and Sark was recently settled (1565) by a group from Jersey and currently has a very small population size (~600) (http://user.itl.net/~glen/sark.html; 30th March 2004) hence sampling from Sark is not expected to reveal any important details about the history of the

Channel Islands not already captured in the other samples. Participants in the study gave appropriate informed consent, had to be over 18 years of age as above, and an inhabitant of Jersey or Guernsey who could also trace their paternal grandfather's birth place to the same island as the donor. Volunteers took the sample themselves using a mouth swab under the instruction of Mr F Falle or Mr W Galliene, (local historians on Jersey and Guernsey respectively), who were trained by JK Abernethy. Mouth swabs were stored in 2ml tubes containing 1ml of 0.05M EDTA/0.05M SDS preservative solution and transported from Jersey and Guernsey to the laboratory where they were stored at 4°C until they were extracted.

2.2.2. DNA Extraction

DNA was extracted using a standard phenol chloroform procedure (Table 2.2) yielding approximately 5ng/μl of DNA.

2.2.3. Y-Chromosome Genotyping

Y-chromosome haplotypes were defined using 6 Y-linked microsatellites (DYS388, DYS393, DYS392, DYS19, DYS390, DYS391) here termed YSTR1, in a PCR multiplex designed by Thomas *et al.* (1999) (see also Table 2.3 for details). All PCR reactions were performed in ABgene® Thermo-Fast® Low Profile 96-well plates. DYS393, 392 and 390 failed to amplify on several samples, including some from Jersey and Guernsey, using the initial conditions, but were successfully amplified using increased concentrations of these primers (Table 2.3). When microsatellite haplotypes are listed they are given as the number of repeats and listed in the following order DYS388-393-392-19-390-391. PCR products were electrophoresed on an ABI PRISM ® 377 DNA Sequencer using the conditions in Table 2.3 and allele sizes determined using ABI PRISM® GeneScan® v3.1. Expected allele sizes are shown in Table 2.3 as

Table 2. 2 DNA Extraction Using Phenol Chloroform[1]

1. Add 0.8ml dH_2O to the mouth swab tube (containing the swab preserved in 1ml of EDTA/SDS solution).
2. Incubate at 60°C for 10 mins.
3. Aliquot 0.6ml phenol/chloroform (1:1) mix to a 2.0ml tube and transfer 0.8ml of the solution from step 1.
4. Mix and centrifuge for 10 mins at maximum speed in a microfuge at room temperature.
5. Aliquot 0.6ml chloroform and 30µl 5M NaCl into a 2ml tube.
6. Transfer the aqueous phase from step 4 into the tube containing the chloroform/ NaCl mix.
7. Mix and centrifuge for 10 mins at maximum speed in a microfuge at room temperature.
8. Aliquot 0.7ml chloroform into a 2ml tube.
9. Transfer the aqueous phase from step 7 into the 2ml tube containing the chloroform.
10. Mix and centrifuge for 10 mins at maximum speed in a microfuge at room temperature.
11. Aliquot 0.7ml of 100% isopropanol into a 2ml tube.
12. Transfer the aqueous phase from step 10 into the 2ml tube containing the isopropanol.
13. Mix and place at -20°C for at least 2 hours.
14. Centrifuge for 12 mins at maximum speed in a microfuge at room temperature.
15. Pour off the supernatant from the sample tube, invert tube at ~45° angle for 1 min to air dry.
16. Add 0.8ml of 70% ethanol to the sample tube.
17. Mix and centrifuge for 10 mins at maximum speed in a microfuge at room temperature.
18. Pour off the supernatant from the sample tube, invert tube at ~45° angle for 20 mins to air dry.
19. Add 200µl TE (1M Tris, 0.5M EDTA, pH 9) mix and incubate at 56°C for 10 mins.
20. Centrifuge briefly and store at -20°C.

[1]. Taken from a protocol written by Mark Thomas.

Table 2.3. YSTR1 PCR Multiplex and Electrophoresis Conditions and Microsatellite Repeat Sizes

(a) PCR Protocol

Primer Mix | **Optimised Primer Mix**

Primer Name	*Fluorescent label (5')*	*Final conc. in PCR (μM)*	*Volume per reaction (μl)*	*Final conc. in PCR (μM)*	*Volume per reaction (μl)*
DYS19L	Tet	0.300	0.060	0.600	0.120
DYS19R	-	0.300	0.060	0.600	0.120
DYS388L	Tet	0.320	0.064	0.640	0.128
DYS388R	-	0.320	0.064	0.640	0.128
DYS390L	-	0.130	0.026	-	-
DYS390R	Fam	0.130	0.026	-	-
DYS391L	Fam	0.380	0.076	-	-
DYS391R	-	0.380	0.076	-	-
DYS392L	-	0.160	0.032	0.320	0.064
DYS392R	Hex	0.160	0.032	0.320	0.064
DYS393L	-	0.090	0.018	-	-
DYS393R	Hex	0.180	0.036	-	-
dH_20	-	-	0.430	-	0.270
Total			1.000		1.000

PCR mix

	Volume per reaction (μl)
Primer mix	1.000
10X Buffer	1.000
0.1 M $MgCl_2$	0.070
10mM dNTP	0.200
dH_20	6.690
Taq[a]	0.040
DNA	1 (~5ng)
Total	10.000

[a] 2HTTaq:1TaqStart™

Cycling Conditions

Temperature (degrees C)	*Duration*[b]	*Cycles*
95	4'	
95	40"	38
57	40"	
72	40"	
72	10'	
4	∞	

[b] Minutes ', seconds "

(b) Electrophoresis Conditions - ABI PRISM ® 377 DNA Sequencer

Dilution (Digestion product:dH_2O)	*Time*	*Filter*	*Acrylamide*	*Standard*
1:4	2.0 hours	C	4.25%	TAMRA™ 350[c]

[c] Consiting of TAMRA™ 350, Dextran blue and de-ionised formamide in the ratio 1:1:9

(c) Expected Allele Sizes (ABI PRISM ® 377 DNA Sequencer) and Microsatellite Conversion

Locus	*Repeat Size*	*Size Range (bp)*	*Conversion from bp to repeat size*[d]
DYS19	Tetranucleotide	182-202	(x-136)/4
DYS388	Trinucleotide	119-145	(x-93)/3
DYS390	Tetranucleotide	192-220	(x-120)/4
DYS391	Tetranucleotide	148-173	(x-124)/4
DYS392	Trinucleotide	148-173	(x-133)/3
DYS393	Tetranucleotide	106-130	(x-71)/3

[d] Where x is the allele size in base pairs

Note: For a full list of suppliers of reagents and equipment see Appendix, Table A.1

is the equation used to convert the size in base pairs to the number of microsatellite repeats. Primer sequences can be found in the Appendix (Table A.2)

Six UEPs (M170, M172, M9, 92R7, M173 and M17) chosen from the literature to include the most frequent Y-chromosome polymorphisms seen in Europe (designed by C Capelli) were combined into a PCR and RFLP multiplex kit (Table 2.4), and typed on all samples to define hgs. All chromosomes found to be M170 derived were typed for M26, a node internal to M170 using a PCR-RFLP approach (Table 2.5). Y-chromosomes that were either derived only at M9 or underived at all EURO loci were further typed using a PCR and RFLP approach for 4 UEPs that define hgs also found in European populations: Tat, M89 and 12f2 (in a multiplex kit, Table 2.6) and M35 as a singleplex (Table 2.7), which subsequently captured the diversity in the entire dataset. Typing M9 derived chromosomes for M35, M89, and 12f2 is evolutionarily redundant because an M9 derived chromosome must be M89 derived and cannot be M35 or 12f2 derived (see Figure 2.6). However this strategy provides a control against the mis-typing of samples either through PCR-RFLP related problems, such as enzymatic failure, or mis-aliquoting of DNA. All of the above UEPs were electrophoresed and allele sizes determined as above, using the conditions in the relevant table. Expected allele sizes are also given. The genealogical relationship and YCC nomenclature of the above UEPs is shown in Figure 2.6. Primer sequences can be found in the Appendix (Table A.2). An example of the GeneScan output for the EURO1 kits can be found in the Appendix (Figure A.4).

2.2.4. Data Analysis

For the common European hgs, R1*(xR1a1), I*(xI1b2), and R1a, previously defined modal haplotypes (Wilson *et al.* 2001a) were identified in the present dataset and the nomenclature of Wilson *et al.* (2001a) was retained for continuity (AMH+1, 2.47+1, and 3.65+1 respectively). The "+1" nomenclature

Table 2.4. EURO1 PCR/RFLP Multiplex and Electrophoresis Conditions and Expected Allele Sizes

(a) PCR Protocol

Primer Mix | **Optimised Primer Mix**

Primer Name	*Fluorescent label*	*Final conc. in PCR (μM)*	*Volume per reaction (μl)*	*Final conc. in PCR (μM)*	*Volume per reaction (μl)*
M9long F	-	0.150	0.030	-	-
M9long R	Tet	0.150	0.030	-	-
92R7 U	Hex	0.150	0.030	0.225	0.045
92R7 R	-	0.150	0.030	0.225	0.045
M17 F	-	0.150	0.030	0.225	0.045
M17R	Tet	0.150	0.030	0.225	0.045
M173II F	-	0.400	0.080	0.600	0.120
M173 R	Fam	0.400	0.080	0.600	0.120
M170F	-	0.500	0.100	0.750	0.150
M170R	Hex	0.500	0.100	0.750	0.150
M172F	Tet	0.200	0.040	0.300	0.060
M172R	-	0.200	0.040	0.300	0.060
dH_20	-	-	0.380	-	1.000
Total			1.000		1.000

PCR mix

Component	*Volume per reaction (μl)*
Primer mix	1.00
10X Buffer	1.00
10mM dNTP	0.20
0.1 M $MgCl_2$	0.14
dH_20	6.62
Taq[a]	0.04
DNA	1 (~5ng)
Total	10

[a] 2HTTaq:1TaqStart™

Cycling Conditions

Temperature (degrees C)	*Duration*[b]	*Cycles*
95	4'	
95	40"	38
55	40"	
72	40"	
72	10'	
4	∞	

[b] Minutes ', seconds "

(b) RFLP Protocol

Enzyme mix

Enzyme	*UEP Restriction Site*	*Volume per reaction (μl)*
HinFI	M9/M172	0.040
Hind III	92r7	0.040
AflIII	M17	0.020
Bcl I	M170	0.040
Bsr G I	M173	0.060
dH_20	-	0.213
Total		0.413

Digestion mix

Component	*Volume per reaction (μl)*
Enzyme mix	0.413
NEB Buffer 2	0.800
10 mg/ml BSA	0.080
dH_20	4.707
PCR Product	2.000
Total	8.000

Incubation

Temperature (degrees C)	*Duration*
37	Overnight
50	2 Hours

continued

Table 2.4 continued

(c) Electrophoresis Conditions - ABI PRISM ® 377 DNA Sequencer

Dilution (Digestion product:dH_2O)	*Time*	*Filter*	*Acrylamide*	*Standard*
1:4	1.5 hours	C	4.25%	TAMRA™ 350[c]

[c] Consiting of TAMRA™ 350, Dextran blue and de-ionised formamide in the ratio 1:1:9

(d) Expected Allele Sizes (ABI PRISM ® 377 DNA Sequencer)

Locus	*Ancestral Allele*	*Derived Allele*
M9	67-G	97-C
92r7	66-C	95-T
M17	123-G	104-G
M173	99-A	118-C
M170	83-A	111-C
M172	172-T	143-A

Table 2.5. M26 PCR/RFLP Singleplex and Electrophoresis Conditions and Expected Allele Size

(a) PCR Protocol

Primer Mix

Primer Name	*Fluorescent label (5')*	*Final conc. in PCR (μM)*	*Volume per reaction (μl)*
M26 F	-	0.500	0.1000
M26 R	Hex	0.500	0.1000
dH_20	-	-	0.8000
Total			1.0000

PCR mix

	Volume per reaction (μl)
Primer Mix	1.000
10X Buffer	1.000
10mM dNTP	0.200
0.1 M $MgCl_2$	0.100
dH_20	6.660
Taq[a]	0.040
DNA	1 (~5ng)
Total	10.000

[a] 2HTTaq:1TaqStart™

Cycling Conditions

Temperature (degrees C)	*Duration*[b]	*Cycles*
95	4'	
95	40"	38
55	40"	
72	40"	
72	10'	
4	∞	

[b] Minutes ', seconds "

(b) RFLP Protocol

Enzyme mix

Enzyme	*UEP Restriction Site*	*Volume per reaction (μl)*
Bcl I	M26	0.04
dH_20	-	0.56
Total		0.6

Digestion mix

	Volume per reaction (μl)
Enzyme mix	0.600
NEB buffer 2	0.800
10 mg/ml BSA	0.080
dH_20	4.520
PCR Product	2.000
Total	8.000

Incubation

Temperature (degrees C)	*Duration*
37	Overnight
50	2 Hours

(c) Electrophoresis Conditions - ABI PRISM ® 377 DNA Sequencer

Dilution (Digestion product: dH_20	*Time*	*Filter*	*Acrylamide*	*Standard*
1:4	4 hours	C	4.25%	TAMRA™ 350[c]

[c] Consiting of TAMRA™ 350, Dextran blue and de-ionised formamide in the ratio 1:1:9

(d) Expected Allele Sizes (ABI PRISM 377 ® DNA Sequencer)

Locus	*Ancestral Allele*	*Derived Allele*
M26	169-G	149-A

Table 2.6. M89/Tat/p12f2 PCR/RFLP Multiplex and Electrophoresis Conditions and Expected Allele Sizes

(a) PCR Protocol

Primer Mix

Primer Name	*Fluorescent label (5')*	*Final conc. in PCR (μM)*	*Volume per reaction (μl)*
M89 F	Hex	0.350	0.0700
M89 R	-	0.350	0.0700
TAT F	-	0.350	0.0700
TAT R	Fam	0.350	0.0700
p12f2D	Hex	0.350	0.0700
p12f2G	-	0.350	0.0700
dH_20	-	-	0.5800
Total			1.0000

PCR mix

Component	*Volume per reaction (μl)*
Primer Mix	1.000
10X Buffer	1.000
10mM dNTP	0.200
dH_20	6.760
Taq[a]	0.040
DNA	1 (~5ng)
Total	10.000

[a] 2HTTaq:1TaqStart™

Cycling Conditions

Temperature (degrees C)	*Duration*[b]	*Cycles*
95	4'	
95	40"	38
59	40"	
72	40"	
72	10'	
4	∞	

[b] Minutes ', seconds "

(b) RFLP Protocol

Enzyme Mix

Enzyme	*UEP Restriction Site*[c]	*Volume per reaction (μl)*
Nla III	M89/TAT	0.3
dH_20	-	0.113
Total		0.413

[c] 12f2 is not assayed by an enzyme, instead by presence (ancestral) or absence (derived) of an 88bp insertion

Digestion mix

Component	*Volume per reaction (μl)*
Enzyme mix	0.413
NEB buffer 4	0.800
10 mg/ml BSA	0.080
dH_20	4.708
PCR Product	2.000
Total	8.000

Incubation

Temperature (degrees C)	*Duration*
37	Overnight
50	2 Hours

continued

Table 2.6 continued

(c) Electrophoresis Conditions - ABI PRISM ® 377 DNA Sequencer

Dilution (Digestion product: dH_20	*Time*	*Filter*	*Acrylamide*	*Standard*
1:4	1.5 hours	C	4.25%	TAMRA™ 350[d]

[d] Consiting of TAMRA™ 350, Dextran blue and de-ionised formamide in the ratio 1:1:9

(d) Expected Allele Sizes (ABI PRISM ® 377 DNA Sequencer)

Locus	*Ancestral Allele*	*Derived Allele*
M89	80-C	98-T
Tat	83-A	112-C
p12f2	88	no band

Table 2.7. M35 PCR/RFLP Singleplex and Electrophoresis Conditions and Expected Allele Size

(a) PCR Protocol

Primer Mix

Primer Name	*Fluorescent label (5')*	*Final conc. in PCR (μM)*	*Volume per reaction (μl)*
M35 F	Fam	0.350	0.070
M35 R	-	0.350	0.070
dH_20	-	-	0.860
Total			1.000

PCR mix

Component	*Volume per reaction (μl)*
Primer mix	1.000
10X Buffer	1.000
10mM dNTP	0.200
0.1 M $MgCl_2$	0.140
dH_20	6.620
Taq[a]	0.040
DNA	1 (~5ng)
Total	10.000

[a] 2HTTaq:1TaqStart™

Cycling Conditions

Temperature (degrees C)	*Duration*[b]	*Cycles*
95	4'	
95	40"	38
58	40"	
72	40"	
72	10'	
4	∞	

[b] Minutes ', seconds "

(b) RFLP Protocol

Enzyme mix

Enzyme	*UEP Restriction Site*	*Volume per reaction (ml)*
Bsr I	M35	0.02
dH_20	-	0.393
Total		0.413

Digestion mix

	Volume per reaction (μl)
Enzyme mix	0.413
NEB Buff. 3	0.800
10 mg/ml BSA	0.000
dH_20	4.788
PCR Product	2.000
Total	8.000

Incubation

Temperature (degrees C)	*Duration*
65	3 Hours

(c) Electrophoresis Conditions - ABI PRISM ® 377 DNA Sequencer

Dilution (Digestion product:		*Time*	*Filter*	*Acrylamide*	*Standard*
1:4	dH_20	1.5 hours	C	4.25%	TAMRA™ 350[c]

[c] Consiting of TAMRA™ 350, Dextran blue and de-ionised formamide in the ratio 1:1:9

(d) Expected Allele Sizes (ABI PRISM ® 377 DNA Sequencer)

Locus	*Ancestral Allele*	*Derived Allele*
M35	130-G	160-C

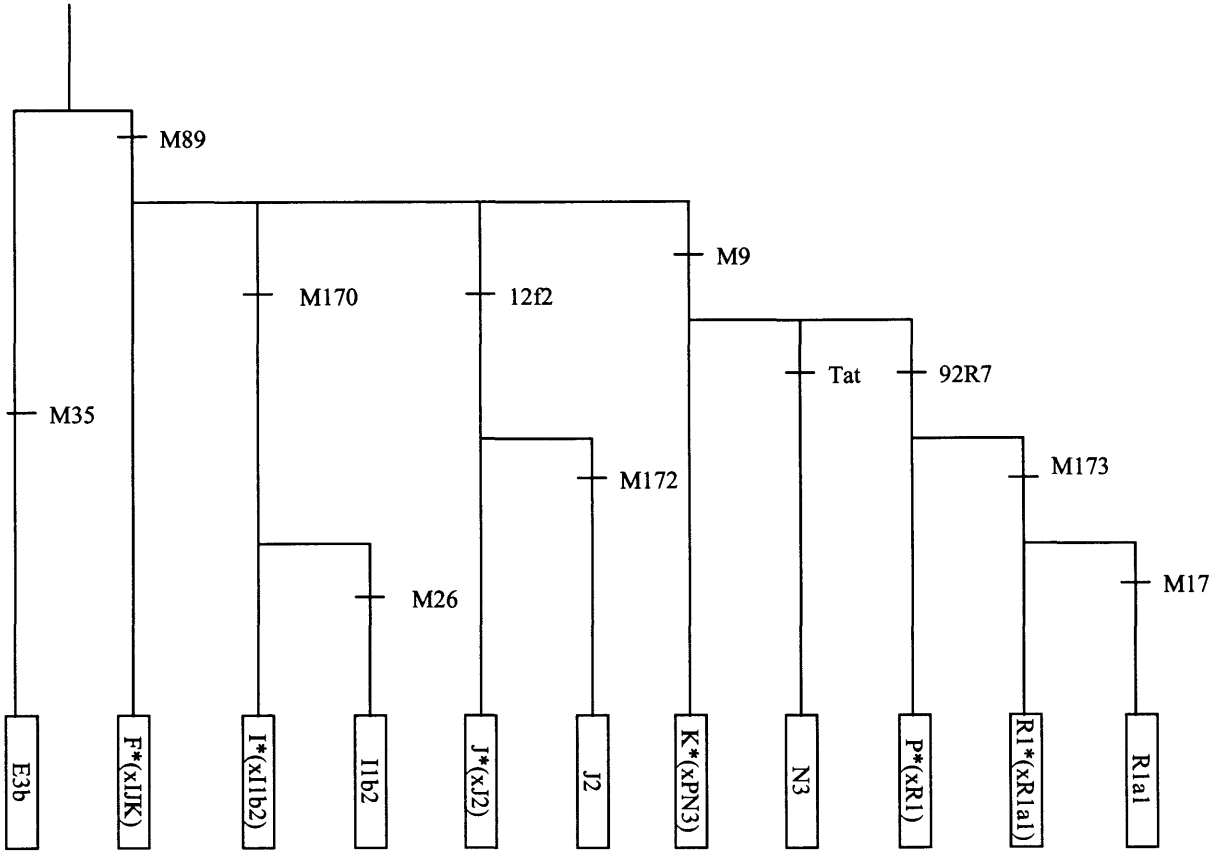


Figure 2.6. Y-Chromosome Genealogy. Genealogical relationship of the UEPs used and the hgs they define. Hg nomenclature is that suggested by the YCC (2002). *Figure modified from Capelli et al. (2003).*

used in this and other contexts indicates that the modal haplotype as well as its one step mutational neighbours have been included as a cluster due to instability in repeat number at microsatellite loci (Thomas *et al.* 1998). Most analyses (see below) were performed using the modal clusters AMH+1, 2.47+1, and 3.65+1. Whilst these microsatellite-defined groups do not strictly reflect genealogical relationships because of homoplasy, they provide further resolution for the purposes of analysis given that these three hgs encompass most British Y-chromosome types.

Exact tests of population differentiation and analysis of molecular variance were carried out in Arlequin 1.1 (Schneider *et al.* 1997) to assess levels of genetic structure, or differentiation, between populations. The presence of differentiation means that there is a non random distribution of haplotypes in the designated populations (although this test assumes panmixia) (Schneider *et al.* 2000). The exact test tests the hypothesis of a random distribution of k different haplotypes in r populations and is analogous to Fisher's exact test (for a 2x2 contingency table) but extended to a rxk contingency table. A Markov chain explores all potential states of the contingency table and the probability of observing a table less or equally likely than the one observed is calculated under the assumption of random mating (Schneider *et al.* 1997). The Markov chain was run for 10,000 steps.

Principal Components (PC) analysis was performed on all populations on the basis of hg+1 frequencies using POPSTR (H Harpending, personal communication) to summarise the variation and infer population affiliations particularly with respect to the relative inputs of the potential source populations. To aid interpretation of the PC plots, simulated admixed British populations were created by drawing varying proportions of individuals from each of the source populations at random and with replacement. Simulated populations were created with 60%, 40%, and 20% input from Norway and conversely 40%, 60% and 80% "indigenous" (Basque and Castlerea combined, see Results, section 2.3.1, and Discussion, 2.4.1) input, respectively. This was repeated for North Germany and Denmark combined (see Results and Discussion) and indigenous. The procedure was repeated 6 times for each of the

60%, 40% and 20% Norway and North German/Danish simulated populations to indicate the range of resultant values. Circles were drawn around the clusters of 60%, 40% and 20% simulated populations to illustrate the range of locations where the (simulated) populations fell and overlap between the populations.

Admixture proportions of the relative inputs of indigenous, Norwegian and North German/Danish Y-chromosomes in the British populations were inferred using a likelihood based approach, LEA (Chikhi *et al.* 2001). The method follows a simple admixture model (Figure 2.7) where two independent parental populations P_1 and P_2 have contributed the proportions p_1 and p_2 to a third hybrid population. From the moment of admixture three populations evolve independently for T generations by drift. LEA calculates the proportion of input for p_1 and p_2 ($1-p_1$) and the time since admixture (t_1, t_2, t_h) scaled by population size, which is used to infer drift since admixture in each population. The method considers sampling variation, drift and uncertainty over the parental allelic frequencies. However, neither mutation or gene flow since admixture are considered. The former of these is unlikely to be problematic because UEP information is used in LEA to calculate admixture proportions; UEPs are believed to have mutated only once in human evolution (Jobling and Tyler-Smith 1995; Thomas *et al.* 1998) barring at least one known reversion (SRY_{10831}, Hammer *et al.* 1998). The effect of the latter is difficult to quantify and has to be considered as an unknown factor that may skew the results. As the Y-chromosome is a single locus, point estimates are anticipated to be unrepresentative of the distribution due to the wide credible intervals associated with single loci (Chikhi *et al.* 2001; Chikhi *et al.* 2002). Hence 95% credible intervals for the proportion of p_1, as well as the median value, are given. Norway and North Germany/Denmark were alternately used to represent p_1, and Castlerea and Basque were combined to represent the indigenous European population, p_2. Simulations were run for 50,000 iterations (100,000 for the Channel Islands). The posterior probability density functions (pdf's) for p_1 and t_1, t_2, and t_h were obtained for the Channel Islands and plotted using the locfit package for R, having removed the first 10% of the runs, the so called "burn in".

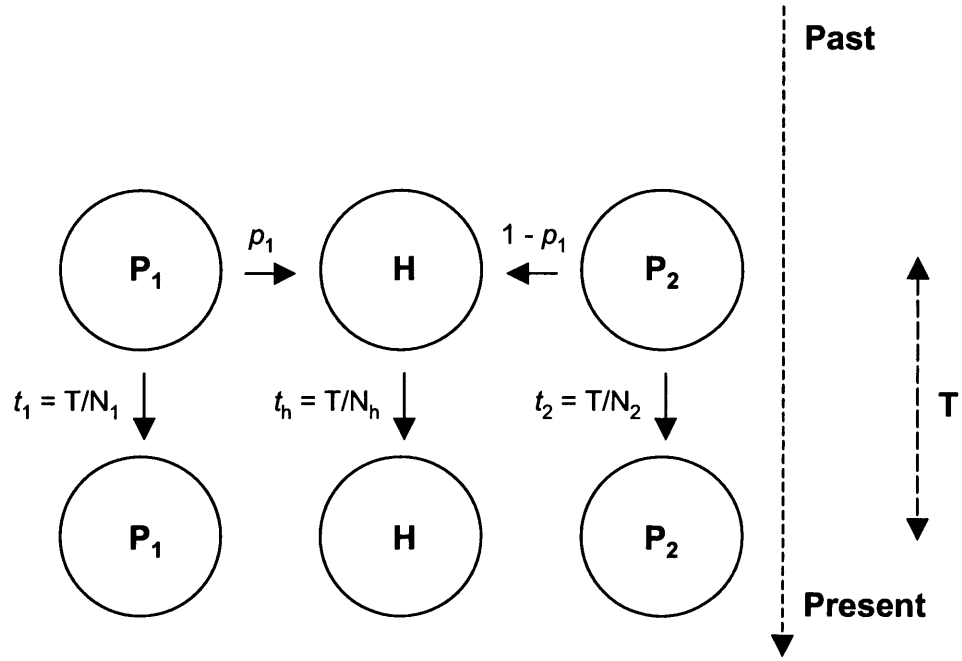


Figure 2.7. The LEA Admixture Model. A single admixture event occurred T generations ago by two parental populations (P_1 and P_2) coming together to form the hybrid population, H. After the admixture event there is no gene flow between the three populations. *Figure adapted from Chikhi et al. (2001).*

To allow a series of questions about the Channel Islands, Jersey and Guernsey to be answered the samples were grouped using 3 criteria: Jersey and Guernsey combined; Jersey and Guernsey separately; and Jersey and Guernsey each split into two groups on the basis of surname, those with Norman surnames and those without Norman surnames. Information about the probable origin of surnames was provided by a local historian from Jersey (F Falle, personal communication). The above analyses were performed on each of these different groups.

2.3. Results

2.3.1. The British Isles and European Populations

Bergen and Trondheim were not significantly different to each other based on hg+1 frequencies, nor were the Danish and North German samples (p=0.08), these populations were thus combined into "Norway" and "North Germany/Denmark" for further analyses. Both of these groups, Norway and North Germany/Denmark are significantly different from the majority of the British populations (Table 2.8, using hg+1 frequencies), indicating significant structure between these continental populations and most of Britain, indeed the only exception is between North Germany/Denmark and York (p = 0.615), There is less structure between Basques and the British Isles, although Basques are still significantly different (hg+1 frequencies) to northern Scottish populations and several northern English and central/eastern English populations (Penrith, Isle of Man, York, Southwell, Norfolk) and one south coast location (Chippenham). The non-significant difference between Basques and Castlerea (p = 0.552) was exploited by combining these two populations to form the "indigenous" source population. The northern most Scottish populations are also highly differentiated from the rest of Britain, but comparisons between English populations reveals minimal differentiation. A summary of hg frequencies can be found in Table 2.9.

Table 2.8: Exact Test of Population Based on Hg+1 Frequencies for the Populations Studied

(a)

Population	*Nrw*	*GD*	*Bas*	*Shet*	*Ork*	*Dur*	*Wls*	*Sth*	*Ptl*	*Oban*	*Mpt*	*Pnt*	*IoM*	*York*	*Sow*	*Utx*	*Ldl*	*Lgf*	*Rsh*	*Cas*	*Nor*	*Hfw*	*Chp*	*Fav*	*Mdh*	*Dcr*	*Cor*	*AllChl*
Nrw	-																											
NG/D	**0.000**	-																										
Bas	**0.000**	**0.000**	-																									
Shet	**0.000**	**0.000**	**0.000**	-																								
Ork	**0.000**	**0.000**	**0.001**	0.829	-																							
Dur	**0.000**	**0.000**	**0.000**	**0.001**	**0.032**	-																						
Wls	**0.000**	**0.002**	**0.000**	0.087	0.182	**0.001**	-																					
Sth	**0.000**	**0.005**	0.077	**0.018**	**0.030**	0.462	**0.016**	-																				
Ptl	**0.000**	**0.000**	0.148	**0.015**	**0.001**	**0.000**	0.057	**0.043**	-																			
Oban	**0.000**	**0.000**	0.849	0.198	0.201	0.095	0.133	0.521	0.446	-																		
Mpt	**0.000**	**0.000**	0.081	0.011	**0.003**	**0.016**	0.260	0.429	0.652	0.804	-																	
Pnt	**0.000**	**0.000**	**0.031**	0.182	**0.041**	**0.025**	0.213	0.199	0.185	0.541	0.392	-																
IoM	**0.000**	**0.011**	**0.018**	0.434	0.311	**0.013**	0.479	0.083	0.064	0.377	0.221	0.812	-															
York	**0.000**	0.614	**0.000**	**0.004**	**0.021**	**0.038**	0.153	0.125	**0.007**	0.052	0.188	0.574	0.214	-														
Sow	**0.000**	**0.003**	0.067	**0.004**	**0.001**	**0.003**	**0.048**	0.124	0.549	0.381	0.495	0.415	0.266	0.303	-													
Utt	**0.000**	**0.001**	0.069	**0.001**	**0.001**	0.115	**0.017**	0.486	0.378	0.358	0.506	0.676	0.078	0.461	0.532	-												
Ldl	**0.000**	**0.024**	0.086	**0.005**	**0.001**	**0.020**	**0.035**	0.698	**0.049**	0.351	0.487	0.739	0.364	0.789	0.624	0.632	-											
Lgf	**0.000**	**0.000**	0.381	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	0.281	0.372	**0.004**	**0.003**	**0.007**	**0.000**	**0.010**	**0.005**	**0.001**	-										
Rush	**0.000**	**0.000**	0.087	**0.000**	**0.003**	**0.061**	**0.000**	**0.014**	0.061	0.120	**0.004**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	-									
Cas	**0.000**	**0.000**	0.568	**0.001**	**0.016**	**0.036**	**0.003**	0.077	0.221	0.619	0.160	**0.028**	**0.009**	**0.005**	0.077	0.171	**0.038**	0.199	0.786	-								
Nor	**0.000**	**0.014**	**0.002**	**0.002**	**0.000**	**0.035**	0.135	0.154	0.038	0.106	0.317	0.426	0.068	0.998	0.294	0.699	0.728	**0.000**	**0.000**	**0.010**	-							
Hfw	**0.000**	**0.000**	0.855	**0.001**	**0.000**	**0.000**	**0.000**	**0.017**	0.057	0.614	**0.008**	**0.018**	**0.005**	**0.000**	**0.010**	**0.005**	**0.009**	0.936	**0.042**	0.402	**0.000**	-						
Chp	**0.000**	**0.024**	**0.020**	0.074	**0.049**	**0.008**	0.058	0.713	0.081	0.406	0.529	0.723	0.603	0.432	0.322	0.252	0.797	**0.000**	**0.000**	**0.009**	0.215	**0.002**	-					
Fav	**0.000**	**0.002**	0.280	**0.031**	**0.008**	0.072	**0.014**	0.747	0.528	0.661	0.482	0.706	0.199	0.142	0.668	0.918	0.669	0.109	**0.003**	0.167	0.169	0.138	0.574	-				
Mdh	**0.000**	**0.000**	0.214	**0.000**	**0.001**	**0.000**	**0.033**	0.127	0.745	0.483	0.873	0.232	0.106	0.061	0.528	0.445	0.244	**0.047**	**0.002**	0.359	0.072	0.056	0.362	0.531	-			
Dcr	**0.000**	**0.000**	0.242	**0.004**	**0.001**	0.064	**0.019**	0.785	0.348	0.690	0.713	0.539	0.428	0.340	0.881	0.721	0.953	0.073	**0.008**	0.283	0.276	0.132	0.605	0.963	0.785	-		
Cor	**0.000**	**0.002**	0.104	0.138	0.217	0.168	0.092	0.929	0.246	0.822	0.712	0.596	0.687	0.098	0.282	0.433	0.535	0.065	**0.012**	0.146	0.141	0.114	0.856	0.853	0.450	0.878	-	
AllChl	**0.000**	**0.000**	0.190	**0.000**	**0.000**	0.170	**0.013**	0.586	**0.025**	0.361	0.215	0.156	0.104	0.813	0.234	0.651	0.823	**0.007**	**0.009**	0.312	0.616	**0.011**	0.337	0.396	0.322	0.675	0.310	-

continued

Table 2.8. Continued

(b)

Population	Nrw	NG/D	Bas	Shet	Ork	Dur	WIs	Sth	Ptl	Oban	Mpt	Pnt	IoM	York	Sow	Utx	Ldl	Lgf	Rsh	Cas	Nor	Hfw	Chp	Fav	Mdh	Dcr	Cor	AllChl
AllChl	**0.000**	**0.000**	0.190	**0.000**	**0.000**	0.170	**0.013**	0.586	**0.025**	0.361	0.215	0.156	0.104	0.813	0.234	0.651	0.823	**0.007**	**0.009**	0.312	0.616	**0.011**	0.337	0.396	0.322	0.675	0.310	-
Jer	**0.000**	**0.002**	**0.002**	**0.000**	**0.000**	0.057	**0.011**	0.203	**0.003**	**0.016**	**0.034**	**0.026**	**0.009**	0.866	**0.022**	0.141	0.382	**0.000**	**0.001**	**0.007**	0.479	**0.001**	0.173	**0.046**	0.100	0.147	**0.040**	-
Gue	**0.000**	**0.000**	0.902	**0.035**	**0.025**	0.092	**0.016**	0.487	0.491	0.876	0.181	0.342	0.145	0.095	0.772	0.584	0.507	0.197	0.078	0.503	0.127	0.626	0.289	0.904	0.397	0.858	0.568	-
JerN	**0.001**	0.055	0.151	**0.028**	0.092	0.148	**0.029**	0.204	0.070	0.137	0.182	0.121	0.086	0.375	0.244	0.224	0.265	**0.011**	0.152	0.078	0.220	0.037	0.420	0.242	0.300	0.240	0.179	-
GueN	**0.007**	0.212	0.124	0.061	0.147	0.184	0.049	0.327	0.061	0.158	0.173	0.328	0.254	0.598	0.565	0.269	0.594	**0.033**	0.077	0.053	0.336	0.046	0.592	0.423	0.254	0.591	0.356	-
JerO	**0.000**	**0.037**	**0.002**	**0.000**	**0.000**	0.068	**0.005**	0.199	**0.001**	**0.014**	**0.021**	**0.038**	**0.011**	0.928	0.057	0.101	0.347	**0.000**	**0.000**	**0.001**	0.617	**0.000**	0.114	**0.015**	**0.045**	0.121	**0.038**	-
GueO	**0.000**	**0.002**	0.781	0.094	0.065	**0.028**	**0.019**	0.253	0.261	0.870	0.126	0.340	0.225	0.079	0.515	0.322	0.450	0.392	0.059	0.213	0.089	0.463	0.134	0.630	0.195	0.690	0.424	-

Population	Jer	Gue	JerN	GueN	JerO	GueO
AllChl	-	-	-	-	-	-
Jer	-		-	-	-	-
Gue	0.066	-	-	-	-	-
JerN	-	-	-	-	-	-
GueN	-	-	1.000	-	-	-
JerO	-	-	0.501	0.624	-	-
GueO	-	-	0.175	0.488	**0.021**	-

Table 2.8 continued. (a) British and European populations, (b) the Channel Islands in more detail (see text). Bold text indicates significant comparisons, $p < 0.05$. Populations as follows: Nrw = Norway, NG/D = North Germany/Denmark, Shet = Shetland Isles, , Ork = Orkney Isles, Dur = Durness, WIs = Western Isles, Sth = Stonehaven, Ptl = Pitlochry, Oban = Oban, Mpt = Morpeth, IoM = Isle of Man, York = York, Sow = Southwell, Utx = Uttoxeter, Nor = Norfolk, Hfw = Haverfordwest, Chp = Chippenham, Fav = Faversham, Mdh = Midhurst, Dcr = Dorchester, Cor = Cornwall, AllChI = Jersey and Guernsey combined, Jer = Jersey, Gue = Guernsey, JerN = Jersey Norman surnames, GueN = Guernsey Norman surnames, JerO = Jersey other surnames, GueO = Guernsey other surnames

Table 2.9. Haplogroups and Modal Haplotypes Encountered in the Populations Studied

Population \ Hg	E3b	F*(xIJK)	J*(xJ2)	J2	I*(xI1b2)	2.47+1	I1b2	K*(xPN3)	N3	P*(xR1)	R1*(xR1a1)	AMH+1	R1a1	3.65+1	n
Shetland	-	-	-	-	3	3	-	-	-	-	11	32	4	10	63
Orkney	-	-	-	-	9	8	1	-	-	2	28	50	9	14	121
Durness	-	-	-	-	2	5	-	-	-	-	24	17	1	2	51
Western Isles	-	-	-	-	16	6	-	-	-	-	15	43	3	5	88
Stonehaven	-	1	-	1	1	5	-	-	-	-	14	20	2	-	44
Pitlochry	-	-	-	3	4	-	-	-	-	-	10	23	-	1	41
Oban	-	1	-	-	2	1	-	-	-	-	11	25	1	1	42
Morpeth	-	2	1	3	11	6	-	-	-	-	20	49	2	1	95
Penrith	3	1	-	2	7	9	-	-	-	-	14	47	2	5	90
Isle of Man	1	-	-	-	5	5	-	-	-	-	9	34	5	3	62
York	2	1	-	-	7	8	-	-	-	-	9	17	1	1	46
Southwell	4	1	-	4	9	3	-	-	-	-	14	31	3	1	70
Uttoxeter	3	1	-	3	7	8	-	-	-	-	22	38	-	2	84
Llanidloes	3	2	-	1	4	7	-	-	-	-	11	27	2	-	57
Llangefni	3	-	-	1	3	-	-	1	-	-	17	54	1	-	80
Rush	-	-	-	-	7	-	2	-	-	-	33	31	1	2	76
Castlerea	-	-	-	-	3	-	1	-	-	-	16	23	-	-	43
Norfolk	4	3	-	2	17	17	-	-	-	-	27	46	2	3	121
Haverfordwest	2	-	-	-	1	-	1	-	-	-	16	38	1	-	59
Chippenham	-	1	-	2	3	7	1	-	-	-	8	25	3	1	51
Faversham	2	-	-	3	2	4	-	-	-	-	14	28	1	1	55
Midhurst	1	-	1	3	9	4	2	-	-	-	16	43	1	-	80
Dorchester	3	1	1	2	5	5	-	-	-	-	17	36	3	-	73
Cornwall	-	-	-	1	2	4	-	-	-	-	13	28	3	1	52
Channel Islands	5	2	1	1	13	14	4	-	-	-	34	50	3	1	128
Jersey	2	1	1	-	11	13	3	-	-	-	21	28	2	-	82
Guernsey	3	1	-	1	2	1	1	-	-	-	13	22	1	1	46
Jersey Norman Surnames	-	-	-	-	1	1	2	-	-	-	4	4	-	-	12
Jersey Other Surnames	2	1	1	-	10	12	1	-	-	-	20	21	2	-	70
Gue Norman Surnames	1	-	-	-	1	1	1	-	-	-	4	5	1	-	14
Guernsey Other Surnames	2	1	-	-	1	0	-	-	-	-	9	17	1	1	32
Norway	-	1	-	1	25	32	-	-	3	8	16	45	26	44	201
North Germany/Denmark	5	3	-	5	37	38	-	-	3	-	25	50	16	8	190
Basques	1	1	-	-	1	-	2	-	-	-	12	25	-	-	42

Notes: Hg nomenclature as per the YCC (2002), except for the modal haplotypes 2.47+1, AMH+1, and 3.65+1 defined by Wilson et al (2001a)

PC plots drawn using hg+1 frequencies reveal some striking patterns. Norway is a clear outlier compared to the British populations (Figure 2.8a) which tend to cluster together (and on the periphery includes Basques) at the opposite end of the axis. Orkney and Shetland are the closest British populations to Norway. North Germany/Denmark is a slight outlier on PC2. The first principal component explains 42.1% of the variation and is driven by frequencies of 3.65+1 which reaches its highest frequencies in Norway, and is absent from Basques who fall at the opposite pole of the axis (Figure 2.8). Populations with low 3.65+1 frequencies also tend to be characterised by high AMH+1 frequencies, and vice versa. The second principal component explains 17.4% of the variation and separates Norway from North Germany/Denmark. The simulated admixed populations with 60%, 40% and 20% Norwegian and North German/Danish input were subjected to PC analysis to aid interpretation of the PC plot in Figure 2.8b. This indicates that the 1st and 2nd principal components are a sensitive indicator to the relative contribution of Norwegian and North German/Danish Y-chromosomes on the British populations: Norwegian input moves populations strictly along the 1st axis and North German/Danish input moves populations at an angle to both axes.

Admixture proportions (median and 95% CIs) for all of the British populations are shown in Table 2.10. As anticipated for the single locus data used here, the range of values observed for p_1 is large (Chikhi *et al.* 2001; Chikhi *et al.* 2002), ranging from almost no input of either Norway or North Germany/Denmark, to almost 100% input, hence the median values must be treated with extreme caution and only used as an indicator to the relative inputs of the parental populations and in conjunction with other data. There are however notable trends. The Scottish islands have the highest median Norwegian input (Shetland: 0.683; Orkney: 0.553; Western Isles: 0.616), which reaches a low in Llangefni in particular (0.141), and Wales in general (0.134). North German/Danish input is typically higher in British populations than Norwegian input.

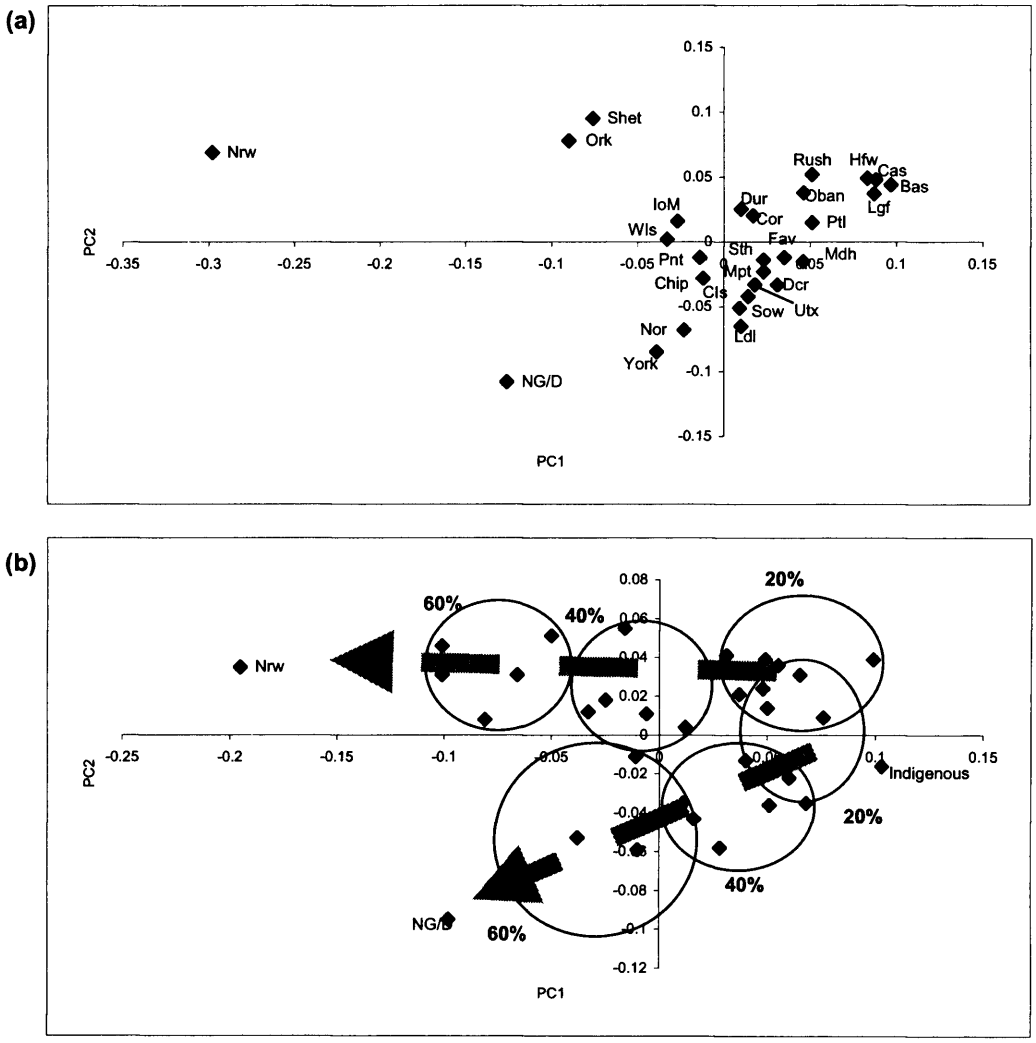


Figure 2.8. PC Plots of the British and European Populations Studied. (a) PC plot based on hg+1 frequencies shown in Table 2.9. PC explained 42.1% of the variation and PC explained 17.4%. (b) Simulated populations with 60%, 40%, and 20% input from Norway and North Germany/Denmark. Abbreviations as Table 2.8.

Table 2.10. Admixture Proportions for the British Populations Calculated by LEA

Map Number	Population	n	Admixture Proportion	Founder	2.5%	97.5%
1	Shetland	63	0.683	Norway	0.099	0.987
			0.716	NG/D	0.101	0.989
2	Orkney	121	0.553	Norway	0.105	0.966
			0.607	NG/D	0.071	0.975
3	Durness	51	0.639	Norway	0.108	0.986
			0.545	NG/D	0.053	0.977
4	Western Isles	88	0.616	Norway	0.099	0.986
			0.746	NG/D	0.099	0.991
5	Stonehaven	44	0.396	Norway	0.027	0.941
			0.576	NG/D	0.07	0.978
6	Pitlochry	41	0.462	Norway	0.023	0.968
			0.453	NG/D	0.027	0.956
7	Oban	42	0.37	Norway	0.024	0.954
			0.357	NG/D	0.023	0.955
8	Morpeth	95	0.429	Norway	0.027	0.957
			0.571	NG/D	0.08	0.974
9	Penrith	90	0.359	Norway	0.033	0.89
			0.544	NG/D	0.1	0.971
10	Isle of Man	62	0.582	Norway	0.066	0.973
			0.757	NG/D	0.18	0.99
11	York	46	0.524	Norway	0.053	0.947
			0.706	NG/D	0.136	0.984
12	Southwell	70	0.405	Norway	0.043	0.925
			0.529	NG/D	0.056	0.971
13	Uttoxter	84	0.266	Norway	0.015	0.876
			0.496	NG/D	0.052	0.971
14	Llanidloes	57	0.251	Norway	0.011	0.909
			0.542	NG/D	0.069	0.974
15	Llangefni	80	0.141	Norway	0.005	0.887
			0.147	NG/D	0.013	0.952
16	Rush	76	0.348	Norway	0.019	0.938
			0.292	NG/D	0.012	0.919
18	Norfolk	121	0.448	Norway	0.048	0.943
			0.725	NG/D	0.143	0.988
19	Haverfordwest	59	0.22	Norway	0.009	0.827
			0.215	NG/D	0.008	0.848
20	Chippenham	51	0.57	Norway	0.068	0.973
			0.708	NG/D	0.17	0.986
21	Faversham	55	0.23	Norway	0.013	0.796
			0.495	NG/D	0.059	0.968
22	Midhurst	80	0.151	Norway	0.006	0.678
			0.244	NG/D	0.015	0.863
23	Dorchester	73	0.227	Norway	0.009	0.829
			0.36	NG/D	0.03	0.926
24	Cornwall	52	0.408	Norway	0.044	0.969
			0.577	NG/D	0.074	0.985
25	Channel Islands[b]	128	0.219	Norway	0.01	0.796
			0.422	NG/D	0.003	0.958
-	Jersey[ab]	82	0.298	Norway	0.014	0.855
			0.577	NG/D	0.058	0.965

continued

Table 2.10. continued

Map Number	*Population*	*n*	*Admixture Proportion*	*Founder*	*2.5%*	*97.5%*
-	Guernsey[ab]	46	0.186	Norway	0.011	0.768
			0.234	NG/D	0.015	0.904
-	Jersey Norman Surnames[ab]	12	0.339	Norway	0.018	0.927
			0.451	NG/D	0.024	0.957
-	Guernsey Norman Surnames[ab]	14	0.324	Norway	0.020	0.878
			0.482	NG/D	0.036	0.960
-	Jersey Other Surnames[ab]	70	0.295	Norway	0.013	0.862
			0.645	NG/D	0.059	0.977
-	Guernsey Other Surnames[ab]	32	0.431	Norway	0.013	0.915
			0.431	NG/D	0.020	0.968
-	Scottish Isles	272	0.63	Norway	0.1	0.979
	(1, 2, 4)		0.515	NG/D	0.051	0.97
-	Scotland	178	0.339	Norway	0.025	0.968
	(3, 5-7)		0.478	NG/D	0.046	0.972
-	England	945	0.243	Norway	0.014	0.823
	(8, 9, 11-13,18, 20- 25)		0.375	NG/D	0.048	0.917
-	Wales	196	0.134	Norway	0.004	0.774
	(14, 15, 19)		0.101	NG/D	0.007	0.695

Shown are the median and 95% confidence intervals calculated on the distribution obtained from 10,000 of 50,000 Markov Chain steps. [a]Populations not included in Capelli *et al.* (2003). [b] Median and 95% confidence intervals calculated on the distribution obtained from 20,000 of 100,000 Markov Chain steps. *Figure adapted from Capelli et al. (2003)*

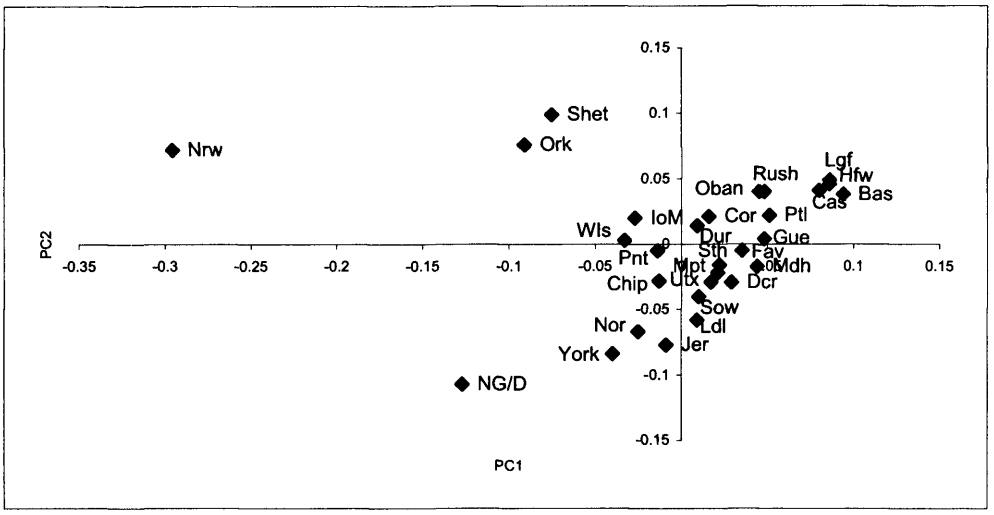
2.3.2. The Channel Islands

Results of the exact test of population differentiation based on hg+1 frequencies are shown in Table 2.8, which reveals several important findings. Jersey and Guernsey are not distinguishable at the hg+1 level, ($p = 0.06$). However to elucidate any differences in history between these islands they are still considered separately, and structured by surname for the purposes of analysis. Apart from the Jersey Norman and Guernsey Norman populations, each different way of clustering the Channel Islands samples reveals some significant structure with North Germany/Denmark, whilst all Channel Island populations are significantly different to Norway.

Considering both PC plots in Figure 2.8 the Channel Islands cluster with populations that have an enrichment of North German/Danish types, falling within the range of (simulated) populations with an estimated 40-60% North German/Danish influence. When Jersey and Guernsey are considered separately (Figure 2.9a) it shows that Jersey had the main effect in moving the Channel Islands towards North Germany/Denmark. When the Channel Islands are stratified by surname (Figure 2.9b) the resultant plot is very different to those in Figures 2.8 and 2.9a. Primarily, the discriminatory power previously seen on the second principal component is lost because of the high frequency of I1b2 chromosomes in Jersey, which forces Jersey as an outlier on this axis. Neither Norman Surname group shows a particular affinity with the North German/Danish population, although the small sample size of these Norman Surname groups must be considered.

Admixture proportions (median and 95% CIs) for the Channel Islands, grouped according to several criteria, are shown in Table 2.10 and posterior pdfs for p_1 and t are shown in Figures 2.10-2.12. As above the range of values observed for p_1 is large (Table 2.10, Figure 2.10), and median values are treated with caution. With these caveats in mind, trends are apparent. For each of the populations the median and 97.5% CI German/Danish input is always higher than the comparable Norwegian estimate. When Jersey and Guernsey are compared,

(a)



(b)

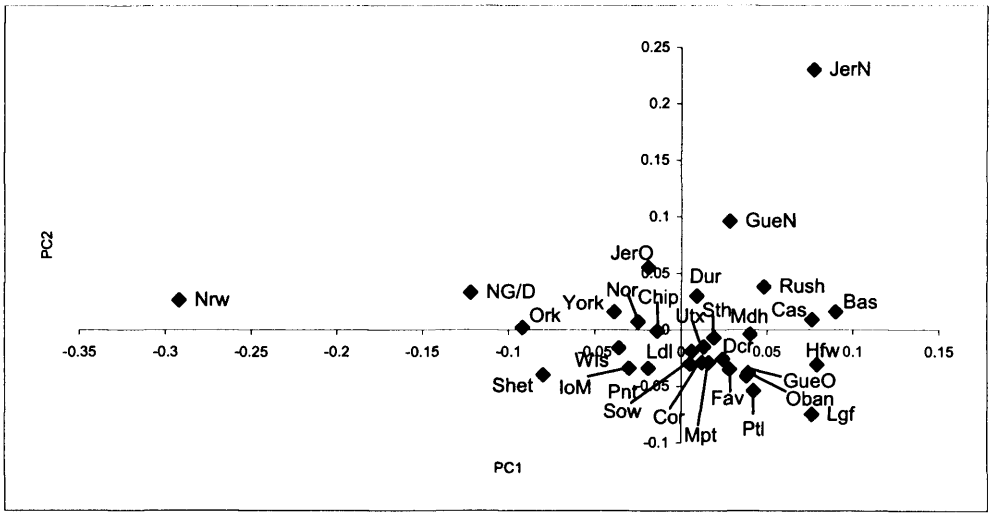


Figure 2.9 PC Plots of British Populations. (a) The Channel Islands have been separated into Jersey (Jer) and Guernsey (Gue).PC1 explained 41.4% of the variation and PC2 explained 17.5%. (b) Jersey and Guernsey have each been separated into two groups on the basis of surname (see text): Norman Surnames (JerN and GueN) and Other Surnames (JerO and GueO). PC1 explained 34.9% of the variation and PC explained 19.6%. Abbreviations as Table 2.8

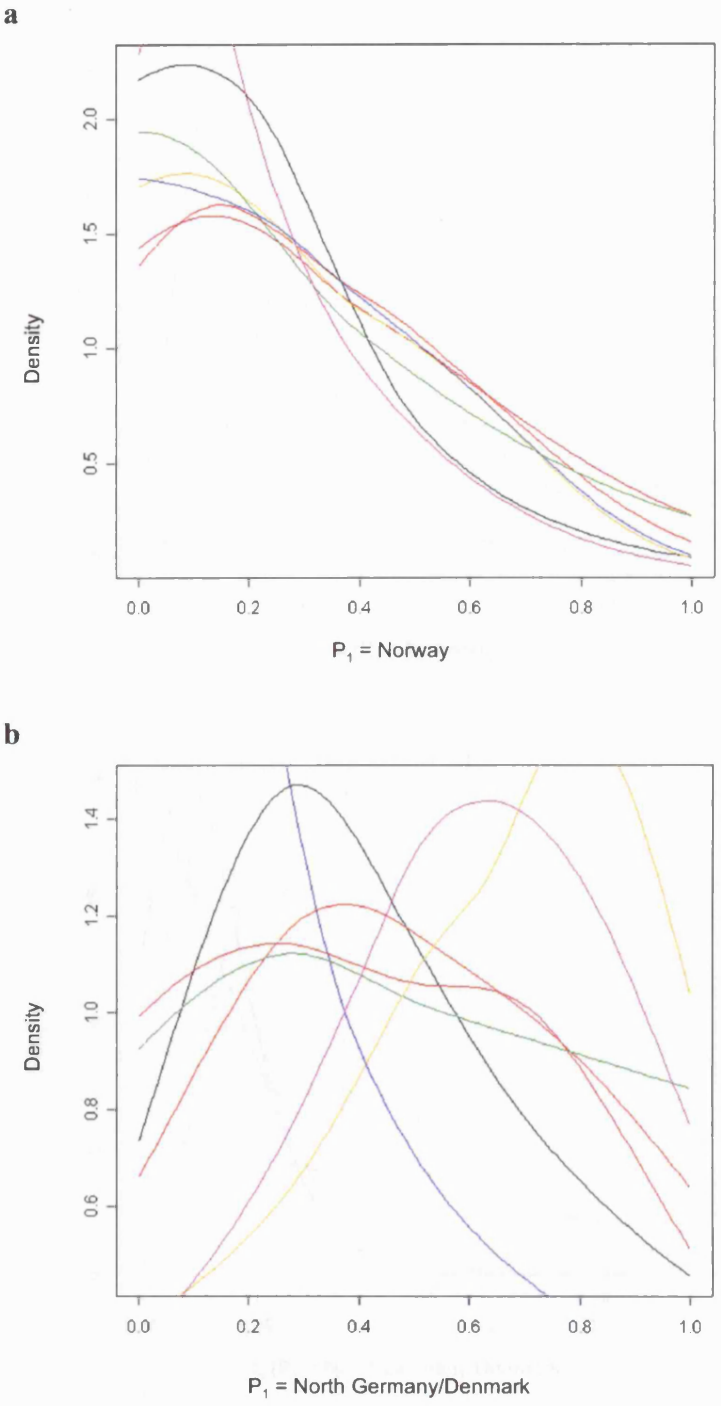


Figure 2.10. Posterior pdf's for p_1. (a) Posterior pdf where P_1 = Norway. (b) Posterior pdf where P_1 = North Germany/Denmark Populations as follows: All Channel Islands (black), Jersey (purple), Guernsey (blue), Jersey Norman (brown), Guernsey Norman (red), Jersey Other (orange), Guernsey Other (green).

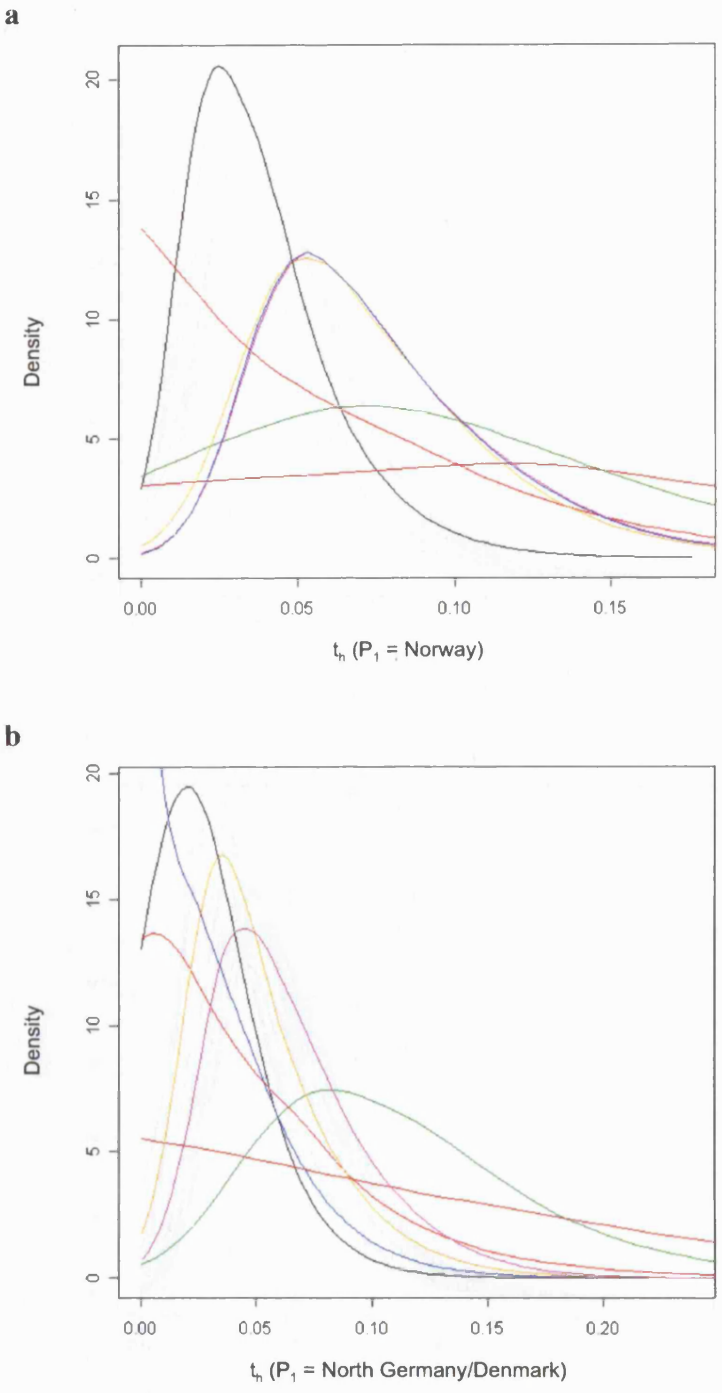


Figure 2.11. Posterior pdf's for t_h. (a) Posterior pdf of t_h where P_1 = Norway. (b) Posterior pdf of t_h where P_1 = North Germany/Denmark. Populations as follows: All Channel Islands (black), Jersey (purple), Guernsey (blue), Jersey Norman (brown), Guernsey Norman (red), Jersey Other (orange), Guernsey Other (green).

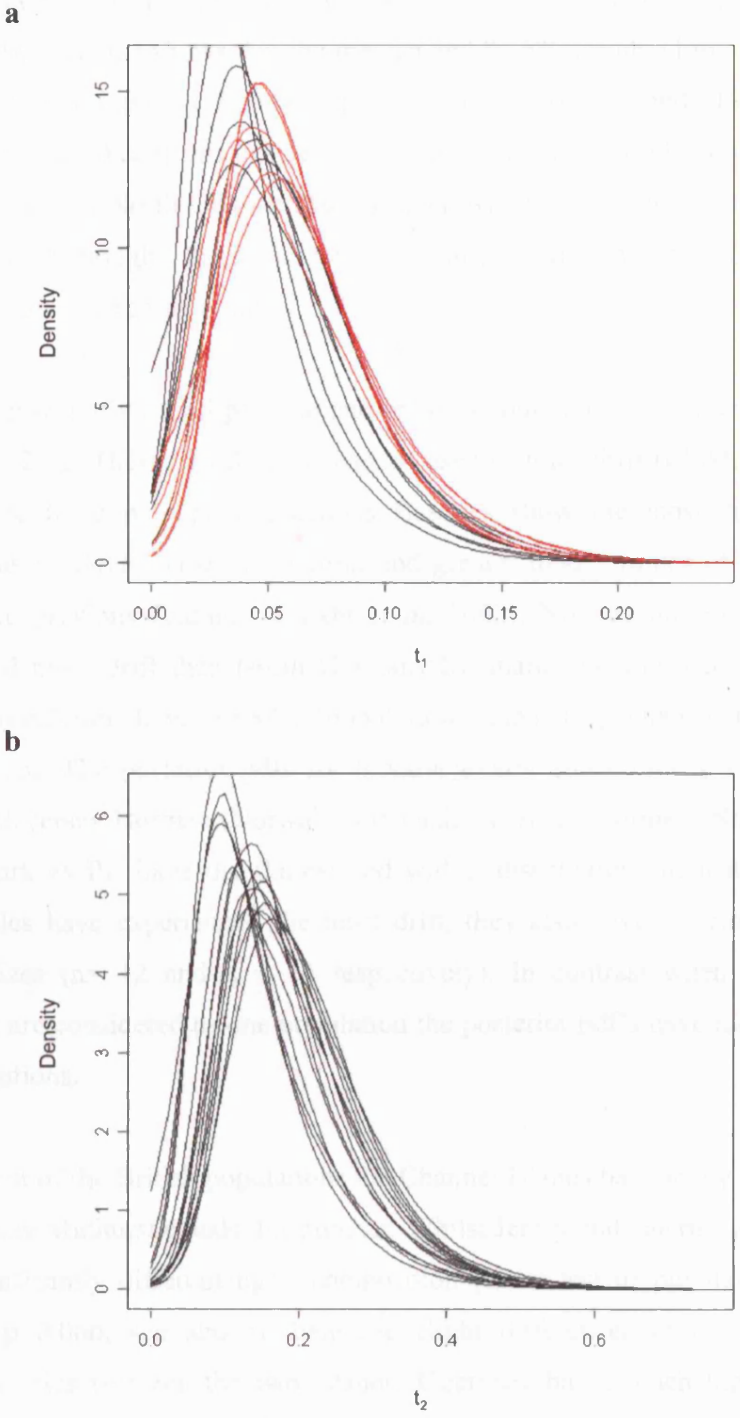


Figure 2.12. Posterior pdf's of t_1 and t_2. (a) Posterior pdf of t_1 where P_1 = Norway (red) and P_1 = North German/Denmark (black). (b) Posterior pdf of t_2 where t_2 = Basques and Castlerea to represent the indigenous Y chromosome gene pool of Europe.

Jersey has a higher German/Danish median value (0.577) and 97.5% credible limit (0.965) than Guernsey (0.234 and 0.904, respectively). Stratifying the Channel Islands by surname shows that the median and 97.5% credible limit for North German/Danish input is highest in Jersey Other (0.645 and 0.977 respectively). Figure 2.10 confirms that Jersey and Jersey Other are outliers with respect to the amount of North German/Danish input, which is greatest in these two populations, although there is clearly more variation in North German/Danish than Norwegian input.

t_1, t_2, and t_h posterior pdf's for all parental and hybrid populations are shown in Figures 2.11 and 2.12. These distributions can be used to infer drift (Chikhi *et al.* 2001). Of the three parental populations Basques show the most drift, inferred from the relatively wide distribution and greater modal values of the pdf's, confirming previous findings (Chikhi *et al.* 2002). Norway appears to have experienced more drift than North Germany/Denmark, possibly because Bergen and Trondheim have smaller populations than Copenhagen and Schleswig-Holstein. The posterior pdfs for t_h vary greatly and correlate with sample size. Both Jersey Norman (Norway as P_1) and Guernsey Norman (North Germany/Denmark as P_1) have the flattest and widest distributions indicating these two samples have experienced the most drift, they also have extremely small sample sizes (n= 12 and n = 14 respectively). In contrast when the Channel Islands are considered as one population the posterior pdf's have much narrower distributions.

Compared to most of the British populations the Channel Islands have many hgs present (n=8), only Midhurst equals this number. Whilst Jersey and Guernsey do not have a significantly different hg+1 composition (exact test of population differentiation, p=0.066, see above) there are slight differences in hg and haplotype frequencies between the two islands. Guernsey has a much higher frequency of AMH+1 than Jersey (47.8% vs 34.1%, although both frequencies are within the range seen in all other British populations; Table 2.9), and a comparatively lower frequency of I*(xI1b2) (6.51% vs 29.26%), with 2.47+1 and non-2.47+1 chromosomes grouped together. Both Norman Surname groups

have a reduced number of hgs compared to the Other surname groups, although this could be a function of sample sizes.

2.4. Discussion

To summarise the findings of this Chapter, the adoption of a systematic sampling approach for the British Isles made it possible to identify distinct patterns in the paternal history of settlement and gene flow from European populations known to have had strong cultural influences on the British Isles (Anglo-Saxons, Norwegian and Danish Vikings). None of the British populations studied here have evidence for complete replacement of indigenous Y-chromosomes by Norwegian or North/German Danish groups, even in locations such as Orkney, Shetland, and York where a strong Viking influence has been documented. Neither Jersey nor Guernsey seem to have an enrichment of Norwegian Y-chromosomes, but these data presented here indicates that the two islands have different histories of contact with North/Germany Denmark.

Results for the British Isles and the Channel Islands will be discussed separately in the following two sections, however before proceeding to discuss the results in more detail it is important to note that the lack of significant differentiation between the North German and Danish populations studied here meant it was impossible to distinguish between the genetic influence of Anglo-Saxons and Danish Vikings in the British Isles. This is unfortunate because some of the main arguments about the relative scale of demic versus cultural movements are focussed on the Anglo-Saxons (Welch 1992). Based on historical and linguistic research the lack of differentiation between North Germany and Denmark does not seem to be due to internal migration between these regions in the last 1,500 years (Forster *et al.* 1995) and may simply be the result of a more ancient shared ancestry due to the close proximity of these populations. Weale *et al.* (2002) recently used Frisia rather than Schleswig-Holstein to represent Anglo-Saxons. An exact test of population differentiation between their Frisian sample and the North Germany/Denmark sample analysed here showed that the two populations

are not significantly different from each other (p = 0.3). Therefore, even if the choice of Schleswig-Holstein rather than Frisia to represent Anglo-Saxons is mistaken, there does not appear to be significant differentiation between North German and Danish Y-chromosomes *per se*. Sampling from many locations in Germany and Denmark in a similar manner to that employed here for Britain could identify possible structure both within and between these countries and clarify matter. The use of additional microsatellites to the 6 routinely typed in this thesis may also reveal more structure within and between the present samples, indeed around 34 polymorphic microsatellites on the Y-chromosome are known (Hurles and Jobling 2001), and this figure is likely to be at least trebled in the near future (Jobling and Tyler-Smith 2003).

2.4.1. The British Isles and European Populations

Despite the problems discussed in the previous paragraph, several trends are apparent in the extent of influence of the European populations studied on the British Isles. In accordance with the well-documented links between Orkney and Shetland and Norway/Norwegian Vikings (Davies 1999), these two British populations have significant amounts of Norwegian Y-chromosome influence, agreeing with earlier findings for Orkney (Wilson *et al.* 2001a). Conversely, there is not evidence for an enrichment of North German/Danish types in Orkney and Shetland, confirming archaeological and historical records which suggest the main Anglo-Saxon (Welch 1992) and Danish Viking focus was England (Richards 1991; Davies 1999). Indeed other lines of evidence such as linguistics also indicate that populations at the northern and western limits of the British Isles did not experience the effects of various invading forces to the same extent as the rest of the Isles, seen for example in linguistic analyses (Bryson 1990; Cavalli-Sforza *et al.* 1994). Norwegian Vikings are an exception to this rule because of the direction from which they approached the British Isles. It is surprising that the Western Isles and the Isle of Man only appear to have a small Norwegian male component, based on their location on the PC plot (tangential to the main Norwegian and North German/Danish axes) as these islands are also

on the hypothesised route taken by Norwegian Vikings along the British west coast (Hill 1981; Davies 1999). This is particularly surprising for the Isle of Man, given the strong ties the island has with Vikings, which are still apparent today, such as the Tynwald (Richards 1991). Indeed the lack of evidence for an increased Norwegian input on the Isle of Man supports arguments against a mass migration of Norse Vikings to the Isle of Man, but instead a replacement of the ruling elite (Richards 1991). Penrith also exhibits a slight shift along this tangent that indicates a slight enrichment of Norwegian Y-chromosome types, which is notable because of the Scandinavian influence on the dialect of this region (Reaney 1927).

All of the English and Scottish sample locations show some degree of North German/Danish influence. For England this is not surprising because of the documented impact of both Anglo-Saxons (Welch 1992) and Danish Vikings (Richards 1991; Davies 1999), however it is extremely surprising for Scotland as neither Anglo-Saxons nor Danish Vikings are documented to have moved as far north as Scotland. Recent gene flow from England to Scotland may explain this pattern. York, Norfolk, Southwell and Llanidloes exhibit most North German/Danish influence. Apart from Llanidloes all of these locations have a strong documented historical presence of Danish Vikings and fell within the Danelaw (Davies 1999). That Llanidloes is within this group may be the result of recent migration within the last 200 years from England (Davies 1999) particularly as this location is closest of the 3 Welsh sample sites to England (Figure 2.5). The Isle of Man also shows some indication of increased frequencies of North German/Danish types, again this may be the result of migration from England. However none of the English populations have evidence for a complete replacement of indigenous Y-chromosomes by North German/Danish types, if the 97.5% credible limit for North German/Danish genetic input estimated using LEA is considered. These upper limits never reach 100% (for any English or British population), hence suggesting that the Anglo-Saxon period was not associated with the complete replacement of indigenous Y-chromosomes, and possibly not a mass migration of peoples.

This however contradicts the findings of Weale *et al.* (2002) who concluded that there had been substantial replacement of indigenous Central English Y-chromosomes by Anglo-Saxons, when Frisia was used as an Anglo-Saxon source population. As noted above, the Frisian sample of Weale and colleagues and the North German/Danish sample used in this study are not significantly different from one another ($p = 0.3$), hence the choice of source population alone cannot explain the different conclusions. Furthermore PC analysis including the Frisians indicates that they cluster most closely with the continental populations and not any English populations (Figure 2.13). Whilst the findings presented here do show that Central English populations have most Continental input in England, the frequency of AMH+1, which is interpreted as a signature of the indigenous European population (see below) in these populations is never lower than 44% (in Southwell). This is higher than the value in Frisia (35%), thus indicating a lack of complete replacement. The discrepancy between these present data and the conclusions of Weale *et al.* (2002) may be influenced by two important factors: (i) Weale *et al.* (2002) did not include representatives of Danish Vikings who had well documented activities in eastern England, although the inclusion of a Danish sample in the analysis of Weale *et al.* might be expected to increase the level of replacement by Anglo-Saxon types given the apparent similarity between North German and Danish Y-chromosomes; (ii) the choice of markers employed to define hgs is not identical, particularly for hg2 (using the terminology of Jobling and Tyler-Smith 2000, or BR*x(DE,JR) in the YCC 2002 terminology) with the present study employing a higher level of resolution by typing M170 which defines I*(xI1b2). As previously noted above in the Introduction (section 2.1.5), these two hgs should be broadly comparable, however in the present context the difference in resolution might be important because the high frequency of I*(xI1b2) and its modal haplotype 2.47+1 is one of the signatures of the North German/Danish sample (Table 2.9).

Although there is clear evidence for some degree of Norwegian and North German/Danish influence on all of the British locations, all have retained a substantial degree of indigenous Y-chromosomes reflected in the high frequency of R1*(xR1a1) and its modal type AMH+1, which is never found below 33%, a higher frequency than in either Norway or North Germany/Denmark.

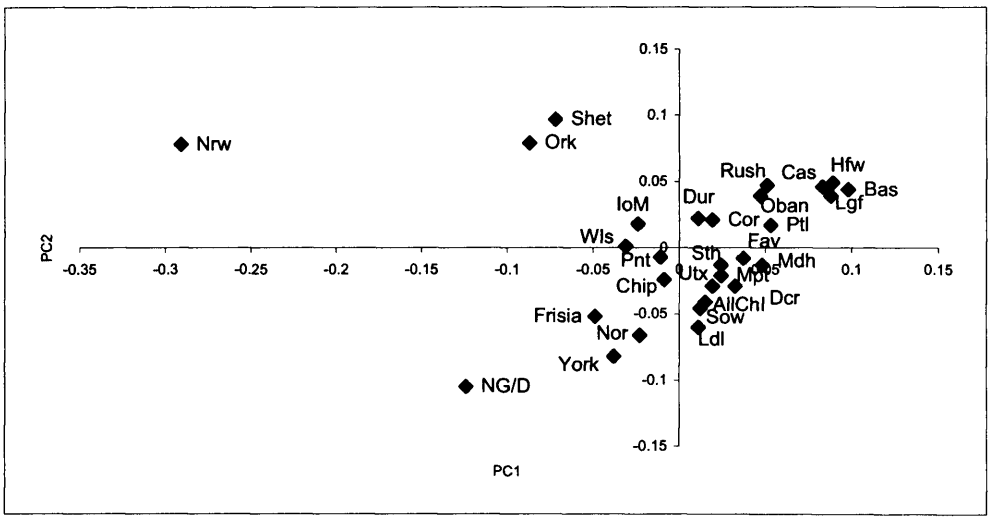


Figure 2.13. PC Plot Including Frisia. PC plot based on the hg+1 frequencies shown in Table 2.9. and including Frisia. Abbreviations as Table 2.8

R1*(xR1a1) and the equivalent hg 92R7 are considered to represent part of the indigenous (or Palaeolithic) component of the European Y-chromosome gene pool (Semino *et al.* 2000; Jobling and Tyler-Smith 2003) because they are at highest frequencies in the Basque population (Rosser *et al.* 2000; Semino *et al.* 2000; Wilson *et al.* 2001a; Wells *et al.* 2001) who are considered to have retained a gene pool that is most representative of the Palaeolithic gene pool based on palaeoanthropology, archaeology, linguistics as well as genetics (Calafell and Bertanpetit 1994; Comas *et al.* 2000; Cavalli-Sforza *et al.* 1994; Chikhi *et al.* 2002). It is simplistic to interpret the frequency of a single haplogroup or haplotype as an indication the degree of admixture (Chikhi *et al.* 2002). However the conclusions based on the frequency of R1*(xR1a1) and AMH+1 are confirmed by admixture proportions which show that the indigenous input is never zero. The high frequency of R1*(xR1a1), and the equivalent hg defined by 92R7 derived chromosomes, in British populations and their similarity with Basques is confirmed by other studies (Rosser *et al.* 2000; Wells *et al.* 2001; Wilson *et al.* 2001a).

2.4.2. The Channel Islands

North German or Danish men have contributed more to the Jersey gene pool than that of Guernsey, reflecting and confirming (Hawkes 1937; Briggs 1995) different histories of contact between these islands and other European populations. Neither Jersey nor Guernsey has evidence for an enrichment of Norwegian Y-chromosomes. Interpreting the PC plot in Figure 2.9a in tandem with Figure 2.7b revels that Jersey clusters near populations with an estimated 40-60% North German/Danish input, whilst Guernsey is with populations that have little North German/Danish input. Admixture estimates confirm these findings in revealing a much higher median North German/Danish input in Jersey than Guernsey, although these point estimates must be treated with caution (Chikhi *et al.* 2001; Chikhi *et al.* 2002) given the large credible intervals observed here. The median North German/Danish input in Jersey is above the

average for England, although much lower than in areas where Danish Viking influence has been well documented, such as York (Davies 1999).

There are two historical and two more recent explanations for the enrichment of North German/Danish types in Jersey, which will be considered in turn. As there is not any history or folklore of Anglo-Saxon settlement on the islands, the addition of North German/Danish Y-chromosomes preferentially to Jersey either happened during the supposed 9th century Viking raids, or during the three centuries that the Channel Islands were part of the Duchy of Normandy (Stevenson 1986), Normandy having been founded by Danish Vikings. Unfortunately it is impossible to differentiate between these two events because any raids were most likely by Danish Vikings due to the route they took to England (Richards 1991; Davies 1999). However, given that the links with the Normandy region of France were strong and remained so after the Channel Islands were no longer part of the Duchy (Stevenson 1986) it is more likely that the ties with Normandy, rather than Viking raids, contributed the North German/Danish Y-chromosomes. That Jersey has been preferentially affected also fits with this interpretation and previous accounts (Hawkes 1937) hypothesising that Jersey has been more influenced by the French mainland than Guernsey. One firm conclusion that can be made both on the basis of PC analysis and admixture analysis is that any Vikings raids on the Islands did not involve large numbers, if any, Norwegian Vikings.

In terms of recent events that may explain the North German/Danish enrichment, the first is the German occupation of Jersey. It is impossible to quantify what, if any, effect this had on the local gene pool, particularly as announcing illegitimate births that resulted from liaisons between German soldiers and local women would have been taboo. It is interesting to note however that the rate of illegitimate births on Guernsey during this period rose to an all time high of 21.8% (Briggs 1995). This of course suggests that Guernsey could have a proportion of Y-chromosomes contributed by Germans, which does not appear to be borne out by these data. Occupation of the Channel Islands was, however, for a relatively short period of time (1940-1945) and the German army would have been a heterogenous mix of men from across

Germany with different Y-chromosomes hgs. Hence given the relatively high frequency of I*(xI1b2) chromosomes in Jersey, and the potential of such Y-chromosomes to be interpreted as a signature of Anglo-Saxon or Danish Viking input, it seems unlikely the occupation contributed substantially to the gene pool of Jersey. A final possibility is that the immigration of British people to Jersey from the end of World War II until the 1970s preferentially introduced these North German/Danish types; indeed migration from England to Scotland, the Isle of Man, and Llangefni is a plausible explanation for their enrichment of North German/Danish types. However such migrations should have been controlled for by the sampling strategy, which only sampled from male donors who could trace their paternal lineage to the same island as far back as their paternal grandfather.

North German/Danish Y-chromosomes therefore have had less influence on Guernsey, possibly due to its location further away from the French mainland. There is indeed subtle evidence to support the notion that Guernsey has been more influenced by the Iberian peninsula than Jersey (Hawkes 1937) in the frequencies of AMH+1 which is at much higher frequency in Guernsey than Jersey. However, the frequency in Guernsey is lower than in Castlerea, Haverfordwest, and Llangefni, all of which are concluded to have retained most so called indigenous Y-chromosomes in Britain. Whilst the trend for Guernsey may be as much due to drift as a consequence of relative isolation *since* the Neolithic as differences in settlement patterns, the haplogroup and haplotype structure of Guernsey argues against total isolation. Small isolated populations are disproportionately affected by drift (Cavalli-Sforza *et al.* 1994), hence isolates are characterised by fewer hgs and haplotypes than seen in less isolated populations. If one assumes that the Basques are an isolated population (Calafell and Bertanpetit 1994; Comas *et al.* 2000; Cavalli-Sforza *et al.* 1994; Chikhi *et al.* 2002; but also see Hurles *et al.* 1999), a comparison between Guernsey and the Basques studied here shows that Guernsey has more haplogroup and haplotype diversity. Guernsey has 7 hgs, whilst Basques have 5 and in terms of haplotypes, Guernsey has 27 whilst Basques have 18 (sample size cannot be a factor here because they are comparable: 46 and 42 respectively), and the most common haplotype in Guernsey (AMH+1) only comprises 15.2% of the total

number of haplotypes, whilst the commonest Basque haplotype (also AHM+1) comprises 31% of all haplotypes.

The presence of the rare hg I1b2 in Jersey and Guernsey implies a common origin, at least in part, for these two islands, particularly as the haplotypes on both islands are closely related. Two different haplotypes comprise the 4 I1b2 chromosomes (13-13-11-17-23-9 and 13-13-11-17-23-10), they are one step neighbours and rare in the other populations studied here (exact matches are found in one Orcadian and one Basque individual, see Appendix, Table A.3). I1b2 is considered as a signature of the indigenous European population because of its high frequency in the Basques (this study, Francalacci *et al.* 2003; Semino *et al.* 2003). In the Channel Islands it may be a signature of the migration of Iberian peoples to both Jersey and Guernsey, although migration between the two islands could also be a factor. It must be remembered that unlike most of the rest of Europe (Klein 1999) the Channel Islands have not been constantly occupied by modern humans since the Palaeolithic. The first evidence for constant settlement is during the Neolithic (Bender 1986), hence the presence of indigenous Y-chromosome types on the Channel Islands must reflect migration events after the Palaeolithic, rather than traces of Palaeolithic populations. A further possibility is that these types have been recently introduced to the islands by the recent Spanish migrants (http://www.bbc.co.uk/legacies/immig_emig/channel_islands/jersey/article_2.shtml; 30th March 2004).

Hill *et al.* (2000) had considerable success in using surname information to differentiate Irish Y-chromosomes into two groups (Gaelic and non-Gaelic surnames), and indeed found significant structure between these groups. A similar approach was used here in an attempt to identify a specific Norman signature on Jersey and Guernsey. Stratification by surname produced a distinct pattern of hg frequencies with both Norman Surname groups having fewer hgs (R and I lineages only) than both Other Surname groups, although this may be a function of sample size. However the only significant difference between these four groups (i.e. Jersey Norman surnames, Jersey other surnames, Guernsey Norman surnames and Guernsey other surnames) was between Jersey Other and

Guernsey Other. Hence there is not significant genetic structure by surname. An interesting finding is that the Norman surname groups are not significantly different to North Germany/Denmark, whereas the other Jersey and Guernsey surnames are significantly different (also note that when Jersey and Guernsey are not stratified by surname, they are significantly different to North Germany/Denmark). In contrast PC plots show that neither of the Norman surname groups are pulled towards the North German/Danish population. Indeed median admixture proportions indicate that the highest North German/Danish input for any of the stratified populations is for Jersey Other, and not the Norman Surname groups. Thus stratifying Jersey and Guernsey by surname is not as straightforward as Hill *et al.* (2000) found for Irish Y-chromosomes. It is possible that the Norman surname sample sizes are too small to be meaningful (indeed indicated by the posterior pdfs for t_h which have the greatest range of values for the Norman Surname groups), or the criteria used to classify surnames is incorrect. Both of these issues could be resolved by larger sample sizes of men with Norman surnames from Jersey and Guernsey.

2.5. Conclusions

Only through the systematic and large scale survey of British Y-chromosomes has it been possible to identify distinct geographical structuring of Y-chromosome types within the British Isles, the distribution of which has been modified by the history of contact with other European populations and more recent gene flow within the Isles. Although it has been shown that European Y-chromosomes exhibit levels of geographic structuring (see for example Rosser *et al.* 2000; Zerjal *et al.* 2001) it was not clear whether this held true over shorter geographical distances.

Chapter 3. What's in a Name: How do Surnames and Y-Chromosomes Correlate?

3.1. Introduction

There are various theories about when and why surnames started to be used in Britain. The most widely accepted is that after the Norman Conquest in 1066 a relatively limited number of Norman first names came to replace the plethora of Old English names so that within one village there could be several men called William, for example, and surnames were needed to differentiate them (Reaney 1997). The adoption of surnames was not a uniform process however, starting to be used first in the south, and by families with land. It took some time before surnames became completely hereditary in the sense that we know today (Lasker 1985). For example, written records from the end of the 13th century show how two brothers could have different surnames, and the same person's surname could also change during their lifetime (Reaney 1997). Spelling was also unstable for much of the early period of surname establishment in the UK (Lasker 1985) and only started to become standardised once the printing press was introduced to England by Caxton in 1476 (Bryson 1990), but only from around the 19th century did surnames start to adopt a standard form as a result of the keeping of public records being passed on from the church to the state (Lasker 1985).

Surnames are used in a variety of contexts to make inferences about population structure and migrations (Jobling 2001), therefore it is important to understand their mode of inheritance. For example the first use of surnames to study inbreeding has been traced to George Darwin in 1875 when he used the occurrence of marriages between people with the same surname (i.e. isonymy) to study the potential deleterious effects of consanguineous mating (Lasker 1985, and references therein, Jobling 2001). Recently surnames have been used to estimate the geographical origin of migrants in France (Degioanni and Darlu 2001), to study the population structure of Austria (Barrai *et al.* 2000), even to classify populations into ethnic groups in epidemiological studies (Abbotts *et al.* 1999). There are over a million different surnames in the world (Lasker 1985), therefore they are an abundant source of information that can be a useful proxy for more detailed genetic information, which is often harder and more costly to

obtain. The value of surnames as suitable proxies, however, depends on the extent to which they mimic an inheritance pattern, or the extent to which they reflect patterns of genetic geographic structure.

Genetics is a powerful tool that can be used to understand more about the history of surnames as one is not reliant on limited written records. Specifically the paternally inherited MSY is ideal for studying surnames in patrilineal societies such as Britain because surnames should be associated with particular Y-chromosome types due to their parallel patterns of inheritance (Jobling 2001). Additionally, as previously described in Chapter 2 (2.1.5), European Y-chromosome diversity is becoming increasingly well characterised with clear indications of geographic structure. The 1,772 British Y-chromosomes typed for UEPs and microsatellites presented in Chapter 2 are an ideal reference database of high resolution Y-chromosome diversity from small regional populations around Britain with which to compare the Y-chromosomes of (British) men with different surnames. Moreover, the fact that most regions in Britain are included in the database means that surnames can be matched with geographically localised control populations of randomly collected Y-chromosomes for comparative purposes. This is particularly pertinent as it is clear that British Y-chromosomes are structured over relatively small distances as a function of both geography and the differential history of contact with other European invading populations (Chapter 2). Given the recent nature of surname establishment in Britain the analysis of Y-linked microsatellites is critical, as argued in the Introduction, therefore the data presented in Chapter 2 are especially suited as comparative data for British surnames. Whilst UEPs provide the unequivocal assignment of hgs the higher mutation rate of microsatellites means they can be used are used to infer closer evolutionary relationships (See for example Hammer and Zegura 2002).

The expected relationship between surnames and the Y-chromosome is not simple however. If one assumes that surname inheritance is to some extent non-random and correlates with the Y-chromosome, several different events during the history of a surname will potentially disrupt its association with the Y-chromosome: multiple origins of the same surname, random adoption of the

name during its lifetime, non-paternity, and subsequent drift, such as the loss of some lineages and the proliferation of others (Jobling *et al.* 2003). Thus, finding disassociation between a particular surname and a Y-chromosome type implies one or more of these events have occurred; genetically differentiating between the events can however be difficult because the net effect on present day genetic diversity within a surname tends to be the same. Multiple origins of a surname (by men with different Y-chromosome types) will appear as distinct clusters of types but so will random adoption of the surname or non-paternity (by men with different Y-chromosome types) if they happened several generations ago and the lineages have not been lost through drift. In contrast, recent random adoption of the name or recent non-paternity will appear in the dataset as low frequency types or singletons. The properties of a particular surname, such as whether the name is rare or common, lead to expectations about its history, hence its correlation with the Y-chromosome. Rare names are expected to have a single origin and common names to have multiple origins, these are intuitive expectations that seem to be supported by surname history studies (Lasker 1985).

An additional caveat to using the Y-chromosome to study surname history is that it relies on the association between Y-chromosomes and surnames to be real and not a genetic artefact of the distribution of Y-chromosome types in Britain. Yet it is known that the distribution of Y-chromosome polymorphisms across Europe (Rosser *et al.* 2000; Wells *et al.* 2001; Casallotti *et al.* 1999) and Britain (Wilson *et al.* 2001a; Weale *et al.* 2002; Chapter 2 and Capelli *et al.* 2003) is non random and has been influenced by ancient and historical events, such as population migrations (Wilson *et al.* 2001a; Weale *et al.* 2002; Rosser *et al.* 2000; Wells *et al.* 2001; Chapter 2 and Capelli *et al.* 2003), and political and geographic boundaries (Weale *et al.* 2002). This will complicate assessments of the historical of surname adoption, although some patterns are expected to be apparent.

A number of publications have recently focussed on surnames and the Y-chromosome, which will now be discussed briefly. The relatively small number of studies, and the very different approaches they have taken to the study of

surnames means that general conclusions about surname inheritance are difficult to make. Hill *et al.* (2000) found significant differences in the frequency of hg P (defined by the mutation 92R7) between clusters of Gaelic and non-Gaelic surnames, with the Gaelic names showing much higher frequencies of hg P. However as the test of significance was based on the groups of surnames it does not show whether or not *individual* surnames would provide enough information to find similar patterns. Soodyall *et al.* (2003) used a combination of UEPs and microsatellites to assess the accuracy of genealogical records pertaining to the people and population history of Tristan da Cunha and identified instances of non paternity as well as the introduction of a novel type from outside the Tristan da Cunha gene pool. This study directly validates the efficacy of Y-linked UEPs and microsatellites (specifically the 6 microsatellites used in this study: DYS388, 393, 392, 19, 390, and 391) to study recent genealogical history and surnames, as direct comparison between historical records and the Y-chromosome types was possible. Note however that within the context of surname research this study is somewhat unusual in having detailed historical records that cover the period in question.

Sykes and Irven (2000) recently investigated the origins of the Sykes surname using four Y-linked microsatellites. Microsatellite haplotypes were obtained for 48 male Sykes, of these 48 men 21 (43.8%) shared the same haplotype that was absent from control populations, leading to the conclusion that this surname had a single origin. This contradicted historical expectations, which suggested multiple origins for the name. However, given knowledge of the distribution of Y-chromosome haplotypes in Britain (Chapter 2, Capelli *et al.* 2003), which although showing evidence for structure do exhibit high frequencies of AMH+1 types, it appears to be fortuitous that the majority of male Sykes shared a *rare* haplotype (which, using the 4 microsatellites that were comparable, is also rare in the data of Chapter 2). Chance predicts that they would share a haplotype within the AMH+1 cluster. Therefore the methodology of Sykes and Irven (2000) is not rigorous enough to be applied to surnames *per se* because in the case of a surname with a (potential) modal cluster in a common hg, UEPs should be used to confirm that all of the haplotypes belong to the same hg. Given that there is expected to be some overlap in the microsatellite haplotypes found in

closely related haplogroups, it is also necessary to analyse UEPs. Furthermore as Sykes and Irven (2000) only addressed the relationship between one surname and the Y-chromosome, employing a limited set of Y-chromosome markers, it was pertinent to gain an understanding of the general patterns of surname inheritance, which the work in this chapter addresses. Furthermore, by assaying additional microsatellites and utilising UEP markers it was possible to provide further resolution and discrimination power between individuals, hence increase the certainty with which conclusions are drawn.

3.1.1. Aims of this Chapter

The work in this chapter addresses the fidelity of surname inheritance in several British surnames. Surname inheritance is assumed to be paternal, hence mimicking the inheritance pattern of the Y-chromosome. Therefore in this chapter the Y-chromosome was used to investigate surname inheritance. The availability of a comprehensive dataset of Y-chromosomes from regions across Britain, as described in Chapter 2 (Capelli *et al.* 2003), allowed detailed comparisons to be made between the chosen surnames and a large number of randomly collected samples that acted as control populations.

3.2. Materials and Methods

3.2.1. The Study Populations

551 DNA samples were collected from 9 different surnames (Barnfather (n=18), Causton (n=55), Folland (n=5), Farrer (n=50), MacLeod (n=365), Sorbie (n=11), Speechley (n=18), Thwaite (n=8), and Whittock (n=21)) distributed across England, Wales and Scotland. Barnfather, Causton, Farrer, and Whittock all have associated spelling variants that are included in the dataset (Table 3.1); where the distinction between the variants is not important each of these surnames will simply be referred to as Barnfather, Causton, etc., otherwise the

Table 3.1. Summary of of the Surnames Studied

Surname and spelling variants	*Count in telephone directory (2003)*	*Count in 1901 Census*	*Sample Size (% sampled of 2003 counts)*	*Surname Type*[c]
Barnfather	167	267	10 (6.0)	Relationship/Nickname?
Banfather	1	5	2 (>100)[a]	Relationship/Nickname?
Bairnsfather	2	30	6 (>100)[a]	Relationship/Nickname?
All variants	170	302	18 (10.6)	-
Causton	113	437	5 (4.42)	Local
Cason	93	408	7 (7.53)	Local
Costen	61	159	1 (1.64)	Relationship
Causon	68	214	4 (5.88)	-
Costin	207	515	5 (2.42)	-
Corston	42	107	7 (16.67)	Local
Cawston	103	216	5 (4.85)	Local
Coston	61	246	6(9.84)	Local
Caston	70	359	12 (17.14)	Local
Corsten	2	3	1(50)	-
All variants	820	2664	55 (6.71)	
Farrer	516	1,912	25 (4.85)	Occupation
Fairer	43	119	8 (18.6)	Occupation
Farrar	1101	3,947	8 (0.73)	Occupation
Farrow	1929	6,312	8 (0.41)	Occupation
Ferrer	68	118	1 (1.47)	Occupation
Pharaoh	57	137	1 (1.75)	Nickname
All variants	3714	12,545	50 (1.35)	
Folland	182	424	8 (4.40)	-
MacLeod	7475	26,321	365 (4.88)	Relationship
Sorbie[b]	37	295	12 (32.43)	Local
Speechly	140	378	18 (12.90)	-
Thwaite[b]	228	770	8 (3.51)	Local
Whittock	111	343	15 (13.51)	Relationship
Whittuck	3	22	1 (0.333)	-
Whytock	98	218	5 (5.10)	Relationship
All Variants	212	583	21 (9.9)	

[a] The telephone directory does not have 100% coverage of the British population, for these rare variants there are samples from more people than are listed in telephone directory.

[b] These names have associated spelling variants, see Reaney (1997), but only the variant listed here is studed

[c] Information from Reaney (1997); a question mark indicates inconclusive evidence and where no information is available the field is left blank

variants are referred to separately. The existence of spelling variants associated with these names was based on information from Reaney (1997).

The surnames that were studied were selected from scores of letters received by Prof. Goldstein as a result of media coverage of other work carried out in the lab. Surnames were chosen on the basis of their frequency, distribution, and hypothesised origins, such that the final set of surnames analysed was heterogenous with respect to these properties (see Table 3.1). For example, Sorbie is extremely rare with only 37 entries in the UK Telephone Directory (ascertained from Directory Enquiries on the bt.com website (see Section 3.2.4 for more information), whilst MacLeod is relatively common with 7475 entries. None of the names in the present study are amongst the commonest in England and Wales however (ascertained from marriage records in 1975; Sokal *et al.* 1992) or in the top 20 commonest surnames in Scotland (ascertained from birth, marriage and death records from 1999-2001; Bowie and Jackson 2003). As summarised in Table 3.1 several of the surnames are so-called local surnames i.e. from a place name (several Causton variants, Sorbie, and Thwaite); others are surnames of relationship (possibly Barnfather, Banfather, Bairnsfather, and Costen, MacLeod, Whittock, Whittuck, Whytock); surnames of occupation or office (Farrer variants except possibly Pharoah); and nicknames (possibly Barnfather, Banfather, Bairnsfather, and Pharoah).

3.2.2. Sample collection

Volunteers within each surname were, to the best of their knowledge, unrelated back to at least paternal grandfather. Contact with the volunteers was initially made by a representative from each surname, a verbal agreement to take part in the study was obtained and subsequently a buccal swab kit was sent to each volunteer by post for the volunteer to take their own sample and return to the lab for analysis. Appropriate informed consent was obtained. The buccal swabs were stored in tubes containing 1ml of 0.05 M EDTA/ 0.05M SDS preservative solution until extraction. DNA was extracted using a standard

phenol/chloroform method and the Promega Wizard ® Genomic DNA Purification Kit following manufacturers instructions, except centrifugation time was increased from 3 mins at 13,000g to 6 minutes at 13,000g. Both methods yielded approximately 5ng/μl of DNA. Samples were redydrated in a TE buffer solution and stored at –20°C.

3.2.3. Y-Chromosome Genotyping

All male DNA samples were typed for the YSTR1 and EURO1 PCR multiplex kits described in Tables 2.3 and 2.4 and the additional 5 UEPs detailed in Tables 2.5–2.7, where relevant. Due to a change in genotyping technology from the ABI PRISM ® 377 DNA Sequencer employed in Chapter 2 to the ABI PRISM ® 3700 DNA Sequencer 3 methodological modifications were required. First the TET fluorescent label on several of the primers needed to be changed to NED (details of these primers can be found in the Appendix, Table A.2). Secondly, due to differences in the migration of PCR fragments between the 377 and 3700 Sequencers, the expected allele sizes (for UEPs and microsatellites) had to be corrected using two control samples with known 377 allele sizes. Sizes obtained using the 3700 sequences were thus altered to the equivalent 377 size using the guidelines listed in Table 3.2. This allowed direct comparison between the Y-chromosomes in Chapter 2 and those presented in this chapter; hence the reported microsatellite repeat sizes in this chapter have had this correction applied. Finally, all samples were kindly electrophoresed by A Smith and M-W Burley, rather than by the author. Briefly, the samples were electrophoresed in 10μl formamide, using the ROX size standard, Filter Set D, and POP 6.

92r7 failed to amplify on the majority of samples and for time constraints the cause of this problem could not be investigated. These samples were not excluded from analysis however because 92r7 derived status can be inferred for most samples from M173 derived state (see Figure 2.6). Moreover, only 3 surname samples were found to only be M9 derived (hence may have been 92r7

Table 3.2. Differences in Allele Size Between the ABI PRISM ® 377 and 3700 DNA Sequencers for Assayed UEPs and Microsatellites

UEPs: Expected ancestral and derived allele sizes for the ABI PRISM ® 377 and 3700 DNA Sequencers

Locus	*Ancestral Allele (377 Size) and* ***3700 Size***	*Derived Allele (377 Size) and* ***3700 Size***
M9	(67) **63-G**	(97) **93-C**
92r7	(66) **62-C**	(95) **92-T**
M17	(123) **118-G**	(104) **100-G**
M173	(99) **95-A**	(118) **114-C**
M170	(83) **76-A**	(111) **105-C**
M172	(172) **170-T**	(143) **145-A**
M26	(169) **166-G**	(149) **143-A**
M89	(80) **74-G**	(98) **95-T**
12f2	(88) **83**	-
Tat	(83) **80-A**	(112) **106-C**
M35	(130) **126-G**	(160) **154-C**

Microsatellites: Expected range of microsatellite repeat sizes for the ABI PRISM ® 377 and 3700 DNA Sequencers and the average difference in allele sizes

Locus	*Repeat Size*	*Size Range in bp (377 sizes) and* ***3700 sizes***	Difference (converting from 377 to 3700)
DYS19	Tetranucleotide	(182-202) **180-200**	-2
DYS388	Trinucleotide	(119-145) **117-143**	-2
DYS390	Tetranucleotide	(192-220) **188-216**	-4
DYS391	Tetranucleotide	(148-173) **145-170**	-3
DYS392	Trinucleotide	(148-173) **145-170**	-3
DYS393	Tetranucleotide	(106-130) **102-126**	-4

Note: Two DNA samples with known (377) allele sizes were included as controls in each 3700 run

derived, but M173 ancestral) therefore only minimal information was lost by not having 92r7 status.

3.2.4. The Geographic Distribution of the Surnames in England, Wales, and Scotland

Regions in Britain where each name is presently most common were identified such that the Y-chromosome diversity in these approximate areas could act as controls (or Geographic Neighbours, see below) against which to test the observed diversity in each surname. The Y-chromosome diversity information was selected from appropriate populations sampled in Chapter 2. A summary of these findings can be found in the Appendix, Figure A.1). This was based on the observation that the place where a surname is thought to have originated (based on historical records) seems to be where the name is still most common (Lasker, 1985). The geographic distribution of each name was achieved by counting the entries for each name in the British Telecom directory in all counties of England, Wales (source: British Telecom telephone directory in 2002/2003; http://www.bt.com). The present day distribution was compared with that of 100 years ago (estimated from 1901 census data http://www.census.pro.gov.uk/index.html and http://www.scotlandspeople.gov.uk/index.php; Crown copyright material is reproduced with the permission of the Controller of HMSO) to control for recent migrations.

It is noted that telephone directory and Census data is not directly comparable as the British Telecom telephone directory does not list 100% of the British population whilst censuses obviously aim to achieve 100% coverage, thus counts taken from the telephone directory will underestimate the true population size of a surname. This should not create a systematic bias within or among the present surnames, however, and the interest in comparing the two sources is to identify whether or not the distribution of the names has changed, not absolute differences in frequency. Male and female entries were counted in both the telephone directory and census records, despite the fact that only Y-

chromosomes were assayed, as it was not always possible to differentiate men from women. Again, interest in mapping the distribution of the surnames was to identify regions of high and low frequency to locate possible centres of origin and identify the correct British populations to use in comparison, rather than to achieve an accurate count of men in Britain with a particular name.

3.2.5. Geographic Neighbours and the Comparison Dataset

Geographic Neighbours for the present surnames were selected from the comprehensive dataset of 23 British populations analysed in Chapter 2 (Shetland, Orkney, Durness, West Isles, Stonehaven, Pitlochry, Oban, Morpeth, Penrith, Isle of Man, York, Southwell, Uttoxeter, Norfolk, Chippenham, Faversham, Midhurst, Dorchester, Cornwall, Llanidloes, Haverfordwest, Llangefni, Channel Islands; note Rush and Castlerea in Ireland were excluded because the distribution of the names in these regions was not ascertained). Appropriate Geographic Neighbours for each surname were selected on the basis of where the names were most commonly observed today and in 1901, yielding the Geographic Neighbours indicated in Table 3.3. The British populations were combined with the 3 European populations from Chapter 2 (Basques [Bosch *et al.*, 1999, 2001], North German/Danish, and Norwegian, [Capelli *et al.* 2003]) for additional analysis where relevant.

3.2.6. Data Analysis

First, for those surnames with sampled spelling variants an association between spelling variants was tested for using an exact test of population differentiation (Arlequin), as described in Chapter 2 (section 2.2.4), except Arlequin 2.000 was employed (Schneider *et al.* 2000). Haplotype frequencies were employed. Variants that were only represented by one sample (Whittuck, Costen, Corsten, Pharoah, and Ferrer) were not included in the analysis of variants, however, if

Table 3.3. Exact Test of Population Dfiferentiation Calculated For the Surnames Studied and the Comparison Populations

Population	*Bnfr (18)*	*Cstn (55)*	*Flld (5)*	*Frph (50)*	*Mcld (365)*	*Spch (18)*	*Srb (11)*	*Thwt (8)*	*Wt (16)*	*Wht (5)*
Shet	**0.000**	**0.000**	**0.047**	**0.001**	**0.049**	0.094	0.067	0.182	**0.000**	**0.000**
Ork	**0.000**	**0.000**	0.461	**0.000**	**0.000**	0.056	**0.019**	0.506	**0.000**	**0.000**
Dur	**0.000**	0.163	0.786	**0.000**	**0.000**	**0.000**	**0.000**	1.000	**0.000**	**0.000**
Wis	**0.000**	0.008	0.153	**0.000**	**0.002**	0.128	**0.045**	0.173	**0.000**	**0.000**
Sth	**0.000**	0.477	0.474	**0.000**	**0.017**	**0.024**	**0.022**	0.772	**0.000**	**0.000**
Ptl	**0.000**	0.013	0.374	**0.000**	**0.012**	0.140	0.048	0.385	**0.000**	**0.000**
Oban	**0.000**	0.021	0.179	**0.000**	0.322	0.359	0.108	0.667	**0.000**	**0.000**
Mpt	**0.000**	0.063	0.421	**0.000**	**0.001**	0.427	0.162	0.564	**0.000**	**0.000**
Pnt	**0.000**	0.099	0.245	**0.000**	**0.031**	0.565	0.170	0.322	**0.000**	**0.000**
IoM	**0.000**	0.045	0.125	**0.000**	0.228	0.436	0.175	0.156	**0.000**	**0.000**
York	**0.019**	0.830	0.422	**0.004**	**0.000**	**0.034**	0.012	0.407	**0.000**	**0.000**
Sow	**0.000**	0.198	0.564	**0.000**	**0.000**	0.265	0.078	0.509	**0.000**	**0.000**
Utx	**0.000**	0.417	0.519	**0.000**	**0.000**	0.147	0.050	0.734	**0.000**	**0.000**
Ldl	**0.000**	0.683	0.298	**0.000**	**0.000**	0.247	0.159	0.512	**0.000**	**0.000**
Lgf	**0.000**	**0.000**	0.117	**0.000**	**0.001**	0.657	0.280	0.219	**0.000**	**0.000**
Nor	**0.000**	0.586	0.585	**0.000**	**0.000**	0.057	**0.014**	0.485	**0.001**	**0.000**
Hwf	**0.000**	0.000	0.118	**0.000**	**0.022**	0.244	0.110	0.521	**0.000**	**0.000**
Chip	**0.000**	0.263	0.195	**0.000**	**0.013**	0.323	0.206	0.336	**0.000**	**0.000**
Fav	**0.000**	0.239	0.396	**0.000**	**0.007**	0.164	0.150	0.695	**0.000**	**0.000**
Mdh	**0.000**	0.052	0.369	**0.000**	**0.000**	0.556	0.167	0.534	**0.000**	**0.000**
Dcr	**0.000**	0.567	0.519	**0.000**	**0.000**	0.337	0.128	0.658	**0.000**	**0.000**
Pnz	**0.000**	0.135	0.277	**0.000**	**0.440**	0.205	0.133	0.572	**0.000**	**0.000**

Notes: Table shows the p-values for the exact test of population differentiation, significant p-values ($p<0.05$) are shown in bold. Geographic Neighbours for each of the surnames are enclosed by a rectangular box (see text for a definition and description of the Geographic Neighbours). Abbreviations for the comparison populations as in Table 2.8. Surname abbreviations as follows: Bnfr = Barnfather, Cstn = Causton, Frph = Farrer, Flld = Folland, Mcld = MacLeod, Spch = Speechley, Srb = Sorbie, Thwt = Thwaite, Wt = Whittock, Wht = Whytock. Sample sizes for the surnames are given in parentheses.

this analysis showed that the variants were not significantly differentiated, they were reincorporated into the dataset for subsequent analyses.

The null hypothesis to be tested was that each of the surnames was a random draw from the region of Britain where it was most frequently observed; the alternative hypothesis was that each name was instead a non-random collection of Y-chromosomes. The analyses employed below were designed to test the null hypothesis sequentially; only if a name was shown to reject the null hypothesis at the first round of analyses did it proceed to the next round and so on, the rounds of analysis were (i) exact test of population differentiation, (ii) population sampling, (iii) Analysis of Molecular Variance (AMOVA), and (iv) Time to the Most Recent Common Ancestor (TMRCA), which will be described below. For surnames where the null hypothesis was rejected by several rounds of analysis it was possible to infer aspects of their history, such as the evidence for single versus multiple origins, and the introgression of other chromosomes through non-paternity and random adoption of the name.

The first level of analysis investigated whether the overall hg+1 distribution of each surname was significantly different to their Geographic Neighbours, as well as the remaining British comparison populations. The latter comparison aimed to control for inappropriate selection of the Geographic Neighbours. Exact tests of population differentiation were performed as detailed in Chapter 2 (Section 2.2.4), using Arlequin 2.000 (Schneider *et al.* 2000). The hg frequencies used in this test are shown in Table 3.4.

The second analysis was a novel approach developed here which has been termed population sampling. A population sampling method was developed to estimate how different each surname was to its Geographic Neighbours on the basis of frequencies of 3 hgs and their modal haplotype clusters: R1*(xR1a1)/AMH+1, I*(xI1b2)/2.47+1, and R1a1/3.65+1, plus any other hgs that appeared to be elevated in a surname. The 3 hgs and modal clusters comprise the commonest hgs and haplotypes (Wilson *et al.* 2001a; Weale *et al.* 2002; Chapter 2 and Capelli *et al.* 2003) typically observed in British populations, and the surnames, therefore it was pertinent to assess their

Table 3.4. Haplogroups and Modal Haplotypes Encountered in the Surnames Studied

Population \ Hg	*E3b*	*F*(xIJK)*	*J*(xJ2)*	*J2*	*I*(xI1b2)*	*2.47+1*	*I1b2*	*K*(xPN3)*	*N3*	*P*(xR1)*	*R1*(xR1a1)*	*AMH+1*	*R1a1*	*3.65+1*	*n*
Barnfather	-	-	-	-	1	11	-	-	-	-	4	1	1	-	18
Causton	3	-	1	1	5	9	-	-	-	-	16	18	2	-	55
Folland	-	-	-	-	1	-	-	-	-	-	5	2	-	-	8
Farrer	-	-	-	1	18	3	-	-	-	-	8	9	4	6	49
MacLeod	1	-	-	2	14	24	2	-	-	-	60	234	11	17	365
Sorbie	-	-	-	-	-	-	-	-	-	-	-	11	-	-	11
Speechley	-	-	-	-	1	-	-	-	-	-	1	16	-	-	18
Thwaite	-	-	-	-	-	-	-	-	-	-	5	3	-	-	8
Whittock	-	-	-	-	11	-	1	-	-	-	3	1	-	-	16
Whytock	5	-	-	-	-	-	-	-	-	-	-	-	-	-	5

Notes: The surnames Whittock and Whytock, although thought to be spelling variants of the same name, have been listed separately as they are significantly differentiated based on hg+1 frequencies (p=0.000). Spelling variants associated with the names Barnfather, Causton and Farrer are not listed separately here as they have not conclusivley been shown to be significantly differentiated. Haplotypes for all surnames, and associate variants can be found in Appendix Table A.4. Hg+1 frequencies for all of the comparative British populations can be found in Table 2.9. The YCC (2002) hg nomenclature has been used.

frequency in the surnames. The population sampling method employed bootstrap resampling to generate simulation surname populations from the appropriate British populations. Populations were generated for each surname by drawing at random from the identified Geographic Neighbours. The sample sizes of the generated populations were matched to that of the actual surname samples. This was repeated 1000 times with replacement to obtain 95% credible intervals. Population simulations for MacLeod were performed separately using English and Scottish populations as Neighbours, then with England and Scotland combined, because the present day distribution of the name shows it is extremely common in Scotland *and* England, whereas in 1901 it was commonest in Scotland, suggesting that English populations may have contributed Y-chromosomes to the MacLeod gene pool in the last 100 years. Surnames with a hg+1 composition that was significantly different to more than half of their Geographic Neighbours were further investigated.

Thirdly, AMOVAs were performed using hg+1 frequencies and implemented in Arlequin 2.000 (Schneider *et al.* 2000) for surnames with evidence to reject the null hypothesis on the basis of the exact test of population differentiation and population sampling. The populations were clustered in three ways: 1) the surname as one group and the relevant Geographic Neighbours as the second group; 2) the surname as one group and the remaining British comparison populations as the second group; and 3) the surname plus Geographic Neighbours as one group and the remaining British comparison populations as a second group. These 3 clustering methods are depicted graphically in Table 3.7. The method of clustering the data that maximised the amount of among group variation was deemed the best way of grouping the data (see for example Hurles *et al.* 2002). By examining the percentage of variation apportioned among groups for each of the ways of clustering the populations it was possible to estimate the extent to which each surname was similar or different to their Geographic Neighbours. High levels of among group variation for clustering methods (1) and (2) above implies that the surname is differentiated from its Geographic Neighbours as well as the rest of Britain, whilst low levels of among group variation suggest less differentiation. Differences in the percentage of variation apportioned among groups for clustering methods (1) and (2) indicate

that the surname is better clustered with either its Geographic Neighbours or the remaining British comparison populations, depending on how the variation is apportioned. If the surname is different to both its Geographic Neighbours and the British comparison populations, clustering method (3) should reflect this by indicating an increase, relative to clustering methods (1) and (2) of variation apportioned among populations within groups. As Geographic Neighbours for MacLeod comprise 3 discrete groups (all Scottish comparison populations, all English comparison populations, or Scotland and England combined), clustering methods (1) and (2) yield the same groups for testing, and clustering method (3) cannot be assessed using the combined Scottish and English Geographic Neighbours. Therefore out of the 9 potential sets of AMOVA results for MacLeod, only 5 are presented.

Finally the TMRCA was calculated (for surnames with evidence to reject the null hypothesis from the previous analyses) using the average squared distance, or ASD (Goldstein *et al.* 1995b), using the programme Y Time (M Weale, personal communication), where ASD is calculated as:

$$ASD = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{n}\sum_{j=1}^{n}(L_{ij} - L_i^0)^2\right)$$

$L_{ij} - L_i^0$ is the difference in repeat size between each sampled allele and the ancestral allele, respectively, m is number of microsatellites used, and n is number of chromosomes. Using Y Time, the ancestral state is defined by the user, which was assumed to be the modal haplotype. A mutation rate of 0.0028 per generation was used (Kayser *et al.* 2000). The ASD method assumes a simple stepwise mutation model and does not take into account length dependence mutation rates, which may be a factor explaining the inter-locus variation in Y-chromosome microsatellites (Forster *et al.* 2000). However all dating methods available are subject to assumptions about generation time and mutation rate and suffer from large confidence intervals (Hurles and Jobling 2001), rendering all estimates prone to inaccuracies. Therefore the decision was made to assume the simplest model, the SMM. To express TMRCA estimates as years, rather than generations, since the common ancestor, a generation time must be assumed; this can never be estimated with accuracy for past

populations, hence generational intervals of 25 and 35 years were used (this rather large estimate of 35 years has been recently suggested by Helgason *et al.* 2003). If the initial TMRCA estimate was outside the expected timescale, the presence of a modal haplotype was investigated. If a modal type was found TMRCA estimates were recalculated using this haplotype and its one-step microsatellite neighbours.

TMRCA estimates were calculated for the observed haplotypes per hg per surname to evaluate whether the degree of haplotype diversity was compatible with an origin within the surname timescale of approximately the last 1,000 years, approximately. High estimates that place the common ancestor over 1,000 years ago therefore suggest that the hg contains evidence for multiple origins or introgression, regardless of whether a potential founding lineage has been identified. In the instance of high TMRCA estimates and the presence of a modal haplotype(s), the calculation was repeated using only the modal haplotype(s) and one step neighbours. Calculations were performed for all hgs present in each surname, not only those that were at high frequency in population simulations, in case the TMRCA test was more sensitive at identifying potential founding lineages. The TMRCA estimates were thus used to infer different histories for the surnames, particularly the extent of introgression and evidence for multiple founding events.

This hierarchical approach to the analysis of the surnames may be prone to ascertainment bias by excluding surnames without evidence to reject the null hypothesis at the first round. However it was felt necessary to impose some structure to the analysis, but the limited number of published studies on genetics and surnames meant there was not an established framework to follow for such analyses.

3.3. Results

3.3.1. Spelling Variants

It was first appropriate to investigate whether the spelling variants associated with Barnfather, Causton, Farrer, and Whittock were spelling variants or different names with independent origins. The answer to this would affect how these names were treated in subsequent analyses. Results are shown in Table 3.5. This revealed no significant differentiation between the Barnfather variants, which are thus concluded to be spelling variants of the same name rather than different names. The remaining surnames revealed some significant comparisons. The two Whittock variants, Whittock and Whytock are highly differentiated from each other (p=0.000), therefore these names have evidence for independent origins, rather that representing spelling variants, hence are treated hereafter as separate names. Although it should be noted that these two names could still be spelling variants of the same name but that non-paternity has occurred in either the Whittock or Whytock variant. Caston, one of the 8 Causton variants analysed, was significantly different from Corston (p=0.048) and Coston (p=0.047), and the variant Fairer was significantly different from Farrer (0.024) and Farrar (0.032). Interpreting these results is far from straightforward and is considered in more detail in the Discussion. However for the purpose of subsequent analyses, as the significant findings do not extend for comparisons between Corston/Fairer and all other Causton/Farrer variants, all Causton and Farrer variants are considered as the same name.

3.3.2. Surname Population Structure

The exact test of population differentiation shows that Causton, Folland, and Thwaite do not have hg+1 compositions that significantly differentiate them from any of their Geographic Neighbours (Table 3.3) and Sorbie and Speechley were only significantly different to one of their Geographic Neighbours. These 5 surnames therefore appear to be random draws, and do not have evidence to reject the null hypothesis. In contrast, the hg+1 compositions of Barnfather, Farrer, MacLeod, Whittock, and Whytock were significantly different to most or all of their Geographic Neighbours, suggesting that the Y-chromosome hg composition of these names is non-random. To control for potentially selecting

Table 3.5. Exact Test of Population Differentiation Calculated for Several Spelling Variants Based on Haplotype Frequencies

Variant	*Barnfather*	*Banfather*	*Bairnsfather*
Barnfather	-		
Banfather	0.590	-	
Bairnsfather	0.377	0.792	-

Variant	*Caston*	*Cason*	*Causon*	*Cawston*	*Corston*	*Costin*	*Causton*	*Coston*
Caston	-							
Cason	0.154	-						
Causon	0.165	0.454	-					
Cawston	0.221	1.000	0.647	-				
Corston	**0.048**	0.090	0.213	0.529	-			
Costin	0.058	0.568	0.239	0.413	0.059	-		
Causton	0.217	1.000	0.532	1.000	0.283	1.000	-	
Coston	**0.047**	0.241	0.133	0.412	0.060	0.119	0.466	-

Variant	*Farrer*	*Fairer*	*Farrar*	*Farrow*
Farrer	-			
Fairer	**0.024**	-		
Farrar	0.291	**0.032**	-	
Farrow	0.182	0.078	0.089	-

Variant	*Whittock*	*Whytock*
Whittock	-	
Whytock	**0.000**	-

Notes: p-values for the exact test of population differentiation are shown. Significant results ($p<0.05$) are indicated in bold. The haplotypes used for these calculations can be found in Appendix Table A.4. The variants Costen, Corsten, Pharoah and Ferrer were not included in the analysis as the sample sizes were 1.

incorrect Geographic Neighbours, the exact test of population differentiation was also performed using all of the British comparisons and all of the surnames, as indicated in Table 3.3, which confirms the previous conclusions. Therefore, Causton, Folland, Sorbie, Speechley and Thwaite still appear to be random draws, whilst Barnfather, Farrer, MacLeod, Whittock and Whytock have evidence to reject the null hypothesis. The former 5 five surnames were therefore not investigated further, whilst the latter 5 were subject to further analyses.

By employing the population sampling method it was possible to investigate whether particular hg frequencies were higher than expected by chance for Barnfather, Farrer, MacLeod, Whittock and Whytock, allowing the identification of potential founding lineages of Y-chromosomes. One expects the frequency of a founding lineage to be elevated compared to other lineages. The presence of a high frequency modal haplotype within the same haplogroup is treated as indicative of a single common origin, whereas finding chromosomes from the same population (i.e. the same surname) falling into different hgs at appreciable frequencies as evidence of multiple origins (Thomas *et al.* 1998; Behar *et al.* 2003). Although it is the haplotypes within hgs that are explicitly used to confirm common ancestry and would more logically be tested by population sampling, the high rate of microsatellite mutation means that it is less robust to test haplotypes than it is to test hgs. Moreover, a potential founding haplotype that is at high frequency is also anticipated to elevate the frequency of the hg within which it is found.

All 5 surnames investigated had at least one hg or modal haplotype that was outside the simulated 95% credible intervals, as well as the simulated range of values, for their Geographic Neighbours (Table 3.6). It is noteworthy that apart from MacLeod, the hgs at elevated frequency were always hgs normally at low frequency in the comparison dataset, thereby increasing the certainty that their high frequency in the surnames was not fortuitous. I*(xI1b2), which despite being the second most common hg in Britain is rarely seen at frequencies above 30% (Table 2.9), is at very high frequency in Farrer and Whittock. The modal haplotype cluster within I*(xI1b2), 2.47+1, is enriched in Barnfather, and the

Table 3.6. Population Simulations for Several Surnames and Their Geographic Neighbours

Surname[a]	Geographic Neighbours[b]	HG/Modal Types[c]	Observed Count	Simulated Range (Lower)	Simulated Range (Upper)	Confidence Intervals 2.50%	Confidence Intervals 97.50%
Barnfather	Mpt, Utx, York, Pnt, Sow.	R1*(xR1a1)	4	0	11	1	7
		AMH+1	1	2	15	5	13
		I*(xI1b2)	1	0	7	0	5
		2.47+1	**11***	**0**	**6**	**0**	**4**
		R1a1	1	0	3	0	2
		3.65+1	0	0	4	0	2
Farrer	Mpt, Utx, York, Pnt, Sow.	R1*(xR1a1)	8	2	20	5	16
		AMH+1	9	13	33	17	30
		I*(xI1b2)	**18***	**0**	**13**	**1**	**10**
		2.47+1	3	0	11	1	8
		R1a1	4	0	5	0	3
		3.65+1	6	0	6	0	3
MacLeod (Scottish)	Shet, Ork, Dur, Wls, Sth, Oban, Ptl.	R1*(xR1a1)	60	80	103	85	99
		AMH+1	**234***	**157**	**182**	**162**	**178**
		I*(xI1b2)	14	21	37	25	34
		2.47+1	24	17	28	19	27
		R1a1	11	10	20	12	19
		3.65+1	17	18	32	22	31
MacLeod (English)	Mpt, Pnt, IoM, York, Sow, Utx, Chp, Fav, Mdh, Dcr, Pnz, Nor	R1*(xR1a1)	60	55	93	64	87
		AMH+1	**234***	**158**	**203**	**168**	**193**
		I*(xI1b2)	14	18	46	24	40
		2.47+1	24	19	41	23	37
		M17	11	2	22	7	16
		3.65+1	**17***	**2**	**14**	**4**	**12**
MacLeod combined	Scottish and English populations combined	R1*(xR1a1)	60	62	102	69	95
		AMH+1	**234***	**155**	**198**	**161**	**192**
		I*(xI1b2)	14	15	49	23	41
		2.47+1	24	17	40	20	35
		R1a1	11	5	22	7	20
		3.65+1	17	6	26	9	21
Whittock	Chp, Mdh, Dcr.	R1*(xR1a1)	3	0	8	1	7
		AMH+1	1	2	13	4	12
		I*(xI1b2)	**11***	**0**	**5**	**0**	**3**
		2.47+1	0	0	6	0	3
		R1a1	0	0	4	0	2
		3.65+1	0	0	1	0	1
Whytock	Oban, Ptl.	R1*(xR1a1)	0	0	5	0	3
		AMH+1	0	0	5	1	5
		I*(xI1b2)	0	0	2	0	2
		2.47+1	0	0	1	0	1
		R1a1	0	0	1	0	1
		3.65+1	0	0	2	0	1
		E3b	**5***	**0**	**0**	**0**	**0**

Notes: Bold text indicates observed frequencies outside the 95% confidence intervals. Abbreviations as in Table 3.3

[a] Only those surnames with evidence to reject the null hypothesis were included in the population simulation analysis.

[b] Geographic Neighbours were selected from the British populations on the basis of the distribution of each of the names in 1901 and 2002. See text for a full description.

[c] The 3 hgs and their modal haplotype clusters most frequently found in British populations were examined in all surnames, plus any other hgs that appeared to be enriched. The observed count of each of these hgs and modal types is given for each

* Observed frequencies also outside the simulated range.

high frequency of E3b in Whytock is extremely unusual as it is completely absent from the Geographic Neighbours (Oban and Pitlochry in Scotland), as well as the rest of Scotland (Table 2.9). Thus the probability of Whytock having a type other than E3b is low, and the probability of the Geographic Neighbours having E3b types is low. In comparison to the Scottish and English Geographic Neighbours, the frequency of the AMH+1 modal cluster is enriched in MacLeod, whilst the 3.65+1 modal cluster is enriched compared to the English Geographic Neighbours only.

All of these high frequency hgs/modal clusters were therefore examined for high frequency haplotypes. In Barnfather all of the 2.47+1 chromosomes share the same haplotype (Appendix, Table A.4 ht 320) and the 1 non-2.47+1 I*(xI1b2) chromosome is a one step neighbour. 5/21 Farrer I*(xI1b2) chromosomes have the same haplotype (ht 269), and a further 6 of the chromosomes are one-step neighbours. A MacLeod modal type is found in 118/235 AMH+1 chromosomes (ht 65) and out of all R1*(xR1a1) chromosomes a further 82/294 are one step neighbours of ht65. In Whittock 4/11 I*(xI1b2) chromosomes share the same haplotype (ht277) and a further 3 haplotypes are one step neighbours. Finally, all Whytock E3b chromosomes share the same haplotype (ht384).

Networks can be used to visually represent the relationship of microsatellite haplotypes and how closely related they are to each other (see for example Jobling 2001 for a hypothetical example relating to surnames). However it was decided not to use networks in the present work as other analyses such as the ASD were used to describe diversity, hence relationship of the different haplotypes, within the surnames. Each of the haplotypes encountered in the dataset are also listed in Appendix Table A.4.

3.3.3. Multiple or Single Origins and the Extent of Introgression

AMOVA was used to assess whether the level of hg diversity in Barnfather, Farrer, MacLeod, Whittock and Whytock meant they could be clustered with

their Geographic Neighbours and/or the remaining British populations, or not, to identify differences in how these surnames were adopted. These results are summarised in Table 3.7. All surnames had most variation apportioned among groups when they were placed on their own (clustering methods 1 and 2 described in Materials and Methods), suggesting that the 5 surnames that were investigated are best placed on their own, although not all of these groupings were significant. Clustering method (3) yields low apportionment of variation among groups, but typically higher apportionment of variation among populations within groups. MacLeod is the exception to these patterns and displays much lower among group variation (0.17-1.99) compared to the other surnames and the highest among population within group variation was observed with MacLeod in one group and Scotland in the other group, which simply shows that Scotland as a whole represents a diverse range of Y-chromosomes.

TMRCA estimates are summarised in Table 3.8. Of the 5 surnames investigated only Barnfather 2.47+1 chromosomes have a low enough haplotype diversity for the TMRCA estimate to fall within the last 1,000 years, whilst the estimate for Whytock is zero as there is no haplotype diversity because all samples share the same haplotype. For the remaining surnames, the TMRCA was calculated using modal haplotypes (and their one step neighbours) that were identified. These calculations produced TMRCA estimates within the surname timescale for Farrer R1*xR1a1) (which was not at elevated frequency using the population sampling method) and I*(xI1b2) chromosomes, MacLeod R1*(xR1a1), I*(xI1b2), and R1a1 chromosomes and Whittock I*(xI1b2) chromosomes. These are the hgs shown to be at high frequency in the surnames relative to other hgs, using population sampling.

3.4. Discussion

Table 3.7. Genetic Strucutre Between the Surnames and Comparison Populations and Geographic Neighbours*, Assessed by AMOVA

Surname	*Barnfather*			*Farrer*			*MacLeod*	*(Neighbours=Scotland*		*Neighbours=England*	
*Clustering Method***	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*	*2*	*3*
Among Groups (Va)	**23.56**	22.49	0.01	**9.83**	11.34	0.2	1.53	1.52	0.17	1.99	0.5
Among Populations Within Groups (Vb)	1.1	-0.07	20.6	1.28	-0.11	1.92	1.5	2.36	1.9	-0.2	1.71
Within Groups (Vc)	75.35	77.59	97.93	88.89	88.77	97.87	96.97	96.12	97.93	98.22	97.78
Va P (random value >= observed value)	0.03842 +- 0.00000	0.16772+- 0.00368	0.28782 +- 0.00436	0.03673 +- 0.00000	0.16960+- 0.00417	0.28782 +- 0.00436	0.11376 +- 0.00366	0.12515+- 0.00331	0.32653 +- 0.00474	0.16465+- 0.00369	0.11257 +- 0.00280
Vb P (random value >= observed value)	0.00000 +- 0.00000	0.52356+- 0.00493	0.00000 +- 0.00000	0.00000 +- 0.00000	0.52396+- 0.00550	0.00000 +- 0.00000	0.00000 +- 0.00000	0.00277+- 0.00049	0.00000 +- 0.00000	0.79881+- 0.00357	0.00000 +- 0.00000
Vc P (random value <= observed value)	0.00000 +- 0.00000	0.00020+- 0.00014	0.00000 +- 0.00000	0.00000 +- 0.00000	0.00000+- 0.00000	0.00000 +- 0.00000	0.00000 +- 0.00000	0.00000+- 0.00000	0.00000 +- 0.00000	0.00604+- 0.00074	0.00000 +- 0.00000

Continued on following page & legend on following page

Table 3.7 continued

Surname	*Whittock*			*Whytock*		
Clustering Method	*1*	*2*	*3*	*1*	*2*	*3*
Among Groups (Va)	**26.37**	28.47	-0.29	**44.59**	53.25	-0.27
Among Populations Within Groups (Vb)	1.06	-0.41	2.16	0.8	-0.69	1.93
Within Groups (Vc)	72.58	71.94	98.13	54.61	47.44	98.34
Va P (random value >= observed value)	0.03505 +- 0.00205	0.24782+- 0.00000	0.3505 +- 0.00205	0.03822 +- 0.00171	0.33475+- 0.00000	0.56010 +- 0.00528
Vb P (random value >= observed value)	0.00000 +- 0.00000	0.77366+- 0.00405	0.00000 +- 0.00000	0.00000 +- 0.00000	0.78020+- 0.00377	0.00000 +- 0.00000
Vc P (random value <= observed value)	0.00000 +- 0.00000	0.00000+- 0.00000	0.00000 +- 0.00000	0.00000 +- 0.00000	0.00010+- 0.00010	0.00000 +- 0.00000

Notes: Bold text indicates p values <= 0.05 for Among Group comparisons, showing which surnames are significantly structured when compared to the British comparison populations.

* See text for a definition and discussion of Geographic Neighbours.

** Clustering methods as follows: Method 1 places only the surname in Group 1 and the entire British comparison dataset in Group 2; Method 2 places the surname in Group 1 and the Neighbours in Group 2; Method 3 places the surname and Neighbours in Group 1 and the entire British comparison dataset minus the Neighbours in group 2. This is depicted in the figure to the right

Graphic Representation of the Three Clustering Methods Employed in the AMOVA Analysis

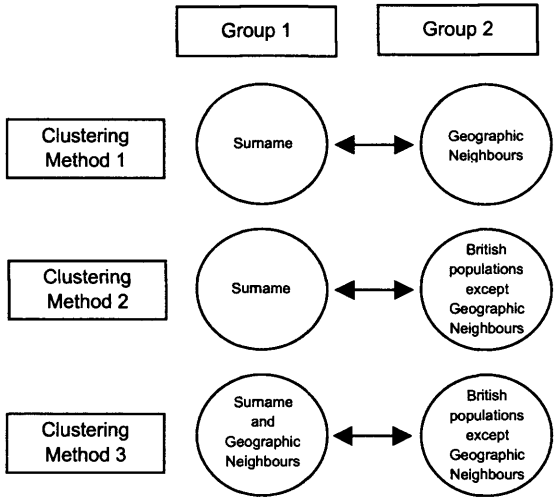


Table 3.8. TMRCA Estimates Calculated Using ASD for Surnames with Evidence of Non-Random Adoption

		Generation Time	
Surname	*HG*	*25yrs*	*35yrs*
Barnfather	R1*(xR1a1)	**	**
	I*(xI1b2)	**124.11**	**173.75**
Farrer	R1*(xR1a1)	2,626	3,676
	R1*(xR1a1) modal+1*	**892.86**	**1,250**
	I*(xI1b2)	7,228	10,119
	I*(xI1b2) modal+1	**811.61**	**1,136**
	R1a1	2,381	3,334
MacLeod	R1*(xR1a1)	1,883	2,636
	R1*(xR1a1) modal+1	**617.86**	**865.00**
	I*(xI1b2)	4,895	6,853
	I*(xI1b2) modal+1	**682.14**	**955.00**
	R1a1	3,933	5,506
	R1a1 modal+1	**612.50**	**857.50**
Whittock	R1*(xR1a1)	3,348	4,688
	I*(xI1b2)	2,435	3,409
	I*(xI1b2) modal+1	**637.50**	**892.50**
Whytock	**E3b**	**0**	**0**

Bold text indicates TMRCA estimates within a plausible historical timescale for surname establishment.

* Farrer R1xR1a1 chromosomes are bimodal (see Appendix, Table A.4), the same TMRCA estimates are calculated regardless of which modal type is used.

** No modal haplotype is present therefore an ancestral haplotype has not been inferred and TMRCA estimates were not calculated.

The aim of this chapter was to study the fidelity of surname inheritance using the Y-chromosome. The presence of the dataset of British Y-chromosomes (Chapter 2) was essential in providing regionally specific control populations for each of the surnames. In summary (Table 3.9) the combination of Y-linked microsatellites and UEPs employed has allowed the identification of those surnames which (with caveats) appear to be random draws from the British population (Causton, Folland, Sorbie, Speechley and Thwaite) and those that contradict the null hypothesis of random adoption (Barnfather, Farrer, MacLeod, Whittock and Whytock). Within this latter group it was first suggested that Whittock and Whytock could indeed be different surnames with independent origins, rather than spelling variants of the same name, which contradicted historical expectations (Reaney 1997). Although as previously stated it is possible that there was a fortuitous case of non-paternity associated with one of the spelling variants which has created a false signal of independent origins. Furthermore, for the 5 names with non-random origins, it was possible to distinguish evidence for a predominantly single origin, with little or no introgression (Barnfather and Whytock), from evidence of multiple origins and introgression (Farrer, MacLeod and Whittock). Further analysis of some of the surnames is required to clarify or confirm current conclusions. These results will now be discussed in more detail below. Consideration will also be given to differences in pattern of inheritance between rare and common surnames and the type of surname, and the extent to which these results can be applied more generally to other British surnames.

Concluding that Causton, Folland, Sorbie, Speechley and Thwaite are random assortments of chromosomes was based upon the finding that these surnames are not significantly differentiated from any (or many) of their Geographic Neighbours or the rest of Britain. There are however some caveats. Folland, Sorbie, Speechley and Thwaite have relatively small sample sizes (n = 8, 11, 18, 8, respectively), which may lower the power of the exact test and show little differentiation between the surnames and the rest of the British populations.

Moreover, either most or all of the Y-chromosomes in Folland, Sorbie, Speechley and Thwaite belong to the common British hg R1*(xR1a1) (Wilson *et*

Table 3.9. Summary of Results Used to Determine Which Surnames Are Random Draws of Y-Chromosomes

	Do the Four Tests Reject (yes) or Support (x) the Null Hypothesis?			
Surname	*Exact test*	*Population Sampling*	*AMOVA*	*TMRCA*
Whytock	yes	yes	yes	yes
Barnfather	yes	yes	yes	yes
Whittock	yes	yes	yes	yes[1]
Farrer	yes	yes	yes	yes[1]
MacLeod	yes	yes	x	yes[1]
Causton	x	N/A	N/A	N/A
Folland	x	N/A	N/A	N/A
Sorbie	x	N/A	N/A	N/A
Speechley	x	N/A	N/A	N/A
Thwaite	x	N/A	N/A	N/A

Notes: "yes" indicates that the surname had evidence to reject the null hypothesis, and "x" indicates that there was not evidence to reject the null hypothesis (see Section 3.2.6) for a given test. The surnames are ranked from those with most evidence to reject the null hypothesis (top) to those with most evidence to support the null hypothesis (bottom). The 4 tests summarised here were performed sequentially (see Section 3.2.6), therefore as the surnames Causton-Thwaite had no evidence to reject the null hypothesis with the first round of tests (exact test of population differentiation) no other tests were performed.

[1] These 3 surnames only had a TMRCA estimate that rejects the null hypothesis when the estimate was calculated using the modal haplotype+one step neighbours (see Section 3.2.6)

al. 2001a; Weale *et al.* 2002; Chapter 2; Capelli *et al.* 2003) which may confound attempts to statistically differentiate the surnames from other British populations. Whilst the present data for these surnames *do* have statistical support for the conclusion that they are random draws, the matter could be clarified by (i) larger sample sizes and (ii) increasing haplotype resolution, particularly for R1*(xR1a1) chromosomes, by assaying more Y-linked microsatellites.

Analysis of Causton is made complex by the presence of spelling variants, one of which (Caston) is significantly different from two of the other variants (Cortson and Coston). This leads to the potential conclusion that the reason Causton appears to be a random assortment of Y-chromosomes is that it is a heterogenous collection of different surnames, some or all of which were founded by men with different Y-chromosome types. Such a conclusion is in keeping with the notion that at least some of the spelling variants actually represent different surnames (Reaney 1997). However, if the current data conclusively supported the notion that Caston was a different surname one would expect Caston to be significantly different from all of the variants, not just Coston and Cortson, or for Coston and Corston to be significantly different from all other variants. As this was not the case, interpretations are certainly complex and further sampling of the different variants seems to be the only way of forming more decisive conclusions. For these reasons it is not possible to reject the null hypothesis for Causton. but it is observed that currently the origins of Causton are unresolved. It is therefore apparent that it is difficult to irrefutably show whether or not a surname is simply a random sample of Y-chromosomes as many factors could create the random appearance of the data, related to the distribution of Y-chromosome types in Britain, limited written records, and issues of sample size.

In contrast, concluding that the remaining 5 surnames, Barnfather, Farrer, MacLeod, Whittock and Whytock have been adopted non-randomly is more certain. When compared to their Geographic Neighbours these surnames have a hg+1 composition that is significantly different to most or all of the Neighbours and most or all of the comparison populations. Additionally the observed

frequencies of hgs/modal clusters (that also contain high frequency haplotypes, i.e. potential founding lineages) in these names are greater than would be expected by chance. This latter analysis is aided by the fact that, apart from MacLeod, the high frequency hgs happen to not be the extremely common R1*(xR1a1) or its modal cluster AMH+1, and in the case of MacLeod the observed frequency of AMH+1 is so elevated compared to their Neighbours that it is highly likely to contain a founding lineage. Apart from MacLeod, Population Sampling only suggests that one hg/modal cluster is identified as being enriched in each of the surnames, suggesting single origins (compared to their English Geographic Neighbours MacLeod have two hgs at high frequency, possibly suggesting two origins).

The total Y-chromosome profile of Barnfather and Whytock, as well as the diversity within the modal hgs strongly suggests that these names have had a single origin, with little or no introgression. Within the modal hgs for these two names, the haplotypes are closely related, as the TMRCA estimates are very recent (the Whytock TMRCA is zero because all of the haplotypes are identical). This confirms that all of the I*(xI1b2) Barnfather Y-chromosomes probably share the same common ancestor, although convergence of microsatellite haplotypes might also mean that unrelated men have the same haplotype. Note that although the Barnfather modal haplotype (ht320) belongs to the common 2.47+1 modal cluster, ht320 is very rare, therefore it is unlikely that the same haplotype is shared by so many men by chance. The presence of non-I*(xI1b2) chromosomes in Barnfather, does however mean that some introgression has been experienced and the present data do not suggest a second founder. In contrast, all Whytock Y-chromosomes belong to the same hg and have the same haplotype, therefore there is no evidence for any introgression. Despite the Whytock sample size being small, the absence of E3b chromosomes in Whytock's Geographic Neighbours (Oban and Pitlochry in Scotland) or indeed in Scotland means it is unlikely that 5 men called Whytock share the same surname by chance. As the exact tests show (Table 3.3) Whytock is indeed highly differentiated from its Geographic Neighbours and the remaining comparison populations. E3b chromosomes are found in Britain, albeit at low frequency (Chapter 2, Capelli *et al.* 2003), but are much more common in

Mediterranean and African regions (Cruciani *et al.* 2004) it is therefore possible that the founder may have migrated to Scotland from Britain or further afield. The degree of Y-chromosome homogeneity in Barnfather and Whytock, and the high frequency of typically rare British Y-chromosome types, compared to their Geographic Neighbours and the rest of Britain is reflected in the high levels of variation apportioned among groups (23.56% and 44.59% respectively), particularly for Whytock.

Whittock also has good evidence for a single origin, although the name appears to have experienced more introgression than either Barnfather or Whytock, which is most apparent in the TMRCA estimates and the observed distribution of hgs and haplotypes. For example, the TMRCA estimate calculated for all I*(xI1b2) chromosomes leads to a high estimate, which only falls within a plausible timescale for surname history when calculated for the modal haplotype and its one step neighbours. The Whittock modal haplotype is rare elsewhere in Britain (see Appendix, Table A.3), therefore it is likely that this represents the Whittock founding lineage, rather than their presence in the sample by chance. Although the modal haplotype and one step neighbours comprise 7/16 of all Whittock chromosomes, the remaining chromosomes are unrelated (see Appendix, Table A.4). It is difficult to assess whether there is a Whittock founding R1*(xR1a1) found lineage due to the small numbers involved.

Finally, Farrer and MacLeod show most evidence for multiple origins and introgression as they display the lowest levels of among group variation for clustering method 2 and high TMRCA estimates. In the case of Farrer it is possible that the presence of variants may confound these conclusions. The Fairer variant may be a different surname with independent origins to the other Farrer variants, suggested by the significant difference in haplotype frequencies between Fairer and Farrer, and Fairer and Farrar. As for the Causton variants discussed above however, the present results are inconclusive because Fairer is not significantly different to all of the variants. Therefore whilst the possibility that Fairer is a different surname should be investigated further, it was not deemed to be conclusive enough to separate the Farrer variants. Documentary sources are not clear about whether the variants are spelling variants or different

names (Reaney 1997). With hindsight therefore it is perhaps apparent that it would have been simpler to study names without spelling variants as this has lead to some complication with analyses. However, without the clarity of hindsight, the motivation to study spelling variants was justified by the fact that the existence or otherwise of spelling variants is an interesting question in the study of surnames.

The potential Farrer founding lineage (ht269) is found on a background of hg I*(xI1b2), which is clearly enriched in the Farrer sample. However the modal haplotype is still at quite low frequency in the whole Farrer sample compared to other surnames with clearer evidence for single origins such as Barnfather, Whittock and Whytock. Furthermore the considerable haplotype diversity in the I*(xI1b2) confirms that even the modal haplogroup is not composed of exclusively the founding lineage. A possible second Farrer founding lineage is observed in R1*(xR1a1) where either of the bimodal haplotypes, which are one-step neighbours, may represent another founder. Note however that neither the frequency of R1*(xR1a1) nor AMH+1 in Farrer is elevated compared to their Geographic Neighbours, but TMRCA estimates using the modal haplotype plus one-step neighbours lead to plausible estimates.

Despite the modal MacLeod haplotype being part of the AMH+1 cluster in R1*(xR1a1), the haplotype is relatively rare in the British populations (Appendix, Table A.3), therefore it is likely that the MacLeods who share the haplotype also share a common ancestor. Diversity within R1*(xR1a1) chromosomes is relatively high however. Of all the surnames studied here MacLeod has the best documented history, which states that all MacLeods descend from Leod who was the son of the Norse King of Man and the Hebrides (Olaf the Black), and is believed to be progenitor the Clan MacLeod (Morrison 1986; Dorward 2000). Although the modal haplotype is not particularly common in the Norwegian or North German/Danish populations studied in this thesis (see Appendix, Table A.3), thereby reducing the chance that the MacLeod progenitor was from one of these countries, it is quite common in Shetland and the Western Isles. The clan's progenitor could have feasibly originated in one of these locations given that the Mac- prefix is a Scottish prefix meaning 'son of'.

Furthermore it is extremely interesting that TMRCA estimates for the modal haplotype and its one step neighbours (617.86-865 years ago) yields a date that approximately matches the date that clan history states the MacLeod lineage was founded (Leod was born ~800 years ago in 1200AD) (Morrison 1986). This straightforward conclusion is somewhat complicated by the fact that the modal haplotype is found at highest frequency in Llangefni, however.

It is possible that sample size has influenced the above conclusions about the degree of introgression in the surnames Barnfather, Farrer, MacLeod, Whittock and Whytock. The sample sizes of Farrer and MacLeod are larger than those of Barnfather, Whittock and Whytock, hence increasing the chance of more variation being assayed. One could argue that the former two names have more evidence for introgression than the latter 3 as a result of the larger number of sampled chromosomes. However, this finding cannot be interpreted only in the context of sample sizes for two reasons: 1) population sampling addresses this issue and shows that all 5 surnames have at least one hg that is at an unusually high frequency, i.e. the distribution of Y-chromosome types in the surnames is non-random; 2) the surname sample sizes collected for this study reflect the frequency of the surnames in Britain, therefore rarer names have small sample sizes and commoner names have larger sample sizes. Hence, it is more correct to conclude that the evidence suggests that rarer names tend to have a single origin, whereas more common names have had multiple origins, or more instances of introgression.

The incidence of non-paternity was not explicitly tested in this Chapter, therefore non-paternity cannot be separated from other factors affecting the amount of introgression. It is noteworthy, however, that apart from Whytock, all surnames have at least one chromosome present that may be the result of non-paternity or some other introgression event, such as the random adoption of the surname. This is even the case for those surnames that are clearly not random draws from the British Y-chromosome gene pool. Estimates of non-paternity rates in the literature cover a broad spectrum (1.5-30%, Cerda-Flores *et al.* 1999), and no formalised research has been carried out on how many people change their name each year. However it seems that around 300,000 cases of

name changes occur in Britain per year. Most of these changes are however thought to be women changing their name through marriage (UK Deed Poll Office, personal communication 2003). Therefore, based on figures in the published literature it is not possible to assess whether non-paternity or random surname adoption is the most likely cause for finding Y-chromosomes that cannot be related to the founding lineage. This question could be addressed by explicitly sampling from men with the same surname who believe they are related.

A final question that can be addressed with the present surnames is whether different types of surnames (local surnames, surnames of relationship and office, and nicknames) have different patterns of inheritance. The surnames that are thought to be local, i.e. derived from a place name (several Causton variants, Sorbie, and Thwaite) appear to be random draws from British populations, albeit with the caveats discussed above, suggesting that local origin names tend to have multiple origins. Surnames of relationship and nicknames, which are considered together because of overlap in assigning names to these two categories, appear in contrast to have single origins: Barnfather, MacLeod, Whittock and Whytock. Farrer represents the only surname of occupation and the name appears to have one or possibly two origins; the etymology derives Farrer from the Old French *ferreor, ferour* meaning a worker in iron, akin to the British surname *Smith*. Given the ubiquity of smiths it is at first surprising that Farrer does not seem to have evidence for multiple origins. However as the name has possible French origins it is possible that only one or two men bearing the surname migrated from France and were the source of all subsequent Farrers in Britain.

3.5. Conclusions

In conclusion, the results of the present study validate the initial findings of Sykes and Irven (2000) that the Y-chromosome can be used to infer aspects of surname history. It is still apparent however that even with the increased number

of Y-chromosome makers employed here further resolution is still required. There seems to be some generalisations that can be made from the data. (i) Of those surnames with evidence for non-random adoption the rarer names have better evidence for single origins, however rare names cannot be concluded to have single origins *per se*. (ii) In contrast local surnames appear to be random draws from the British population, implying that local names have typically been founded in many different locations across Britain, although this conclusion might be modified for local surnames that derive from a very localised dialect. Finally, it is also evident that even in the absence of well-kept historical records, such as those available for the study of Tristan da Cunha (Soodyall *et al.* 2003), it was still possible to make inferences about the history of the surnames, which is important given the paucity of available records for much of the period of interest.

Chapter 4. Y-Chromosome and mtDNA Diversity in Present Day Inhabitants of London

4.1. Introduction

Most studies of genetic history concentrate on sampling from small towns and villages to make inferences about past historical processes (see for example Cavalli-Sforza *et al.* 1994, Richards *et al.* 1996; Weale *et al.* 2002; Chapter 2 and Capelli *et al.* 2003)), such as the sampling strategy employed in Chapter 2. The assumption is that cities contain high levels of genetic diversity which will obscure past events (Cavalli-Sforza *et al.* 1994), and are therefore ignored in sampling strategies. To the author's knowledge however, there has not been a study to specifically address the genetic history of a city within a genetically well-characterised country or region, such as Britain. In particular the composition of Y-chromosomes from many *rural* British populations is particularly well known (Chapter 2 and Capelli *et al.* 2003), providing an excellent comparative dataset against which to compared the Y-chromosomes of Londoners. Like many other capital cities, London has been, and still is, a centre for the immigration of people from around Britain, Europe, and the rest of the world to the extent London has often been dependent on migrants for its prosperity (Inwood 1998). This creates a wealth of history and a diverse ethnic and social background to life in London. As the aim of this chapter was to examine the genetic diversity of London, the following paragraphs will review the history of immigration to London, within the context of the British history.

4.1.1. A Brief History of London

London, or at least the area known as London today, seems to have been settled from the Neolithic period (10,000 years ago) onwards, initially by people indigenous to Britain. For example a site at Uxbridge in north-west London has finds dated from around 9,000-7,000 years ago which suggest that it was a butchery site (Cotton and White 1998). The first mass immigrant presence in London however, the invading Roman armies of the 1st century AD, are more famously considered as the first inhabitants of London (for example Inwood 1998 p.1). It is during their second invasion of Britain that the Romans settled in

London, taking advantage of its good strategic location and commercial advantages over other locations, thus London became the centre of administration of Roman Britain and started to grow and flourish. In 410AD the Romans withdrew from Britain and in the same century Saxons started to take power in Britain; by the middle of the 6th century London was under Saxon rule. Three centuries later Viking invasions were seen across Britain (described in more detail in Chapter 2) and in London. The final large scale invasion of Britain was by the Normans in 1066; again London was captured, and subsequently ruled by Normans for 300 years (Ackroyd 2000). Therefore, even prior to the 19th century which saw an "open door" policy to immigrants (Kershen 1997) and the post World War Two period, which has been regarded as *the* age of immigration (McAuley 1993), London, like the rest of Britain, has been subject to large numbers of immigrants from across Europe, and possibly further afield as Roman armies were known to use people from Africa as slaves (McAuley 1993).

The important point to remember is that many regions within Britain were affected by these invasions, not only London, hence all of these various immigration (or maybe more precisely invasion) events had the potential to affect the gene pool of London as well as the rest of Britain. However in the context of British history, London has had a unique history of immigration due to a combination of factors: its role as a capital city and the concomitant status associated with this, its history of being a busy port (Holmes 1997), and its location today close to many large international ports and airports making it literally the first port of call for many people arriving in Britain. The more detailed written records of the last ~ 800 years show that London received many immigrants from across Britain and further afield as the city grew increasingly cosmopolitan, indeed London has become dependent on migrants for its continued prosperity (Inwood 1998), particularly in trades that Londoners were (and are) not willing pursue themselves (Ackroyd 2000). During the 12th century London housed merchants from Brabant, Rouen, and Ponthieu on the Thames waterfront, and their numbers were so great that the Thames was repeatedly reclaimed and the banks extended to house the migrants. The first evidence for a distinct Jewish district in London appeared during the 12th century (Ackroyd

2000) and today there are around 50,000 self-designated Jews in London (2001 Census, Source: National Statistics website: www.statistics.gov.uk). During the 16th century it is estimated that around 1/6 of all Englishmen had migrated to or were resident in London, and a large proportion of the city was occupied by immigrants from England and abroad (Ackroyd 2000). 16th century London also saw Huguenot refugees arriving who were fleeing Catholic persecution.

Over the centuries different trades often became associated with discrete groups. 19th century bakers were from Scotland, shoemakers from Southampton, and many sugar refiners were German (Ackroyd 2000). Certain areas in London have now acquired distinctive ethnic characteristics. An Italian Quarter emerged in the areas of Clerkenwell and Holborn in central London (McAulay 1993) and this is still evident in these areas by the conglomeration of Italian delicatessens and the annual "Italian Procession" in honour of Our Lady of Mount Carmel. The East End, particularly around Brick Lane ("Banglatown"), has a strong Bangladeshi community and is famous for its collection of Bangladeshi restaurants, and Brixton in south London has a large African and Caribbean population. The present ethnic diversity in London can be quantified in the 2001 Census records (see also Table 4.1) which show that 40% of Londoners are not classified as British (using self designation of ethnicity). In this collection of "non-British" Londoners, 15 different ethnic groups are found, with overall frequencies within the London population that range from 8.29% (other [i.e. non-British] White) to 0.476% (White and Black African). In terms of recent trends in international immigration to the UK, London consistently receives the single largest proportion of migrants, usually around 30%; apart from the South East no other region reaches double figures (Dobson and McLaughlan 2001). In contrast it is interesting to note that when movement between different British regions is considered, London actually makes a net loss of people each year (although when this is balanced with international migration flows, London makes a net gain in population numbers) (Vickers 1998).

Is it possible to estimate the impact Romans, Saxons, Vikings, and Normans had on London's gene pool? History and archaeology cannot produce absolute figures, although it can provide some clues. For example, based on skeletal

Table 4.1. Frequency of Self Defined Ethnic Group from 2001 Census Records and Populations Analysed in this Study

Self-Defined Ethnic Group	*Greater London**	*Liverpool**	*Present London Sample*
British	0.598	0.918	0.785
Irish	0.031	0.012	0.027
Other White	0.083	0.013	0.137
White and Black Caribbean	0.010	0.005	0.005
White and Black African	0.005	0.005	0
White and Asian	0.008	0.003	0.005
Other Mixed	0.009	0.005	0.014
Indian	0.061	0.004	0.005
Pakistani	0.020	0.002	0
Bangladeshi	0.021	0.001	0
Other Asian	0.019	0.003	0.005
Black African	0.053	0.007	0.005
Black Caribbean	0.048	0.002	0.014
Other Black	0.027	0.003	0
Chinese	0.038	0.012	0
Other Ethnic Group	0.016	0.004	0

*Source: National Statistics website: www.statistics.gov.uk

remains, archaeologists have concluded that most Londoners during the Roman period were actually indigenous Britons (Hall and Conheeney 1998). There are many variables that will affect whether such events are evident genetically: the degree to which the incoming population replaced the existing population; the size of the incoming population; the amount of gene flow between these populations; and the extent to which the incoming and existing populations were genetically differentiated. These points will be further considered in the Discussion (section 4.4). However to augment what is known from the historical and archaeological sources genetic can be a useful resource. Due to the unique features of the Y-chromosome and the mitochondrial genome, such as their uniparental mode of inheritance and lack of recombination (see section 2.1.2 and 4.1.2 below) much is known about the geographic distribution of Y-chromosome mutations (Jobling and Tyler-Smith 2003) and mtDNA sequence motifs (Richards *et al.* 2002), making them the well suited to questions relating to population history. European Y-chromosomes have been particularly well characterised in recent years, as described in more detail in Chapter 2 (Section 2.1.5). This allows the genetic investigation of Londoners to be placed within a relatively well characterised framework. The results presented in Chapter 2 also provide an important comparative dataset. British mtDNA diversity has not been studied in as much detail, however European populations are amongst the best defined for mtDNA variation and the available data provide a useful starting point for the analyses performed here. Before proceeding to present a summary of these findings, an overview of mtDNA will be presented.

4.1.2. mtDNA – An Overview

Unlike the Y-chromosome which has only recently been (almost) fully sequenced (Skaletsky *et al.* 2003), the entire sequence of the mitochondrial genome has been known for over 20 years in what has become known as the Cambridge Reference Sequence, or CRS (Anderson *et al.* 1981). The mitochondrial genome (Figure 4.1) is circular and around 16,569bp in length

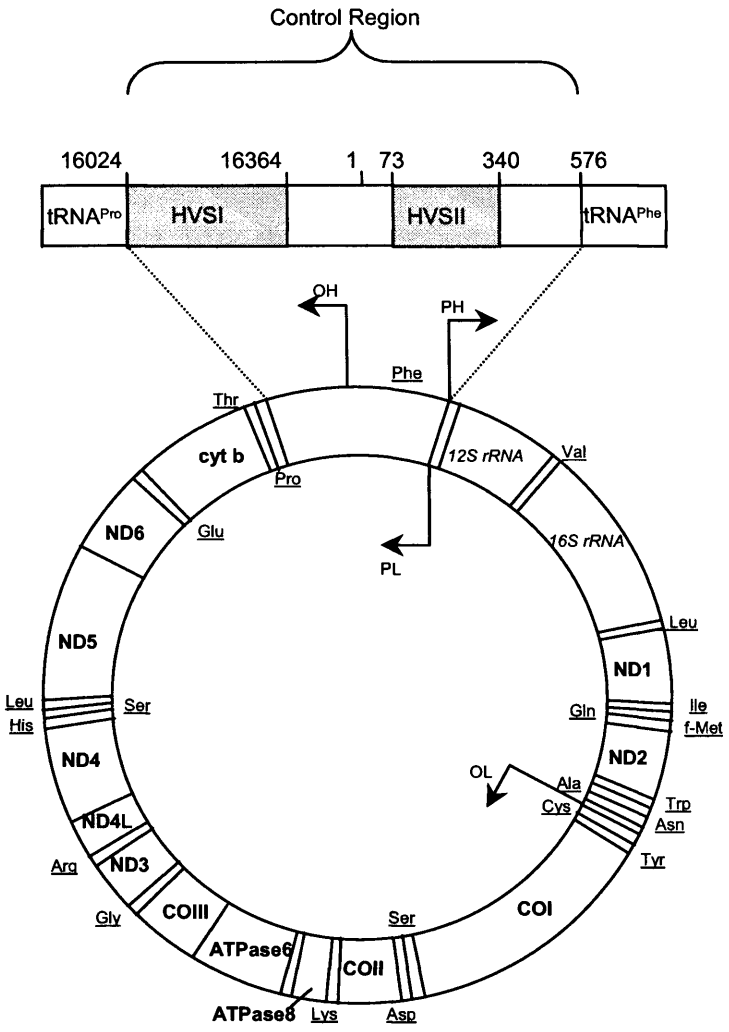


Figure 4.1. The Human Mitochondrial Genome. The mitochondrial genome is circular and around 16,569 bp in length. HVSI and HVSII are shown at the top of the diagram. tRNA genes are indicated by underline, protein coding genes are in bold, and the two rRNA genes are in italics. The origins of replication of the light and heavy strands (OL, OH) and the promoters for transcription for the two strands are shown (PL, PH). *Figure modified from Jobling et al. (2003) and Strachan and Read (1999).*

(Anderson *et al.* 1981, Andrews *et al.* 1999). It is a double stranded molecule composed of a heavy and light strand. Mitochondria generate energy by oxidative phosphorylation. There are two sections to the mitochondrial genome, the coding region, which comprises around 93% of the sequence, and the remaining 7%, which is non-coding (also known as the control region) (Strachan and Read 1999). The mutation rate of the whole mitochondrial genome is around 5-10 times higher than for the nuclear geneome (Brown *et al.* 1979; Budowle *et al.* 2003), hence providing a large array of polymorphisms available for population studies. Such high levels of mutation are generally thought to be generated during oxidative phosphorylation, and these mutations are allowed to accumulate because histones, and the highly efficient DNA repair mechanisms seen in the nuclear genome, are not present in the mitochondrial genome (Fliss *et al.* 2000). There are a high copy number of mtDNA molecules in each cell (e.g. Budowle *et al.* 2003), unlike the nuclear genome where only two copies are present in somatic cells and one in the germ-line. This can result in heteroplasmy, the presence of more than one (mtDNA) type in the same individual, however as this phenomenon is rarely observed in the germ-line (Awadalla, 2003) it does not present a problem for population studies.

mtDNA has a long track record of use in human population studies with the level of resolution employed continually increasing. The earliest study of mtDNA variation in human populations was published in 1981 by Denaro and colleagues, followed by Johnson and colleagues in 1983 (reviewed by Cavalli-Sforza *et al.* 1994), who employed Southern blots to assay RFLPs. Since these early studies the number of individuals, populations, and markers has increased, as has the ability to differentiate populations and detect more polymorphisms. Today most studies of human populations use a combination of RFLP and sequence based assays; RFLPs assay around 20% of the entire mitochondrial genome whilst most sequence-based analyses assay two regions (hypervariable segments I and II; HVSI and HVSII respectively, also know as HVRI and II) within the non-coding or control region. More recently whole genome sequencing has been employed (Ingman *et al.* 2000; Finnilä *et al.* 2001; Torroni *et al.* 2001a; Maca-Meyer *et al.* 2003).

The mutation rate in HVSI and II is higher than the rest of the mitochondrial genome, as the name might suggest, with some bases in particular appearing to be mutational hotspots (e.g. Heyer *et al.* 2001). Indeed several sites within HVSI are known to have back-mutated in the human phylogeny, thus clearly challenging the infinite alleles model which is a fundamental tenet of many analytical models (see for example Richards *et al.* 2000). Recently an average rate of 1 transition per 20,180 years was used in the literature (Richards *et al.* 2000). Due to the high rate of mutation of HVSI and II sequences RFLPs are usually preferred for assigning sequences to hgs (e.g. Torroni *et al.* 1996; Finnilä *et al.* 2001); the RFLPs that define lineages in human populations can be found in Table 4.2, and a description of the geographic distribution of these lineages in populations pertinent to this thesis are found in Section 1.4 below. HVSI and II sequence variation is then used to further subdivide the hgs as diagnostic hg-specific sequence motifs can often be found (Graven *et al.* 1995, and see Figure 4.2 and Table 4.2)

A lack of geographic resolution has been found for mtDNA compared to the Y-chromosome, and is a problem that has somewhat hindered the value of mtDNA analysis in human population studies, a fact acknowledged by recognized proponents of mtDNA research (Richards and Macaulay 2000). The lack of resolution is due in part to the higher mutation rate of the HVSI region and its use in constructing phylogenetic relationships of mtDNA lineages (Richards and Macaulay 2001) as the degree of homoplasy leads to considerable reticulation hence many trees are equally parsimonious. (Finnilä *et al.* 2001). At the other extreme, many of the slower mutating RFLP sites do not contain enough information for detailed phylogenetic resolution (Richards and Macaulay 2000; Finnilä *et al.* 2001). However, the new approach of sequencing the entire mitochondrial genome of individuals known to belong to particular RFLP-defined hgs (Finnilä *et al.* 2001; Torroni *et al.* 2001a; Maca-Meyer *et al.* 2003) will likely help by clarifying and refining phylogenetic and geographical relationships of mtDNA lineages.

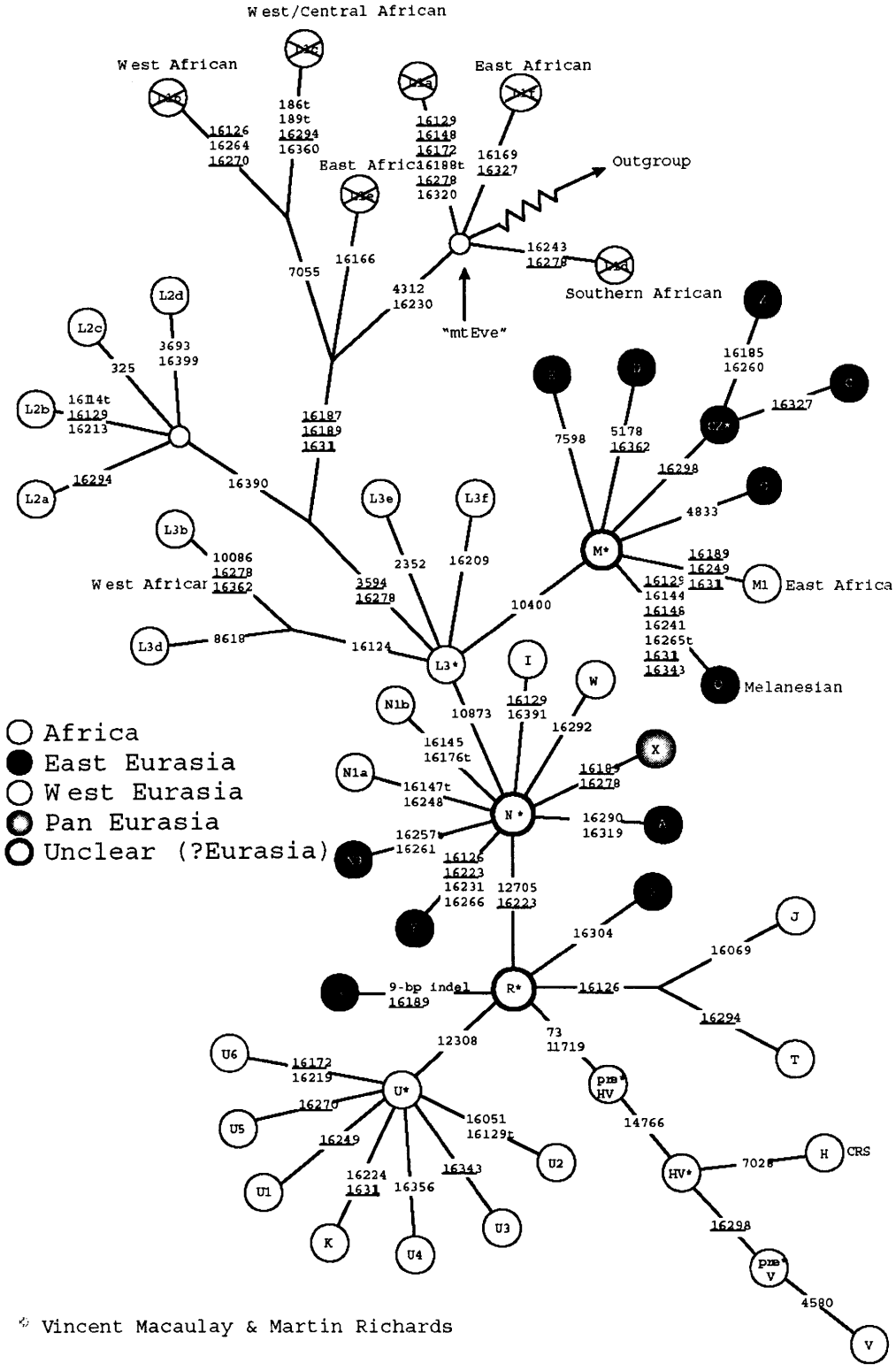


Figure 4.2. Skeleton Network of the mtDNA Phylogeny. Some diagnostic sequence motifs are shown (see Table 4.2 for more complete RFLP and HVSI sequence motifs), transversions are indicated by the –t postfix and recurrent mutations by an underline. L1 hgs have been struck through because the previous incorrect rooting of L1 and the conclusion that it was a clade means that those hgs labelled L1a-L1f should be re-named, however, they are still shown to retain continuity. *Figure taken from Richards and Macaulay (2000).*

Table 4.2 mtDNA RFLP Sites and HVSI Sequence Motifs Used to Assign mtDNA Sequences to Haplogroups

Haplogroup	HVS-I motif	73 status	Coding-region mutations
L1a *L1a1a* *L1a2*	148 172 187 188G 189 223 230 311 320 *278* *129*	A	+3592*Hpa* I
L1b	126 187 189 223 264 270 278 311	G	+3592*Hpa* I
L1c *L1c1* *L1c2* *L1c3*	129 187 189 223 278 294 311 360 *274* *265c 286g* *187 218*	G	+3592*Hpa* I
L1e	129 148 166 187 189 223 278 311	G	+3592*Hpa* I
L1f	169 187 189 223 230 278 311 327	G	+3592*Hpa* I
L2 *L2a* *L2a1* *L2a1b* *L2c2*	223 278 *390* *294* *309* *290* *264*	G	+3592*Hpa* I
L3	223	G	
>L3b *L3b1* *L3b2*	124 223 278 362 *124* *311*	G	+10084*Taq* I
>L3d *L3d1*	124 223 *319*	G	-8616*Mbo* I
>L3e *L3e1* *L3e1a* *L3e2b* *L3e3* *L3f*	223 *327* *185* *172 189* *265T* *209 311*	G	+2349*Mbo* I
>M	223	G	+10397*Alu* I
>>M1	129 189 223 249 311	G	+10397*Alu* I
>>C	223 298 327	G	+10397*Alu* I -13259*Hinc* II/+13262*Alu* I
>>D	223 362	G	-5176*Alu* I +10397*Alu* I
>>E	223 227 362	G	-7598*Hha* I +10397*Alu* I
>>G	*017* 129 223	G	+4830*Hae* II/+4831*Hha* I +10397*Alu* I
>>Z	185 223 224 260 298	G	+10397*Alu* I
>N	223	G	+10871*Mnl* I (present also in all below)
>>N1	223	G	+10237*Hph* I
>>>N1a	147A/G 172 223 248 355	G	+10237*Hph* I
>>>N1b	145 176G 223	G	+10237*Hph* I
>>>N1c	223 265	G	+10237*Hph* I
>>>I	129 223 *391*	G	+10032*Alu* I
>>A	223 290 319	G	+663*Hae* III
>>W	223 292	G	-8994*Hae* III

continued

Table 4.2 continued

>>X	189 223 278	G	+14465*Acc* I
>>R	CRS	G	
>>>R1	278 311	G	
>>>R2	*71*	G	
>>>B	189	G	9bp del COII-tRNA^{Lys}
>>>F	304	G	-12406*Hinc* II/-12406*Hpa* I
>>>Y	126 231 266	G	+7933*Mbo* I -8391*Hae* III
>>>JT	126	G	+4216*Nla* III
>>>>J	*069* 126	G	+4216*Nla* III -13704*Bst* OI
>>>>>J1	*069* 126 261	G	+4216*Nla* III -13704*Bst* OI
>>>>>>J1a	*069* 126 145 231 261	G	+4216*Nla* III -13704*Bst* OI
>>>>>>J1b	*069* 126 145 222 261	G	+4216*Nla* III -13704*Bst* OI
>>>>>>>J1b1	*069* 126 145 172 222 261	G	+4216*Nla* III -13704*Bst* OI
>>>>>J2	*069* 126 193	G	+4216*Nla* III -13704*Bst* OI
>>>>T	126 294	G	+4216*Nla* III +13366*Bam* HI/- 13367*Ava* II/+13367*Mbo* I +15606*Alu* I -15925*Msp* I
>>>>>T1	126 163 186 189 294	G	+4216*Nla* III -12629*Ava* II +13366*Bam* HI/- 13367*Ava* II/+13367*Mbo* I +15606*Alu* I -15925*Msp* I
>>>>>T2	126 294 304	G	+4216*Nla* III +13366*Bam* HI/- 13367*Ava* II/+13367*Mbo* I +15606*Alu* I -15925*Msp* I
>>>>>T3	126 292 294	G	+4216*Nla* III +13366*Bam* HI/- 13367*Ava* II/+13367*Mbo* I +15606*Alu* I -15925*Msp* I
>>>>>T4	126 294 324	G	+4216*Nla* III +13366*Bam* HI/- 13367*Ava* II/+13367*Mbo* I +15606*Alu* I -15925*Msp* I
>>>>>T5	126 153 294	G	+4216*Nla* III +13366*Bam* HI/- 13367*Ava* II/+13367*Mbo* I +15606*Alu* I -15925*Msp* I
>>>U	CRS	G	+12308*Hinf* I

continued

Table 4.2 continued

>>>>U1	249	G	-4990*Alu* I +12308*Hinf* I -13103*Hinf* I/+13104*Mbo* I +14068*Taq* I
>>>>>U1a	189 249	G	-4990*Alu* I +12308*Hinf* I -13103*Hinf* I/+13104*Mbo* I +14068*Taq* I
>>>>>U1b	249 327	G	-4990*Alu* I +12308*Hinf* I -13103*Hinf* I/+13104*Mbo* I +14068*Taq* I
>>>>U2	*051* 129C	G	+12308*Hinf* I +15907*Rsa* I
>>>>U3	343	G	+12308*Hinf* I
>>>>U4	356	G	+4643*Rsa* I +11329*Alu* I +12308*Hinf* I
>>>>U5	270	G	+12308*Hinf* I
>>>>>U5a	192 270	G	+12308*Hinf* I
>>>>>>U5a1	192 256 270	G	+12308*Hinf* I
>>>>>>>U5a1a	256 270 *399*	G	+12308*Hinf* I
>>>>>U5b	189 270	G	+12308*Hinf* I
>>>>>>U5b1	144 189 270	G	+12308*Hinf* I
>>>>U6	172 219	G	+12308*Hinf* I
>>>>>U6a	172 219 278	G	+12308*Hinf* I
>>>>>>U6a1	172 189 219 278	G	+12308*Hinf* I
>>>>>U6b	172 219 311	G	+12308*Hinf* I
>>>>U7	318T	G	+12308*Hinf* I
>>>>K	224 311	G	-9052*Hae* II/-9053*Hha* I +12308*Hinf* I
>>>pre-HV	CRS	A	+11718*Hae* III
>>>>pre-HV	126 362	A	+11718*Hae* III
>>>>HV	CRS	A	+11718*Hae* III -14766*Mse* I
>>>>>HV1	*67*	A	+11718*Hae* III -14766*Mse* I
>>>>>H	CRS	A	-7025*Alu* I +11718*Hae* III -14766*Mse* I
>>>>>V	298	A	-4577*Nla* III +11718*Hae* III -14766*Mse* I

Notes. L Hgs in italics have been taken from Salas *et al.*(2002) as these provide increased resolution for some of the African hgs. For brevity only those L hgs that have been found in the populations analysed in this thesis (in this Chapter and Chapter 5) have been included. HVSI sequences are less 16,000, and the coding-region mutations indicate the name of the restriction enzyme and the location of the restriction site. The geographic distribution of mtDNA lineages is shown in Figure 4.2

Table modified from Richards *et al.* (2000) (supplementary material)

4.1.3. mtDNA - Evidence for Paternal Inheritance, Recombination and Selection?

As stated above, one of the most fundamental assumptions made in using mtDNA for population studies is that it only maternally inherited and therefore does not recombine. The assumption was that the paternal mitochondria do not penetrate the egg during fertilization. Whilst Schwartz and Vissing (2002) presented evidence for paternal inheritance of mitochondria in the case of one male with mitochondrial myopathy, a larger sample of individuals suggests that paternal inheritance is rare or non existent.

Recently evidence has been presented that contradicts the no-recombination assumption (Awadalla *et al.* 1999; Eyre-Walker *et al.* 1999, Hagelberg *et al.* 1999) with obvious implications for the use of mtDNA in population studies. mtDNA recombination could happen in at least 3 ways: in heteroplasmic individuals, with regions of nuclear DNA that have sequence identity, or with paternal mitochondria. Hagelberg *et al.* (1999) argued that the presence of a rare point mutation in several different mtDNA lineages in a Melanesian population was most reasonably explained by the presence of recombination, however resequencing indicated it was a sequencing error (Hagelberg *et al.* 2000). Eyre-Walker *et al.* (1999) argued that the amount of homoplasy, or parallel evolution, in human mtDNA sequences was too great to occur by mutation alone, whilst Awadalla *et al.* (1999) examined the decay of linkage disequilibrium (LD; the non-random association between alleles in a population that are more likely to be inherited together because of limited recombination between them [Jobling *et al.* 2003]) over increasing distances in the mitochondrial genome and found that LD did indeed decay with distance, which is the typical pattern expected for normally recombining autosomes. The results of both of these studies have however been refuted. For example, Elson *et al.* (2000) did not find evidence for excessive homoplasy, nor a decay of LD with distance, the latter finding was also supported by Ingman *et al.* (2000). Several studies (Elson *et al.* 2000; Kivisild and Villems 2000; Jorde and Bamshad 2000) have also questioned the methodology of Awadalla *et al.*'s (1999) study.

The answer to whether mtDNA do recombine, with its implications for the possibility of paternal inheritance is thus far from clear, although even original proponents of the arguments in favour of recombination are swaying away from their original stance (Eyre-Walker and Awadalla 2000; Awadalla 2003). Undoubtedly it is correct to maintain an open mind and not simply dismiss possible evidence for paternal inheritance and recombination simply because it would force a re-evaluation of the accepted paradigm. Given the uncertainty over the claims for recombination however, the work in this thesis assumes that mtDNA is only maternally inherited as there is no overwhelming evidence to reject this assumption at present, whilst acknowledging that paternal inheritance is a possibility.

A second potential confounding factor in the use of mtDNA in population studies is the possibility that natural selection affects the pattern of diversity within and between populations. The assumption of the neutral evolution of the mitochondrial genome has been based on early inferences made from the observance of the high rate of evolution of the genome compared to nuclear genes (reviewed by Gerber *et al.* 2001). Recently, however, tests of selection applied to the mitochondria of various species suggest that the conclusion of neutrality might need to be reconsidered, although an apparent signal of selection may also be caused by factors such as population size and structure (Gerber *et al.* 2001). Most studies that questioned the neutrality of mtDNA focused on genes in the coding region whilst human population studies also use information from the control region, which is not thought to code for genes and is thus not expected to show evidence for selection. However (if one assumes that mtDNA does not recombine), selection acting on a gene(s) in the coding region would affect the frequency of variants in the non-coding region through a selective sweep. Mishmar *et al.* (2003) concluded that the non-African distribution of mtDNA lineages used in population studies could be due to selection, associated with adaptation to different climates, acting on the mitochondrial genome, as modern humans migrated out of Africa and faced greater climatic stress than in Africa. The role of mtDNA in human adaptation to climatic stress is highly plausible given its function of energy production (Strachan and Read 1999).

4.1.4. mtDNA Nomenclature

As illustrated in Figure 4.2 and Table 4.2, an alternating letter and number system is used to name a series of nested clades of mtDNA lineages in a similar way to the YCC (2002) nomenclature. The basal lineages M and N are described as macrohaplogroups because they give rise to several other groups of lineages; L1, L2 and L3 are classified as superhaplogroups and the remaining lineages defined by a single letter are called hgs. Lineages derived from these hgs and defined by a single letter are classified as subgroups, and lineages derived from the subgroups are called derivatives. An asterisk is used to indicate a potential paraphyletic group (Richards *et al.* 2000). The terms lineage and hg are used to generically refer to any of the classifications just described when differentiating between these classifications is either obvious from the context or unimportant. The geographic distribution of lineages relevant to the work in this thesis are described in the Introductions to this Chapter and Chapter 5. The nomenclature system for mtDNA lineages has often been subject to revision making it somewhat confusing when referring to different studies, indeed mtDNA could do with a similar overhaul to that imposed on the Y-chromosome by the YCC (2002).

4.1.5. mtDNA Hg Distribution in Europe

All non-African mtDNA lineages derive from the macrohaplogroups M and N. M contains hgs that are mainly found in South and East Asia, whilst N contains hgs predominantly distributed in West Asia and Europe. The latter of which is broken down into the following clusters: HV, UK, TJ, and WIX (Finnilä *et al.* 2001, see also Figure 4.2). mtDNA diversity in Europe does not exhibit high levels of geographic structuring. Most European populations tend to have the same hgs present albeit at different frequencies, even populations such as the Basques and Saami who are typically considered to be outliers on the basis of many factors have similar mtDNA hgs to the rest of Europe (Simoni *et al.* 2000).

A summary of the most frequent mtDNA hgs seen in British populations is found in Table 4.3, which will be briefly described here. Hgs H and V are sister lineages within HV. Several studies have found H to be the single commonest lineage in Britain. The CRS HVSI sequence motif is common in H, but is also found in other hgs (see below), and the RFLP sites that describe H are seen in Table 4.2. Estimates for the frequency of H in Britain are ~0.50, and quite similar for Europe (González *et al.* 2003; Helgason *et al.* 2001; Richards *et al.* 1996), although the frequencies decrease towards the Middle East (Al-Zahery *et al.* 2003, González *et al.* 2003; Richards *et al.* 1996). Although several subclusters within H have been identified on the basis of whole mtDNA genome sequencing (Finnilä *et al.* 2001), not all studies today use these sublineages, hence they are not considered here. H is thought to have originated in the Middle East, but to have reached its present day high frequency as a result of population expansions after the LGM, rather than being a signature of the Neolithic Expansion (Richards *et al.* 1996; Richards *et al.* 2000); the high frequency of H in north western Europe is thought to be because the expansion was from the Iberian peninsula or southern France (Forster, 2004). Although V derives from the same root as H it is much more rare in Britain and Europe, ranging in frequency from 0.013-0.043 in Britain. The distribution of V in Eurasia is strongly skewed with the highest frequencies being in the west and the hg is virtually absent in the Middle East, Caucasus, Turkey and the Balkans. This distribution along with the older age of the hg in the west than the east was interpreted as a signal of population expansion after the LGM from a refuge in Iberia, which is where the hg originated, whilst pre*V originated in Europe before the LGM (Torroni *et al.* 2001b).

U is an important western Eurasian hg (Quintana-Murci *et al.* 2004), although the presence of many subgroups means that U lineages are found across Europe, the Middle East, India and Africa. Of the many U subgroups, U4 and U5 appear to be most common in Britain, although U4 is clearly more common in Scotland. U5 has coalescent ages in the Upper Palaeolithic (González *et al.* 2003; Dubut *et al.* 2004) and has expanded greatly across Europe (Maca-Meyer *et al.* 2001). U3 is also quite common in Scotland, this lineage is hypothesised to be a Neolithic founder that has migrated from the Middle East (Richards *et al.* 2000). U6 is the

Table 4.3 mtDNA Haplogroup Frequencies in British Populations

Hg	Frequency in Population						
	Ireland[a]	Orkney[a]	Scotland[a]	England/ Wales[a]	Western Isles/Isle of Skye[a]	Scotland[b]	England[b]
allH	0.477	0.507	0.457	0.522	0.346	-	-
H/CRS	-	-	-	-	-	0.268	0.340
CRS	-	-	-	-	-	0.164	0.191
V	0.070	0.013	0.043	0.037	0.020	0.035	0.031
U1	-	-	-	-	0.020	0.006	-
U2	0.008	-	0.008	0.007	0.004	0.008	-
U3	-	-	0.012	0.007	0.024	0.019	-
U4	0.023	0.007	0.025	0.016	0.004	0.026	0.023
U5	0.063	0.118	0.072	0.065	0.081	0.084	0.061
U6	-	-	-	-	-	-	-
U7	-	-	0.001	-	-	-	-
K	0.078	0.066	0.066	0.061	0.134	0.080	0.073
J	0.117	0.079	0.086	0.107	0.106	0.076	0.118
J1	-	-	0.006	0.005	0.012	0.008	0.004
J1a	0.008	-	0.004	0.016	-	0.006	0.011
J1b	-	-	0.001	-	-	-	-
J1b1	0.008	0.020	0.035	0.014	0.012	0.023	0.023
J2	0.008	-	0.011	0.002	0.016	0.009	-
allT	0.094	0.059	0.101	0.077	0.126	0.103	0.065
W	0.023	0.020	0.009	0.016	0.004	0.013	-
I	0.023	0.033	0.044	0.030	0.065	0.042	0.038
X	-	0.072	0.017	0.009	0.020	0.021	0.015
Other	-	0.007	0.002	0.007	0.004	0.010	0.008

Notes - the pre-fix "all" in hg names indicates that several subgroups have been collasped into one hg due to low frequencies, for example "allH" includes H, H1, H3, H4, H5, and H8 lineages indentified by Helgason *et al.* (2001). The "allH", "H" and "H/CRS" are all broadly comparable due to the high frequency of CRS sequences in hg H. The term "other" is a pooled group of disparate hgs observed at very low frequencies

[a] Helgason *et al.* (2001)

[b] Gonzalez *et al.* (2003)

main U subgroup found in Africa, it is generally rare in Europe, although it has been found in populations such as Portugal, where there is a known history of contact with Africa, and a history of the slave trade (González *et al.* 2003). U6 is considered in more detail in Chapter 5 (section 5.1.5). Hg K is proposed to be a lineage of U (Macaulay *et al.* 1999; Maca-Meyer *et al.* 2001) and is thought to have arrived before the LGM (i.e. pre-Neolithic) and then suffered as a result of the LGM and have subsequently re-expanded (Richards *et al.* 2000). This hg is typically at low frequency in Europe and the Middle East, although it has been found at high frequency in Ashkenazi Jews (Behar *et al.* 2004), this latter finding is considered in more detail in Chapter 5 (5.1.3).

Within the TJ cluster González *et al.* (2003) found that the subhaplogroup J was most common in Britain compared to other European populations. The J derivatives J1a and J1b1 were particularly common in Britain and other northern areas such as Scandinavia and Germany, whilst J2 is absent from northerly populations apart from Scotland. Elsewhere in Europe J is commonest in the Middle East, where the coalescence dates are older than for the rest of Europe, hence its presence in Europe is thought to be a signature of the Neolithic expansion, as is the rarer hg T (Richards *et al.* 2000). Representatives of the WIX cluster are very rare across Britain and much of Eurasia (Helgason *et al.* 2001; González *et al.* 2003).

4.1.6. Aims of the Chapter

Most genetic history studies do not focus on cities and specifically aim to sample DNAs from small rural populations (Cavalli-Sforza *et al.* 1994, Richards *et al.* 1996; Weale *et al.* 2002; Chapter 2 and Capelli *et al.* 2003) in order to study *past* historical events. This chapter asked whether the Y-chromosome and mtDNA diversity assayed from a sample of Londoners was comparable to historical predictions and those from the 2001 Census. In addition, a comparison was made, for the Y-chromosome data, with the rest of the British population

(make possible by the work described in Chapter 2), an independently collected sample of Londoners, and another metropolitan district (Liverpool).

4.2. Materials and Methods

4.2.1. Sample Collection

DNA samples were collected from 107 men and 124 women who visited the Museum of London during the National Archaeology Days on the 19th and 20th July 2003. People who participated were over 18 years of age, to the best of their knowledge had no same-sex blood line relatives participating, and had to provide a residential address within Greater London. Volunteers did not have to be able to trace their ancestry to London as the aim was to assess diversity in present day Londoners. Appropriate informed consent was obtained from all volunteers before the sample was taken; volunteers were also asked to fill in a voluntary, anonymous, form detailing their self designated ethnic origin to enable a comparison between the ethnic composition of the samples collected at the Museum and the ethnic make-up of London, ascertained from the 2001 Census. Due to the anonymous nature of the ethnic identity questionnaire it was not possible to correlate an individual's self designated ethnic identity with their genetic results. For reasons of racial equality and sensitivities associated with collecting information on ethnic background, these questionnaires had to remain anonymous. Samples were collected by mouth swab by the volunteer under direction from JK Abernethy or two trained assistants and stored in tubes containing 1ml of 0.05M EDTA/0.05M SDS preservative solution. Samples were transported to the lab on the day of collection and stored at 4°C until extraction.

Often, in studies of genetic history, only male DNA samples are collected even if both the Y-chromosome and mtDNA wish to be studied. Such a sampling strategy is not only time effective it also allows a direct comparison of the maternal and paternal history of the same individuals in a population, which can often reveal interesting disparities in their history (see for example Seielstad et al, 1998; Hurles *et al.* 1998; Carvajal-Carmona *et al.* 2000). This approach

would have been pursued in the present work. However, for reasons of sex discrimination the Museum of London were not happy to only allow male samples to be collected. Therefore to allow the project to proceed at all it was necessary to collect male samples for Y-chromosome typing and female samples for mtDNA analysis. It was decided to not type the male samples for mtDNA polymorphisms because of time restraints. Hence the male London samples are hereafter referred to as LondonY, and the female as LondonMT

4.2.2. DNA Extraction

DNA was extracted using the Promega Wizard ® Genomic DNA Purification Kit, using the modifications described in Section 3.2.2.

4.2.3. Y-Chromosome Genotyping

All male DNA samples were typed for YSTR1 and EURO1 as described in Tables 2.3 and 2.4, and a further 5 UEPs detailed in Tables 2.5 –2.7. Additionally SRY_{10831a} and M201 were typed on 5 chromosomes (Table 4.4). Samples were electrophoresed using the 3700 Automated Sequencer by M-W Burley using the modifications described in Section 3.2.3. Difficulties arose in typing M89 on LondonY; for the 8 samples that required M89 typing (using the criteria in section 2.2.3) all displayed both ancestral and derived alleles making it impossible to distinguish between the derived and underived states. The two control DNAs included in the M89 PCR multiplex as routine displayed the anticipated derived state (both samples were known to be Tat derived, hence M89 derived) thus the multiple peaks present in the LondonY samples was not related to either a PCR amplification or restriction enzyme problem. Similar problems with M89 had been experienced in another laboratory on an independent dataset where preliminary experiments suggested ambiguous M89 results correlated with poor quality DNA (Cristian Capelli, personal communication). This may explain the difficulties encountered here as the DNA

Table 4.4. SRY_{10831a} PCR and RFLP Conditions and Expected Allele Sizes

(a) Primer Mix

Primer Name	*Fluorescent label (5')*	*Conc (μM)*	*Volume per reaction (μl)*
SRY10831L	FAM	0.150	0.0300
SRY10831R	-	0.150	0.0300
dH_20	-	-	0.9400
Total			1.0000

PCR mix

Component	*Volume per reaction (μl)*
Primer mix	1.00
10X Buffer	1.00
10mM dNTP	0.20
dH_20	6.76
Taq[a]	0.04
DNA	1 (~5ng)
Total	10.00

[a] 2HTTaq:1TaqStart™

Cycling Conditions

Temperature (degrees C)	*Duration*[b]	*Cycles*
95	4'	
95	40"	38
58	40"	
72	40"	
72	10'	
4	∞	

[b] Minutes ', seconds "

(b) RFLP Protocol

Enzyme mix

Enzyme	*UEP Restriction Site*	*Volume per reaction (μl)*
DraIII	SRY10831	0.024
dH_20	-	0.276
Total		0.3

Digestion mix

Component	*Volume per reaction (μl)*
Enzyme mix	0.3
NEB buffer 3	0.8
10 mg/ml BSA	0.08
PCR Product	2
dH_20	4.82
Total	8

Incubation

Temperature (degrees C)	*Duration*
37	Overnight

(c) Expected Allele Sizes (ABI PRISM ® 3700 DNA Sequencer)

Locus	*Ancestral Allele*	*Derived Allele*
SRY10831	44-G	77-A

extraction method used had consistently given lower DNA yields for other independently collected DNA samples (Helen Roberts, personal communication). Three of the M89-unknown chromosomes were subsequently found to be M35 derived and an M89 underived state could be inferred. The remaining five chromosomes were hierarchically tested for the markers SRY_{10831a} and M201. All 5 chromosomes were SRY_{10831a} derived, excluding them from hg A and placing them within BR (see Figure 2.3a). M201 was successfully typed on 4/5 of these chromosomes, all of which were found to have the derived allele. M201 is a branch internal to M89, therefore in analyses these 4 chromosomes will be considered as M89 derived. The remaining chromosome whose UEP status is presently underived or unknown for all of the markers used here was excluded from analyses involving hg frequencies, but included in analyses involving haplotypes. In total 93 Y-chromosomes could be used in hg analyses, and 94 in analyses using haplotypes.

4.2.4. mtDNA HVS1 Procedures

The hypervariable segment 1 (HVSI) region of the mitochondrial genome was successfully amplified on 117 female DNA samples, using the primers conH1 and conL2 (see Appendix, Table A.2 for sequences) and a standard PCR protocol (Table 4.5). 4µl of PCR product was electrophoresed on a 1% (w/v TBE) agarose gel to visualise PCR products and ascertain which samples could be subsequently sequenced. Several DNAs failed to amplify successfully using ~5ng of DNA; these were repeated using ~10ng of DNA, leading to successful amplification. PCR products and sequencing reactions were cleaned, and the sequencing reactions performed, using a protocol modified from an original idea by Dr M Thomas (Table 4.5). Sequences were electrophoresed on an ABI PRISM® 3700 DNA Sequencer by M-W Burley. Briefly, 10µl of formamide was added to each sample, the plate centrifuged at 1000rpm for 1 minute, and electrophoresed according to the manufacturers instructions using POP6. Sequence alignments were performed in Sequencher ™. Polymorphisms were called with respect to the Cambridge Reference Sequence, CRS, (Anderson *et.*

Table 4.5. PCR and Sequencing Protocol Used for mtDNA HVSI Analysis

(a) PCR Protocol

Component	*Volume per reaction (μl)*
ddH_2O	5.47
Qiagen PCR Buffer (contains 15mM $MgCl_2$)	1
Qiagen 25mM $MgCl_2$ (final Mg^{2+} conc. 2.5mM)	0.4
dNTP Mix (25mM per dNTP)	0.08
Qiagen HotStart Taq (5U/μl)	0.05
U' Primer (5μl)	1
R' Primer (5μl)	1
DNA (~5ng/μl)	1
Total	10

Cycling Conditions

Temperature (degrees C)	*Duration*[a]	*Cycles*
95	4'	
94	40"	35
57	40"	
72	40"	
72	10'	
4	∞	

[a] Minutes ', seconds "

PCR Clean Up

1. Add 1 volume of MicroCLEAN to each PCR reaction. Mix and incubates at room temperature for 10 min.

2. Centrifuge at 1,800g for 60 min at room temperature.

3. Invert the plate and place back in the centrifuge on tissue paper. Centrifuge at 120g for 1min at room temperature.

4. Add 150 μl of 70% Ethanol to each reaction and centrifuge at 1,800g for 10 min at room temperature.

5. Invert the plate and place back in the centrifuge on a piece of tissue paper. Centrifuge at 120g for 1min at room temeperature. Remove the plate from the centrifuge air-dry for 15 min at room temperature or for 5 min at 65°C.

6. Add 5μl H_2O to each sample.

continued

Table 4.5 continued

(b) Sequencing Protocol

Component	*Volume per reaction (μl)*
ddH_2O	5.36
ABI Ready Reaction Mix	1
ABI 5X Buffer	2
Sequencing Primer (5μl)	0.64
PCR Product	1
Total	10

Cycling Conditions

Temperature (degrees C)	*Duration* [a]	*Cycles*
96	10"	
50	5"	25
60	4'	
4	∞	

[a] Minutes ', seconds "

Sequencing Clean Up

1. Add 40 μl of 80% ethanol to each reaction, cover and mix by inversion. Incubate at room temperature for 15 min.

2. Centrifuge at 1,800g for 60 min at room temperature.

3. Invert the plate on a piece of tissue paper and tap gently to remove ethanol.

4. Add 150 μl of 70% ethanol to each sample, invert, and place in the centrifuge on tissue paper. Centrifuge at 1000rpm for 1 min. Remove from the centrifuge air-dry for 15 min at room temperature for 5 min at 65°C. The samples are ready for electrophoresis.

al. 1981). Financial and time constraints did not allow the author to use RFLP assays, therefore clade and hg assignments are on the basis of HVSI sequence motif only

4.2.5. mtDNA Clade and Hg Assignment

Most sequences (102/117) could be assigned to hgs (see Appendix, Table A.5) using HVSI sequence motif and the criteria of Richards *et al.* (2000, supplementary data). For H and U hg assignment was somewhat tentative as there is a degree of sequence sharing between these groups that most authors resolve using RFLPs (see for example, Torroni *et al.* 1996; González *et al.* 2003). The CRS, for example, appears in pre-HV, HV, H, and U* hgs, the distinction between these hgs being made with the following restriction sites: pre-HV: +11718*Hae*III; HV: +11718*Hae*III and -14766*Mse*I; H: -7025*Alu*I, +11718*Hae*III, and -14766*Mse*I, U*: +12308*Hinf*I. In this study all CRS sequences were placed into H and no attempt was made to distinguish the sub-lineages of H. The remaining sequences that could not be assigned to a clade or hg were omitted from analyses requiring this information. Polymorphisms are given as a string of nps where mutation occurs, listing the base change when it was a transversion (G/A-C/T) but only the np for transitions (G-A and C-T).

4.2.6. Y-Chromosome Comparison Populations

The British populations described in Chapter 2 were used as comparison populations. Additionally, chromosomes from Liverpool, another British city, and an independently collected London sample (hereafter referred to as London2), have been used. These 2 city samples had been previously typed for the same Y-linked markers (Cristian Capelli, unpublished data). A full dataset of published European Y-chromosomes (Rosser *et al.* 2000) was also used a comparison for Europe, hereafter referred to as the Rosser Dataset (RD). This enabled comparisons with the following populations: Iceland, Saami, Northern Sweden, Gotland, Norway, Denmark, Finland, Estonia, Latvia, Lithuania,

Russia, Belarus, Ukraine, Mari, Chuvashia, Georgia, Ossetia, Armenia, Turkey, Cyprus, Greece, Bulgaria, Czech Republic, Slovakia, Hungary, Poland, Italy, Sardinia, Bavaria, German, Holland, France, Belgium, Western Scotland, Scotland, Cornwall, East Anglia, Ireland, Basque, Spain, Southern Portugal, Northern Portugal, Algeria, and North Africa. Please refer to Rosser *et al.* (2000) for hg frequencies.

There are differences between the RD and this thesis in the markers that have been typed: microsatellites were not typed by Rosser and colleagues, and although most of the UEPs define comparable lineages they are not identical. Therefore the hg+1 nomenclature cannot be used and certain hgs have to be placed into lower- resolution phylogenetically related groups. Hence analyses employing the RD involve some inevitable loss of resolution, but as the RD represents one of the most comprehensive study of European Y-chromosomes, the comparisons this enables should more than outweigh any loss of resolution. The hg clustering methodologies applied to both datasets are shown in Table 4.6.

4.2.7. mtDNA Comparison Populations

Data from 3 published studies (Al-Zahery *et al.* 2003; González *et al.* 2003; Helgason *et al.* 2001, hereafter referred to as AD, GD, and HD respectively) were used for comparative purposes, providing information for the following populations: (AD) Iraqi, Iranian, Arabian, Syrian, Palestinian, Georgian, Armenian, Anatolian, Italian, Slavic speakers, Finno-Ugric speakers, Germans, Central Asians, Indian; (GD) Finns, Norwegian, Scottish, English, North Germans, South Germans, French, Galicia, North, South and Central Portugal, and North Africa; (HD) Austrians/Swiss, European Russians, Finns/Estonians, French/Italian, Germans, Icelandic, Irish, Orcadians, Scandinavians, Scottish, Bulgarians/Turks, Spanish/Portuguese, English/Welsh, Western Isles/Isle of Skye, and Saami. There is some overlap between these studies in the original source populations (for details please refer to each publication). Overlap is never

Table 4.6. Clustering Criteria Applied to Comparisons Between the Current Study and the RD

Markers used in this study	M173, 92R7	M170, M26, M89	M17	12f2, M172	M9	Tat	YAP, M35
Rosser *et al.* (2000) equivalent	hg1, hg22	hg2	hg3	hg9	hg12, hg26	hg16	hg21, hg4, hg8, hg25

populations respectively: Ireland, Orkney, Western Isles/Isle of Skye, Scotland, complete, therefore no total duplication of results occurred. Unfortunately there is not an mtDNA database of Britain to match that presented in Chapter 2 for the Y-chromosome; the highest level of mtDNA resolution for Britain is afforded by the Helgason and González datasets which have data for the following England/Wales; and Scotland, England.

To allow comparison between these three datasets and LondonMT it was necessary to cluster all of the datasets into broadly defined clades: H/CRS, V, TJ, K, U, I, X, W, L, Z, and "other" (see Table 4.7). All J and T chromosomes from the LondonMT data and the comparative datasets were placed into the group "allTJ" to allow the 5 LondonMT chromosomes designated as JT (defined by the mutation at 16126) to be included in the analysis, as the comparative datasets placed chromosomes into either the J *or* T lineages, and not JT.

4.2.8. Y-chromosome Data Analysis

The hg+1 composition of LondonY, London2 and Liverpool, was compared with the British and European comparison populations using PC plots (POPSTR, H Harpending personal communication). Exact tests of population differentiation (Arlequin 2.000, Schneider *et al.* 2000) were calculated as in Chapter 3 (section 3.2.6). Additionally, pairwise Fst values were calculated (Arlequin 2.000, Schneider *et al.* 2000) as a measure genetic distance, where Fst = 0 indicates that populations are identical, and Fst = 1 implies no sharing of alleles and populations are completely different. p-values were calculated using 10,000 iterations. The p-value is the proportion of permutations leading to an Fst greater than or equal to that observed.

Hg (and haplotype, where relevant) diversity (*h*), and its sampling variance [(V(H)], were calculated in (Arlequin 2.000, Schneider *et al.* 2000).

Table 4.7. mtDNA Haplogroups Encountered in LondonMT and Comparison Populations

Population\Clade	H	V	JT	K	U	I	X	W	L	Z	Misc	Other	Sample size
LondonMT [1]	**47**	**4**	**16**	**11**	**16**	**2**	**1**	**2**	**2**	**0**	**1**	**0**	**102**
Iraqi [2]	75	1	20	19	7	41	4	6	5	9	4	25	216
Iran [2]	113	0	61	38	34	97	9	13	28	10	9	39	451
Arabia [2]	123	0	81	18	14	41	3	7	30	41	7	24	389
Syrian [2]	24	2	7	7	3	11	0	0	1	4	2	8	69
Palestinian [2]	41	0	11	15	8	9	0	4	2	6	3	18	117
Georgian [2]	35	1	5	18	14	30	3	14	4	0	2	13	139
Armenian [2]	74	0	17	22	15	43	3	4	1	0	2	10	192
Anatolia [2]	122	0	42	46	23	75	9	17	20	1	15	17	388
Italian [2]	35	5	7	9	8	22	4	3	0	0	2	4	99
Slav [2]	134	10	34	40	12	63	9	2	3	0	3	14	324
Finno-Ugrie [2]	68	3	18	9	5	34	2	3	1	0	4	2	149
German [2]	100	5	15	17	13	27	5	1	0	0	2	15	200
C-Asian [2]	29	0	5	7	1	16	2	0	93	0	2	50	205
Indian [2]	46	0	10	14	3	156	8	3	862	0	20	179	1300
Austria/Switzerland [3]	101	5	31	17	25	4	1	3	0	0	0	0	187
European/Russia [3]	91	9	39	6	50	3	0	5	0	1	8	3	215
Finland/Estonia [3]	89	13	28	5	45	5	3	11	0	0	1	2	202
France/Italy [3]	133	7	51	15	25	2	5	2	0	0	2	6	248
Germany [3]	258	27	97	35	74	12	4	11	0	0	3	6	527
Iceland [3]	222	8	113	36	55	22	7	1	0	1	2	0	467
Ireland [3]	61	9	30	10	12	3	0	3	0	0	0	0	128
Orkney [3]	77	2	24	10	19	5	11	3	0	0	0	1	152
Scandinavia [3]	313	37	123	32	105	12	4	10	0	4	2	3	645
Scotland [3]	407	38	218	59	105	39	15	8	0	0	1	1	891
Bulgaria/Turkey [3]	39	0	25	6	12	2	4	4	0	0	7	3	102
Spain/Portugal [3]	206	21	42	16	37	2	6	7	0	0	8	7	352

continued

Table 4.7. continued

Population\Clade	*H*	*V*	*JT*	*K*	*U*	*I*	*X*	*W*	*L*	*Z*	*Misc*	*Other*	*Sample size*
England/Wales[3]	224	16	95	26	41	13	4	7	0	0	1	2	429
Western Isles/Isle of Skye[3]	85	5	67	33	33	16	5	1	0	0	1	0	246
Saami[3]	10	70	0	0	80	0	0	1	0	6	9	0	176
Finland[4]	18	6	6	2	6	5	5	0	0	0	2	0	50
Norway[4]	151	14	51	18	72	8	6	1	0	0	2	0	323
Scotland[4]	378	31	124	70	196	37	11	18	1	0	8	0	874
England[4]	139	8	22	19	58	10	0	4	1	0	1	0	262
North Germany[4]	73	8	13	11	27	1	2	1	1	0	3	0	140
South Germany[4]	137	12	38	15	47	8	3	3	0	0	3	0	266
France[4]	121	7	16	12	36	3	6	2	3	0	7	0	213
Galicia[4]	86	7	10	7	15	0	3	1	4	0	2	0	135
North Portugal[4]	84	12	29	7	34	5	6	0	4	0	3	0	184
Centre Portugal[4]	83	5	17	13	26	0	2	3	7	0	6	0	162
South Portugal[4]	93	8	16	12	31	1	4	7	17	0	7	0	196
North Africa[4]	111	17	78	12	34	0	2	4	76	0	16	0	350

Notes: Where sublineages have been identified in the comparison datasets they have been grouped into a single clade, for example L1, L2, L3 etc are here placed into "L". Additionally "H" contains all H and CRS sequences, "JT" contains all J and all T sequences. "Miscellaneous" refers to the non-specific grouping of hgs by the 3 authors listed (please refer to original publications for details), which cannot be assigned to one single hg. The LondonMT hg C sequence was also placed in this group. "Other" refers to the categories of the same name used by the 3 publications to place undesignated sequences.

[1]. Present Study, [2]. Al-Zahery *et al.* (2003), [3]. Helgason *et al.* (2001), [4]. Gonzalez *et al.* (2003).

$$\hat{H} = \frac{n}{n-1}\left(1 - \sum_{i=1}^{k} p_i^2\right)$$

$$V(\hat{H}) = \frac{2}{n(n-1)}\left\{2(n-2)\left[\sum_{i=1}^{k} p_i^3 - (\sum_{i=1}^{k} p_i^2)^2\right] + \sum_{i=1}^{k} p_i^2 - (\sum_{i=1}^{k} p_i^2)^2\right\}$$

where n is the number of gene copies in the sample, k is the number of haplotypes, and p_i is the sample frequency. H is defined as the probability of two randomly chosen chromosomes from a sample are different. Calculations of h can be elevated by high frequency hgs and haplotypes and do not reflect diversity represented by low frequency types, thus hg and haplotype diversity can be better assessed by simply considering their frequencies. Analyses using the BCD were performed using hg+1 frequencies, and those with the RD were calculated hg frequencies as shown in Table 4.6.

4.2.9. mtDNA Data Analysis

PC analysis and exact tests of population differentiation were performed as described above for the Y-chromosome (section 4.2.8) and using hg frequencies. Haplotype (sequence) diversity was calculated using h as above (4.2.8) and two population parameters, θ_π (mean pairwise differences), and θ_k (Arlequin 2.000, Schneider *et al.* 2000) and compared to published results for the HD (Helgason *et al.* 2001). θ_π is calculated as

$$\theta_\pi = \sum_{i=j}^{k} \sum_{j<i} p_i p_j d_{ij}$$

estimated from

$$E(\hat{\pi}) = \theta$$

where

$$\hat{\pi} = \sum_{i=j}^{k} \sum_{j<i} p_i p_j \hat{d}_{ij}$$

and $\hat{d}_{ij}$ is an estimate of the number of mutations that have occurred since the divergence of haplotypes i and j, k is the number of haplotypes, and p_i is the frequency of the haplotype i. θ_π measures the mean number of pairwise difference between HVSI sequences in a given population.The measure is sensitive to factors affecting allele frequencies, such as recent admixture (Árnason 2003). $\hat{\theta}_k$ is estimated from

$$E(k) = \theta \sum_{i=0}^{n-1} \frac{1}{\theta + i}$$

where k is the expected number of alleles, n is the sample size, and $\theta = 2nN_e\mu$ for haploid loci, such as the mitochondrial genome, where N_e is the effective population size and μ mutation rate. θ_k is used as an estimator of the effective female population size, or N_{fe} (Helgason *et al.* 2001). The value of θ_k increases with sample size (Árnason 2003). Both θ_π and θ_k can thus be used to infer the relative levels of sequence diversity between LondonMT and the comparison populations: a high θ_π score could thus indicate recent admixture θ_k on the other hand indicates those populations with relatively larger population sizes of females. These calculations rely on several assumptions: neutrality, a constant population size, the infinite-sites mutation model, and an equal mutation rate across populations (Helgason *et al.* 2001). Whilst it might be fair to assume that any departures from these assumptions will be equal across all populations (e.g. Helgason *et al.* 2001), there is some evidence to suggest this is not the case. For example, a much used Norwegian mitochondrial sample is drawn from infants who died from Sudden Infant Death Syndrome, but there is evidence that this is not a selectively neutral sample (Árnason 2003), therefore violating a basic assumption of the theta parameters in only one population, not across all populations.

To maximise the number of sequences that could be included in haplotype diversity calculations for LondonMT, bases 16061-16368 (inclusive) were used. This should result in the minimal loss of resolution, as some 82% of the total variation between 16010-16400 (i.e. most of the HVSI that is sequenced and

reported in the literature) is actually contained within a small stretch from 16090-16324 (Helgason *et al.* 2001). This allowed 117 individual LondonMT sequences to be included in the calculations.

4.2.10. Relative Y-Chromosome and mtDNA Diversity

AMOVA was implemented in Arlequin (Schneider *et al.* 2000) to assess the relative levels of hg diversity in LondonY and LondonMT in relation to each other and the rest of Britain. AMOVA allows the hierarchical assessment of the variation present and depending on the number of groups specified apportions it to the percentage of variation observed. When one group is specified the percentage of variation 1) among populations, and 2) within populations is estimated. When two or more groups are specified the percentage of variation observed 1) among groups, 2) among populations within groups, and 3) within populations is estimated. Hg frequencies were employed because comparable data were available for both British Y-chromosome and mtDNA lineages. To enable calculations for the Y-chromosome and mtDNA to be as equal as possible the BCD was placed into the following groups (the equivalent British populations from the HD and GD are given in parentheses): Orkney (Orkney); Scottish Isles (Western Isles/Isle of Skye); Scotland (Scotland HD, Scotland GD), England (England); Wales (England/Wales); Ireland (Ireland). AMOVA was performed separately for the Y-chromosome and mtDNA first by placing all British populations and London(Y/MT) into one group to assess the overall diversity in British Y-chromosome and mtDNA lineages, and secondly by placing all British populations into "Group 1" and London(Y/MT) into "Group 2".

4.3. Results

4.3.1. Y-Chromosome Comparisons with Britain

As the exact test of population differentiation showed that LondonY and London2 were significantly different at the hg+1 level (p=0.002), these two

London samples were not pooled and will be considered as separate populations (Table 4.8). The implications of this are considered further in the Discussion (section 4.6).

PC analysis of LondonY and the BCD placed LondonY with populations with low frequencies of M17 (towards the negative extreme of the x-axis) and moderate 2.47+1 frequencies (towards the centre of the y-axis) (Figure 4.3a). Compared to the PC plot drawn for the BCD only (Figure 2.8) the positioning of most populations remains unchanged, as LondonY fits within the general pattern of British diversity. LondonY is drawn towards the negative extreme of PC1 with an absence of R1a1 chromosomes, a trait which also characterises British populations described as indigenous in Chapter 2 (see section 2.3.1 for example). A plot of PCs 2 and 3 was also drawn as these components still explained ~35% of the variation (Figure 4.3b). Shetland and Orkney no longer fall as outliers and LondonY does not fall to any extreme. The addition of London2 and Liverpool to this analysis places London2 as an outlier to the rest of Britain (Figure 4.4a and b). The YAP+ chromosomes in this sample separates populations into two distinct groups: those with YAP+ (London2) and those without (the remaining populations) predominantly falling into an undifferentiated group. As the 2nd and 3rd components explained 34.6% of the variation and were not affected by the frequency of YAP+ chromosomes a plot of these components was also drawn. London2 is no longer an outlier and falls close to Oban whilst Liverpool is slightly drawn towards York and Norfolk.

Results of the exact test of population differentiation are summarised in Table 4.8. LondonY is significantly different to approximately half of the BCD populations, the only apparent geographic pattern is that LondonY is not significantly different to any of the south coast populations. London2 is significantly different to most of the BCD, as well as LondonY, suggesting a high level of structure between London2 and the rest of Britain. Liverpool is indistinguishable from most of the BCD making it difficult to identify any patterns. Pairwise Fst scores (Table 4.9) show that none of the British populations are highly differentiated, the most differentiation is between Durness and Llangefni (0.136) and Durness and London2 (0.134). The highest

Table 4.8. Y-Chromosome Exact Test of Population Differentiation: British Cities and BCD Using Hg+1 Frequencies

Population	*Shet*	*Ork*	*Dur*	*Wls*	*Sth*	*Ptl*	*Oban*	*Mpt*	*Pnt*	*IoM*	*York*	*Sow*
LondonY	**0.000**	**0.000**	0.171	**0.000**	0.707	0.091	0.516	0.057	0.088	**0.012**	0.052	**0.040**
London2	**0.007**	**0.000**	**0.000**	**0.000**	**0.023**	**0.003**	0.055	**0.001**	**0.016**	0.103	**0.004**	**0.015**
Liverpool	**0.005**	**0.007**	0.080	0.151	0.276	0.108	0.222	0.893	0.270	**0.046**	0.758	0.471

Population	*Utx*	*Ldl*	*Lgf*	*Rsh*	*Cas*	*Nor*	*Hfw*	*Chp*	*Fav*	*Mdh*	*Dcr*	*Cor*
LondonY	0.520	0.403	**0.005**	**0.000**	0.137	**0.021**	**0.030**	0.157	0.746	0.056	0.473	0.365
London2	**0.000**	0.126	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**	0.172	**0.026**	**0.003**	0.078	0.064
Liverpool	0.731	0.601	**0.000**	**0.006**	0.075	0.971	**0.000**	0.334	0.214	0.527	0.564	0.110

Population	*Chl*	*LonY*	*Lon2*	*Liv*
LondonY	0.058	-		
London2	**0.000**	**0.002**	-	
Liverpool	0.887	0.203	**0.003**	-

Notes: Shown are the p-values. Bold text indicates significant comparisons, $p < 0.05$. Abbreviations as follows: Shet = Shetland Isles, Ork = Orkney Isles, Dur = Durness, WIs = Western Isles, Sth = Stonehaven, Ptl = Pitlochry, Oban = Oban, Mpt = Morpeth, IoM = Isle of Man, York = York, Sow = Southwell, Utx = Uttoxeter, Nor = Norfolk, Hfw = Haverfordwest, Chp = Chippenham, Fav = Faversham, Mdh = Midhurst, Dcr = Dorchester, Cor = Cornwall, LonY = LondonY, Lon2 = London2, Liv = Liverpool

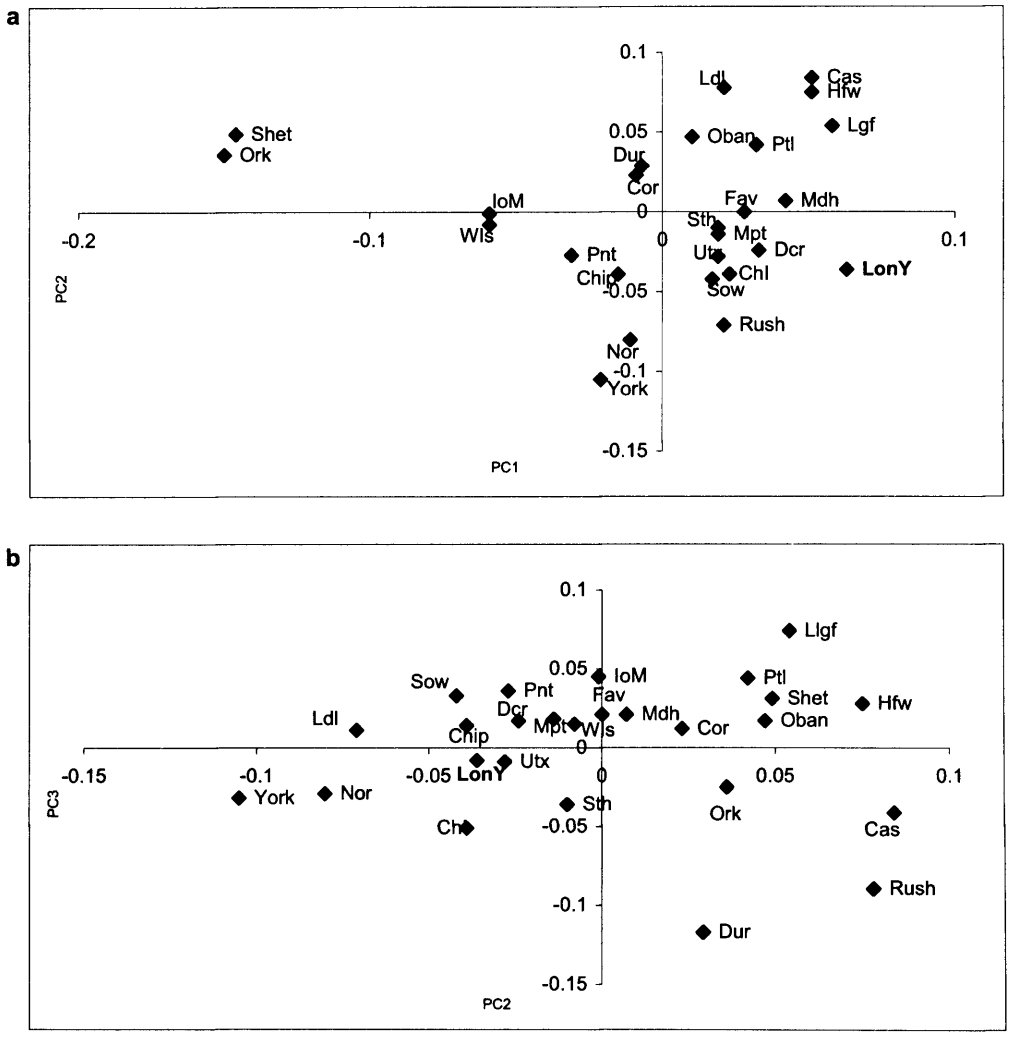


Figure 4.3. Y-Chromosome PC plots for LondonY and the BCD Using Hg+1 Frequencies. (a) PCs 1 and 2, PC1 explains 24% of the variation and PC2 20.5% (b) PCs 2 and 3, PC3 explains 15.1%. Abbreviations as Table 4.8.

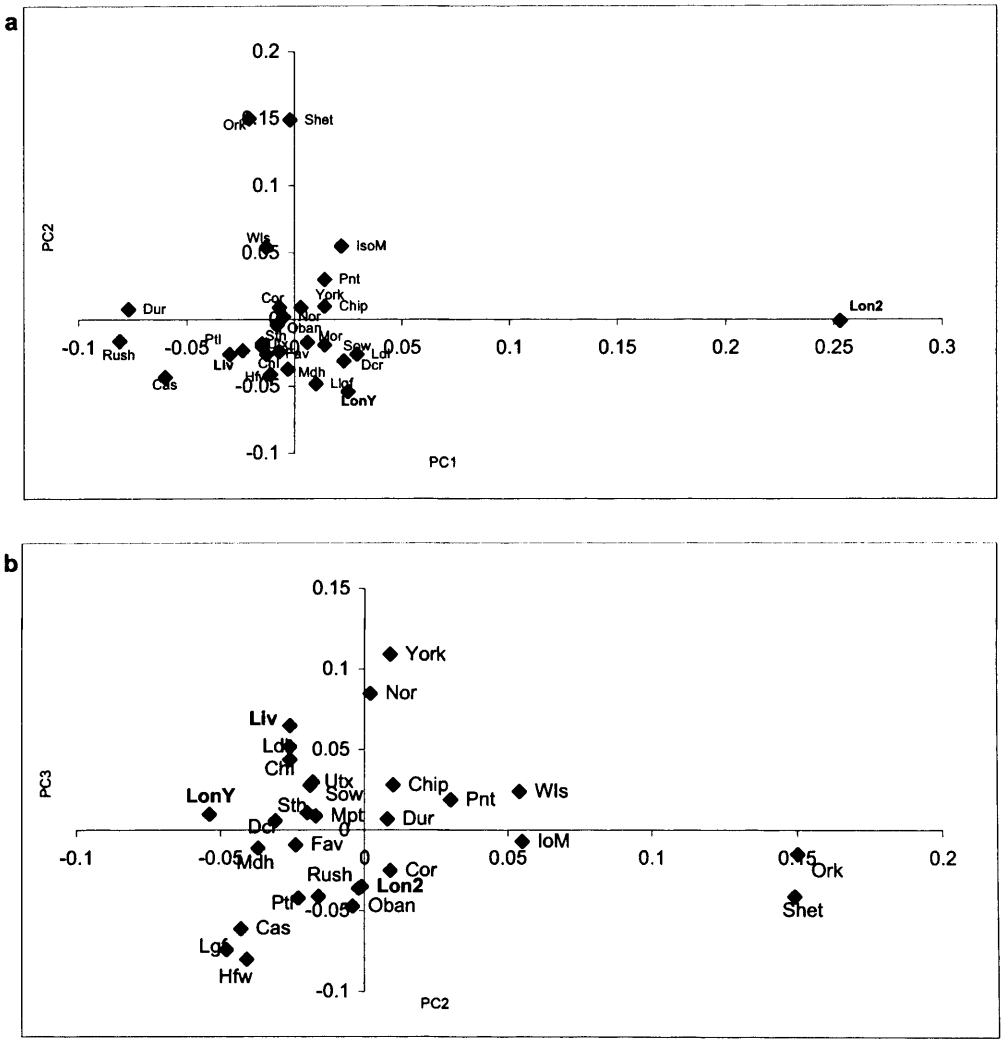


Figure 4.4. Y-Chromosome PC Plots for British Cities and the BCD Using Hg+1 Frequencies. (a) PCs 1 and 2, PC1 explains 23.8% of the variation and PC2 18.6%. (b) PCs 2 and 3, PC3 explains 16% of the variation. Abbreviations as Table 4.8, additionally: Lon2 = London2, Liv = Liverpool

Table 4.9. Y-Chromosome Pairwise Fst comparisons: British Cities and BCD Using Hg+1 Frequencies

	Shet	*Ork*	*Dur*	*Wls*	*Sth*	*Ptl*	*Oban*	*Mpt*	*Pnt*	*IoM*	*York*	*Sow*	*Utx*	*Ldl*	*Lgf*	*Rsh*	*Cas*	*Nor*	*Hfw*	*Chp*	*Fav*	*Mdh*	*Dcr*	*Cor*	*Chl*	*LonY*	*Lon2*	*Liv*
Shet	-																											
Ork	-0.001	-																										
Dur	0.077	0.037	-																									
Wls	0.008	0.007	0.077	-																								
Sth	0.020	0.006	0.009	0.022	-																							
Ptl	0.010	0.015	0.067	0.004	0.008	-																						
Oban	0.008	0.018	0.067	0.014	0.003	-0.015	-																					
Mpt	0.010	0.010	0.063	-0.003	0.003	-0.010	-0.005	-																				
Pnt	0.000	0.009	0.079	0.000	0.011	0.001	0.002	-0.004	-																			
IoM	-0.003	0.009	0.092	-0.001	0.016	0.002	0.000	-0.003	-0.010	-																		
York	0.028	0.008	0.045	0.004	0.009	0.032	0.041	0.011	0.010	0.018	-																	
Sow	0.012	0.004	0.053	-0.002	0.006	-0.002	0.011	-0.005	0.000	0.003	0.000	-																
Utx	0.014	0.004	0.029	0.006	-0.009	-0.001	0.004	-0.003	0.001	0.009	-0.001	-0.005	-															
Ldl	0.010	0.007	0.053	0.002	-0.005	0.005	0.005	-0.006	-0.008	-0.005	-0.005	-0.007	-0.008	-														
Lgf	0.037	0.058	0.136	0.042	0.047	0.003	-0.006	0.020	0.022	0.017	0.089	0.040	0.039	0.034	-													
Rsh	0.059	0.029	-0.001	0.052	0.010	0.032	0.035	0.039	0.062	0.069	0.049	0.034	0.020	0.045	0.089	-												
Cas	0.035	0.024	0.028	0.031	-0.001	-0.004	-0.007	0.010	0.030	0.033	0.047	0.019	0.006	0.022	0.027	-0.001	-											
Nor	0.026	0.008	0.038	0.006	0.007	0.023	0.032	0.008	0.012	0.020	-0.013	-0.001	-0.001	-0.001	0.073	0.036	0.034	-										
Hfw	0.032	0.045	0.098	0.041	0.025	-0.001	-0.015	0.016	0.022	0.019	0.077	0.034	0.027	0.027	-0.009	0.058	0.004	0.062	-									
Chp	0.004	0.006	0.068	0.002	0.000	0.006	0.007	-0.005	-0.010	-0.010	0.000	-0.002	-0.002	-0.014	0.036	0.060	0.033	0.005	0.032	-								
Fav	0.007	0.007	0.041	0.010	-0.011	-0.011	-0.009	-0.006	-0.004	0.001	0.014	-0.004	-0.010	-0.009	0.017	0.026	0.000	0.011	0.006	-0.006	-							
Mdh	0.012	0.015	0.073	-0.001	0.009	-0.012	-0.006	-0.010	-0.004	-0.004	0.017	-0.004	0.000	-0.004	0.013	0.044	0.010	0.014	0.010	-0.003	-0.006	-						
Dcr	0.009	0.006	0.046	0.004	-0.008	-0.007	-0.006	-0.008	-0.004	-0.003	0.007	-0.008	-0.009	-0.012	0.021	0.029	0.005	0.006	0.012	-0.007	-0.013	-0.007	-					
Cor	0.003	0.006	0.049	0.008	-0.010	-0.009	-0.015	-0.007	-0.004	-0.006	0.020	0.002	-0.005	-0.007	0.011	0.031	0.000	0.016	0.001	-0.008	-0.015	-0.006	-0.012	-				
Chl	0.026	0.006	0.022	0.012	-0.004	0.018	0.022	0.007	0.013	0.020	-0.006	0.000	-0.005	-0.002	0.063	0.019	0.018	-0.004	0.047	0.005	0.003	0.012	0.000	0.008	-			
LonY	0.018	0.009	0.023	0.022	-0.012	0.007	0.007	0.005	0.010	0.018	0.011	0.004	-0.007	-0.003	0.044	0.019	0.007	0.008	0.027	0.004	-0.008	0.009	-0.004	-0.002	0.000	-		
Lon2	0.016	0.035	0.134	0.024	0.043	0.024	0.021	0.017	0.006	-0.002	0.043	0.019	0.034	0.010	0.030	0.109	0.067	0.045	0.037	0.004	0.021	0.015	0.015	0.016	0.045	0.038	-	
Liv	0.027	0.005	0.023	0.000	-0.005	0.009	0.019	-0.002	0.010	0.018	-0.013	-0.006	-0.010	-0.005	0.067	0.016	0.013	-0.012	0.052	0.003	0.001	0.004	-0.002	0.006	-0.011	0.000	0.050	-

Abbreviations as Table 4.8, additionally, Lon2 = London2, Liv = Liverpool

Table 4.10. Y-Chromosome Haplogroup and Haplotype Diversity (*h*) for British Cities and British Comparison Populations, Ordered by Hg *h* score

				Haplogroups			*Haplotypes*		
Population	*n*	*# hgs*	*#hts*	*h*	*Standard Error (+/-)*	*Rank*	*h*	*Standard Error (+/-)*	*Rank by ht*
York	46	5	24	0.79	0.036	1	0.9575	0.0131	3
Nor	121	6	59	0.77	0.0232	2	0.9493	0.0119	6
Chl	128	8	54	0.76	0.0247	3	0.9542	0.0102	4
Ork	121	5	51	0.75	0.0269	4	0.9525	0.0089	5
Sow	70	6	40	0.75	0.0409	5	0.9445	0.0169	10
Liv	**44**	**4**	**26**	**0.74**	**0.0397**	6	**0.9461**	**0.0199**	8
Ldl	57	5	28	0.73	0.0493	7	0.9398	0.0201	12
LonY	**93**	**8**	**48**	**0.72**	**0.0344**	8	**0.9595**	**0.0103**	2
Chp	51	6	25	0.72	0.056	9	0.9325	0.0203	15
Utx	84	6	39	0.72	0.0362	10	0.9461	0.0122	7
Dcr	73	7	39	0.7	0.0453	11	0.9456	0.0155	9
Wls	88	3	32	0.7	0.0387	12	0.9287	0.0164	18
Sth	43	5	23	0.69	0.0472	13	0.9355	0.0252	13
Pnt	90	6	36	0.69	0.0458	14	0.9296	0.0155	17
Shet	63	3	27	0.69	0.0492	15	0.9437	0.014	11
Lon2	**63**	**8**	**33**	**0.68**	**0.0631**	16	**0.9307**	**0.022**	16
Mpt	95	6	45	0.68	0.0412	17	0.9286	0.0165	19
Fav	55	5	28	0.68	0.0525	18	0.9138	0.0295	23
Dur	51	3	22	0.67	0.044	19	0.8894	0.0299	28
IoM	62	4	26	0.67	0.0581	20	0.926	0.0183	20
Mdh	80	7	36	0.66	0.0478	21	0.9161	0.019	22
Cor	52	4	24	0.65	0.0565	22	0.9344	0.0184	14
Rush	76	4	39	0.64	0.0313	23	0.9628	0.0098	1
Ptl	41	4	18	0.63	0.0644	24	0.8939	0.0325	27
Oban	42	4	18	0.59	0.0677	25	0.9106	0.0263	25
Cas	43	2	20	0.58	0.0457	26	0.9136	0.0271	24
Hfw	59	4	25	0.52	0.0571	27	0.9223	0.0232	21
Lgf	80	6	26	0.5	0.0565	28	0.8987	0.0239	26

Bold text highlights the 3 British cities

Abbreviations as Table 4.8, additionally, Lon2 = London2, Liv = Liverpool

pairwise Fst between LondonY and another population is with Llangefni (0.04356). However, few of the comparisons are significant. Haplogroup and haplotype diversity (*h*, Table 4.10) does not rank any of the cities highest compared to the BCD. Hg diversity ranks Liverpool 6th, LondonY 8^{th,} , and London2 16th. York has the highest hg diversity (0.7855) and Llangefni the lowest (0.5022). This result somewhat contradicts the fact that LondonY and London2 both have more hgs present than most of the populations ranked above them, but reflects the relatively low frequency of several hgs in London. Haplotype diversity calculated for the same populations does not drastically alter the ranking of the cities, except LondonY which moves to second place and better reflects the number of different haplotypes in the sample. Therefore a count of the hgs and haplotypes can be more informative (Table 4.11).

4.3.2. Y-chromosome Comparisons with Europe

The BCD represents a very detailed picture of Y-chromosome diversity in one small part of Europe, therefore PC plots drawn with the entire BCD and the RD have very poor resolution for non-British populations (see Appendix, Figure A.2). To control for this effect the BDC populations were clustered together into "AllBrit"; as the relationship between LondonY and the BCD is known the loss of resolution for British Y-chromosomes is not important. The PC plot drawn with these data (Figure 4.5) shows 3 main poles towards which populations are drawn: Northern Africa (N Africa and Algerian samples), Western Europe (Basques, Cornish), and Finno-Ugric/Eastern [Indo-European] speakers Finland, Lithuania etc). The North African populations and Finnish/Lithuania group are clear outliers within European Y-chromosome diversity. The Basques and Cornish appear as less extreme outliers because they simply fall at the limit of a trend followed by most of the western European populations. The positioning of LondonY, London2, and Liverpool confirms the findings above that the overall hg composition of these cities is very similar to that in Britain. Due to the hg clustering methodology that was applied to the data for comparisons with the

Table 4.11. Y-Chromosome Haplogroups and Modal Haplotypes Encountered in the British Cities and British Comparison Populations

Population \ Hg	*E3b*	*F*(xIJK)*	*J*(xJ2)*	*J2*	*I*(xI1b2)*	*2.47+1*	*I1b2*	*K*(xPN3)*	*N3*	*P*(xR1)*	*R1*(xR1a1)*	*AMH+1*	*R1a1*	*3.65+1*	*DE*(xE3b)*	*n*
LondonY	**3**	**4**	**2**	**3**	**2**	**8**	**-**	**2**	**-**	**-**	**26**	**41**	**-**	**2**	**-**	**93**
London2	**3**	**3**	**2**	**1**	**2**	**3**	**-**	**-**	**-**	**-**	**4**	**35**	**6**	**1**	**3**	**63**
Liverpool	**-**	**1**	**-**	**1**	**7**	**5**	**-**	**-**	**-**	**-**	**12**	**18**	**-**	**-**	**-**	**44**
Shetland	-	-	-	-	3	3	-	-	-	-	11	32	4	10	-	63
Orkney	-	-	-	-	9	8	1	-	-	2	28	50	9	14	-	121
Durness	-	-	-	-	2	5	-	-	-	-	24	17	1	2	-	51
Western Isles	-	-	-	-	16	6	-	-	-	-	15	43	3	5	-	88
Stonehaven	-	1	-	1	1	5	-	-	-	-	14	20	2	-	-	44
Pitlochry	-	-	-	3	4	-	-	-	-	-	10	23	-	1	-	41
Oban	-	1	-	-	2	1	-	-	-	-	11	25	1	1	-	42
Morpeth	-	2	1	3	11	6	-	-	-	-	20	49	2	1	-	95
Penrith	3	1	-	2	7	9	-	-	-	-	14	47	2	5	-	90
IoM	1	-	-	-	5	5	-	-	-	-	9	34	5	3	-	62
York	2	1	-	-	7	8	-	-	-	-	9	17	1	1	-	46
Southwell	4	1	-	4	9	3	-	-	-	-	14	31	3	1	-	70
Uttoxeter	3	1	-	3	7	8	-	-	-	-	22	38	-	2	-	84
Llanidloes	3	2	-	1	4	7	-	-	-	-	11	27	2	-	-	57
Llangefni	3	-	-	1	3	-	-	1	-	-	17	54	1	-	-	80
Rush	-	-	-	-	7	-	2	-	-	-	33	31	1	2	-	76
Castlerea	-	-	-	-	3	-	1	-	-	-	16	23	-	-	-	43
Norfolk	4	3	-	2	17	17	-	-	-	-	27	46	2	3	-	121
Haverfordwest	2	-	-	-	1	-	1	-	-	-	16	38	1	-	-	59
Chippenham	-	1	-	2	3	7	1	-	-	-	8	25	3	1	-	51
Faversham	2	-	-	3	2	4	-	-	-	-	14	28	1	1	-	55
Midhurst	1	-	1	3	9	4	2	-	-	-	16	43	1	-	-	80
Dorchester	3	1	1	2	5	5	-	-	-	-	17	36	3	-	-	73
Cornwall	-	-	-	1	2	4	-	-	-	-	13	28	3	1	-	52
Channel Islands	5	2	1	1	13	14	4	-	-	-	34	50	3	1	-	128

Modified from Table 2.9. Shown are the counts of all of the hgs and modal haplotypes found in the 2 London samples, Liverpool and the BCD

RD London2 does not fall as an outlier in this PC. The YAP+ chromosomes of London2 are placed with M35 derived chromosomes into one group; as M35 derived chromosomes are more common in Britain (and Europe) than YAP+ chromosomes (Table 4.11) London2 is thus not an outlier to Britain.. Based on the results of the exact test of population differentiation, LondonY shows less differentiation with Western European populations than other European groups; this is also true for London2 and Liverpool (Table 4.12). It is interesting to note that with the lower level of hg resolution that is used in comparisons with the RD, LondonY and London2 are not significantly different. Pairwise Fst scores indicate that the most differentiation between LondonY and the RD populations is with North Africa (0.58857) and Algeria (0.48775) (Table 4.13), which is not surprising given that North Africa and Algeria are clear outliers in Figure 4.5. When the pairwise Fst scores are ordered geographically it reveals a general trend for LondonY to be less differentiated compared to the north- and south-western European populations than the other groups (Central and Eastern Europe, Middle East, and Africa). Amongst the lowest pairwise Fst scores are those with other British populations.

The British cities are ranked low in terms of hg diversity compared to most of the populations of the RD. However, most of the populations that are ranked towards the lower end of the range of diversities are western European populations (Table 4.14), so the cities are not abnormal within the context of western Europe. The bias of *h* that was seen in section 4.2.9 above also has to be taken into account.

4.3.3. Y-chromosome Hg Distributions

Three hgs, R1*(xR1a1), I*(xI1b2), and R1a1 comprise the highest frequency hgs in most British populations with an assortment of other hgs being found at low frequencies (Table 4.11). In LondonY the former two of these hgs are also the commonest observed, however R1a1 chromosomes are only the 4th commonest. LondonY has 3 hgs that are rarely seen in the BCD: J*(xJ2),

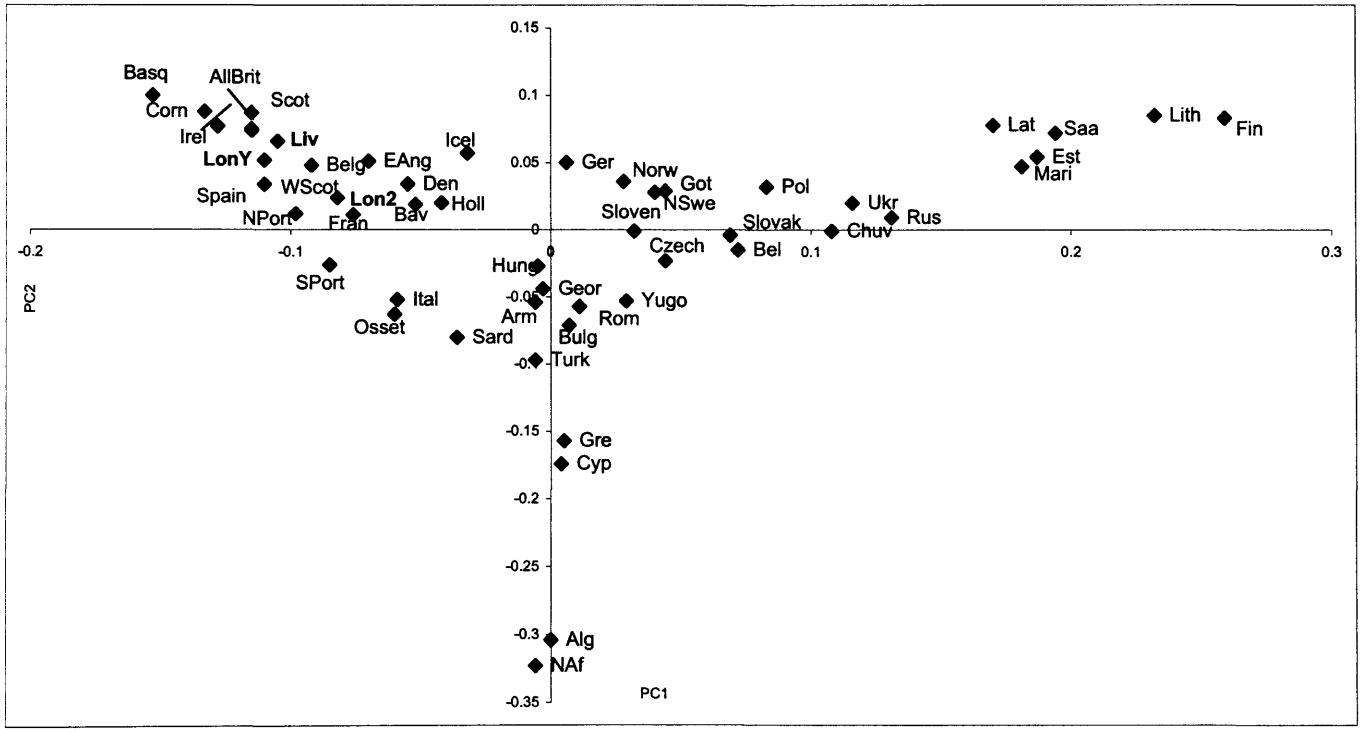


Figure 4.5. Y-Chromosome PC Plot of the British Cities and RD Using Hg Frequencies. PC1 explained 39.1% of the variation and PC2 explained 29.5%. Abbreviations for Rosser Dataset as follows: Alg = Algeria, Arm = Armenia, Basq = Basque, Bav = Bavaria, Bel = Belarus, Belg = Belgium, Bulg = Bulgaria, Chuv = Chuvash, Corn = Cornwall, Cyp = Cyprus, Czech = Czech Republic, Den = Denmark, EAng = East Anglia, Est = Estonia, Fin = Finland, Fran = France, Geor = Georgia, Ger = Germany, Got = Gotland, Gre = Greece, Holl = Holland, Hung = Hungary, Icel = Iceland, Ire = Ireland, Ital = Italy, Lat = Latvia, Lith = Lithuania, Mari = Mari, NPort = Northern Portugal, NAf = Northern Africa, NSw = Northern Sweden, Norw = Norwegian, Osset = Ossetia, Pol = Poland, Rom = Romania, Rus = Russia, Saa = Saami, Sard = Sardinia, Scot = Scotland, Slovak = Slovakia, Sloven = Slovenia, SPort = Southern Portugal, Spai = Spain, Turk = Turkey, Ukr = Ukraine, WScot = Western Scottish, Yugo = Yugoslavia. Other abbreviations as Table 4.8, additionally, AllBrit = All Britain, the entire BCD combined

Table 4.12. Y-Chromosome Exact Test of Population Differentiation: British Cities and British and European Comparison Datasets Using Hg Frequencies Ordered by Geographical Location

	North-western Europe											
Population	*Iceland*	*Wscot*	*Scotland*	*Cornwall*	*Eanglia*	*Ireland*	*Saami*	*NSweden*	*Gotland*	*Norway*	*Denmark*	*Finland*
LondonY	**0.002**	**0.026**	0.359	0.294	**0.001**	**0.016**	**0.000**	**0.000**	**0.000**	**0.000**	**0.041**	**0.000**
London2	**0.034**	**0.007**	0.107	**0.003**	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.046**	**0.000**
Liverpool	**0.001**	0.064	**0.045**	0.134	0.141	0.154	**0.000**	**0.000**	**0.000**	**0.000**	0.185	**0.000**

	North-western Europe						South-western Europe					
Population	*Estonia*	*Bavaria*	*Germany*	*Holland*	*France*	*Belgium*	*Italy*	*Sardinia*	*Basque*	*Span*	*SPort*	*NPort*
LondonY	**0.000**	**0.002**	**0.000**	**0.000**	0.189	0.694	**0.000**	**0.008**	0.600	0.438	**0.047**	**0.032**
London2	**0.000**	0.506	**0.015**	**0.033**	0.549	0.076	**0.001**	**0.049**	0.064	0.174	0.166	**0.000**
Liverpool	**0.000**	**0.004**	**0.001**	**0.011**	0.077	0.653	**0.000**	**0.006**	**0.049**	**0.048**	**0.004**	**0.017**

	Central and Eastern Europe											
Population	*Russia*	*Belarus*	*Ukraine*	*Mari*	*Chuvash*	*Georgia*	*Ossetia*	*Latvia*	*Lithuania*	*Czech*	*Slovakia*	*Romania*
LondonY	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
London2	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Liverpool	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

	Central and Eastern Europe									Middle East	Africa
Population	*Yugoslav*	*Slovenia*	*Hungary*	*Poland*	*Armenia*	*Turkey*	*Cyprus*	*Greece*	*Bulgaria*	*Algeria*	*NAfrica*
LondonY	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
London2	**0.000**	**0.000**	**0.024**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.000**	**0.000**
Liverpool	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

	British Comparison Dataset											
Population	*Shet*	*Ork*	*Dur*	*Wls*	*Sth*	*Ptl*	*Oban*	*Mpt*	*Pnt*	*IoM*	*York*	*Sow*
LondonY	**0.000**	**0.000**	0.274	**0.002**	0.727	0.820	0.273	0.361	0.245	**0.028**	**0.046**	0.540
London2	**0.009**	**0.000**	0.054	**0.003**	0.119	0.069	**0.025**	**0.001**	0.313	0.145	**0.030**	0.575
Liverpool	**0.000**	**0.000**	**0.018**	0.087	0.260	**0.037**	**0.017**	0.474	0.167	**0.007**	0.286	0.164

continued

Table 4.12. continued

	British Comparison Dataset											
Population	*Utx*	*Nor*	*Chp*	*Fav*	*Mdh*	*Dcr*	*Cor*	*Chl*	*Hfw*	*Ldl*	*Lgf*	*Rsh*
LondonY	0.849	**0.018**	0.248	0.900	0.825	0.889	0.342	0.112	**0.025**	0.529	**0.047**	**0.072**
London2	0.108	**0.007**	0.131	0.337	**0.010**	0.346	0.125	**0.011**	**0.001**	0.248	**0.001**	**0.003**
Liverpool	0.502	0.557	0.281	0.070	0.567	0.204	**0.026**	0.635	**0.000**	0.444	**0.000**	**0.007**

	Cities		
Population	*Cas*	*London*	*London2*	*Liverpool*
LondonY	0.305	-		
London2	**0.004**	0.080	-	
Liverpool	**0.035**	0.275	**0.004**	-

Bold text indicates $p<0.05$

Abbreviations: NSweden = Northern Sweden, Czech = Czech Republic, Yugoslav = Yugoslavia, WScot = Western Scotland, EAnglia = East Anglia, SPort = Southern Portugal, NPort = Northern Portugal, NAfrica = A Africa. Abberviations for the British Comparison Dataset as in Figure 4.3

Table 4.13. Y-Chromosome Pairwise Fst comparisons: British Cities and the RD Using Hg frequencies Ordered by Geographical Location

	North-western Europe											
	Iceland	*Wscot*	*Scotland*	*Cornwall*	*Eanglia*	*Ireland*	*Saami*	*NSweden*	*Gotland*	*Norway*	*Denmark*	*Finland*
LondonY	0.101	-0.001	-0.003	0.005	0.039	0.007	0.378	0.246	0.331	0.205	0.059	0.466
London2	0.041	0.014	0.023	0.054	0.023	0.056	0.289	0.168	0.260	0.118	0.031	0.390
Liverpool	0.056	0.005	0.036	0.025	0.005	0.038	0.331	0.174	0.248	0.158	0.017	0.439

	North-western Europe						South-western Europe					
	Estonia	*Bavaria*	*Germany*	*Holland*	*France*	*Belgium*	*Italy*	*Sardinia*	*Basque*	*Span*	*SPort*	*NPort*
LondonY	0.309	0.051	0.134	0.087	0.025	0.002	0.075	0.213	0.039	-0.004	0.030	0.006
London2	0.236	0.006	0.052	0.042	-0.001	0.005	0.036	0.119	0.092	0.003	0.000	0.004
Liverpool	0.289	0.037	0.116	0.046	0.006	-0.007	0.080	0.155	0.111	0.023	0.044	0.017

	Central and Eastern Europe											
	Russia	*Belarus*	*Ukraine*	*Mari*	*Chuvash*	*Georgia*	*Ossetia*	*Latvia*	*Lithuania*	*Czech*	*Slovakia*	*Romania*
LondonY	0.343	0.336	0.392	0.339	0.284	0.275	0.126	0.352	0.424	0.259	0.304	0.227
London2	0.246	0.225	0.287	0.243	0.179	0.210	0.080	0.241	0.323	0.153	0.194	0.139
Liverpool	0.321	0.289	0.328	0.323	0.246	0.206	0.136	0.331	0.402	0.234	0.283	0.191

	Central and Eastern Europe									Middle East	Africa
	Yugoslav	*Slovenia*	*Hungary*	*Poland*	*Armenia*	*Turkey*	*Cyprus*	*Greece*	*Bulgaria*	*Algeria*	*NAfrica*
LondonY	0.301	0.241	0.172	0.333	0.197	0.208	0.317	0.298	0.270	0.488	0.589
London2	0.223	0.143	0.078	0.232	0.140	0.151	0.226	0.197	0.171	0.384	0.515
Liverpool	0.237	0.206	0.136	0.317	0.157	0.184	0.289	0.269	0.208	0.494	0.615

	Cities		
	LondonY	*London2*	*Liverpool*
LondonY	-		
London2	0.009	-	-
Liverpool	0.011	0.031	0.000

Abbreviations as Table 4.8

Table 4.14. Y-Chromosome Haplogroup Diversity *(h)* for British Cities and the RD Using Hg Frequencies

Population	*h*	*Standard Error (+/-)*	*Rank*	*Population*	*h*	*Standard Error (+/-)*	*Rank*
Chuv	0.8824	0.0407	1	Geor	0.6815	0.0406	26
Rom	0.802	0.0214	2	Ukr	0.6809	0.062	27
Gre	0.7984	0.0299	3	Bav	0.6772	0.0397	28
Czech	0.7794	0.034	4	Icel	0.6587	0.0447	29
Turk	0.7781	0.0151	5	Lith	0.6558	0.0467	30
Mari	0.7757	0.0314	6	Den	0.6474	0.0447	31
Hung	0.773	0.0267	7	Pol	0.6448	0.039	32
Cyp	0.7727	0.03	8	Sport	0.6372	0.0576	33
Bulg	0.7717	0.0593	9	Fran	0.6359	0.0639	34
Est	0.7578	0.0163	10	Got	0.5987	0.0564	35
Arm	0.7572	0.0204	11	**Lon2**	**0.5863**	**0.0636**	36
Sloven	0.7383	0.025	12	Eang	0.5788	0.0279	37
Sard	0.7333	0.0764	13	Fin	0.5689	0.06	38
Ger	0.731	0.0465	14	Nport	0.5523	0.0277	39
Bel	0.728	0.0435	15	Algerian	0.5442	0.05	40
Ital	0.7277	0.0311	16	Belgian	0.537	0.0497	41
Norw	0.727	0.0243	17	Spanish	0.4759	0.0496	42
Rus	0.7259	0.0327	18	**LonY**	**0.4584**	**0.0584**	43
Slovak	0.7168	0.042	19	**Liv**	**0.4577**	**0.0598**	44
Lat	0.7112	0.0426	20	Wscot	0.4366	0.0474	45
NSw	0.7092	0.0447	21	Scot	0.3643	0.0874	46
Yugo	0.7051	0.0376	22	Ire	0.3288	0.0339	47
Osset	0.7003	0.0423	23	Naf	0.324	0.05	48
Saa	0.6959	0.0326	24	Corn	0.2965	0.0696	49
Holl	0.6873	0.0304	25	Basq	0.1477	0.0888	50

Bold text highlights the 3 British cities

Abbreviations as Table 4.8

Note, calculations were not performed on AllBrit as more detailed calculations are presented in Table 4.10

K*(xPN3), and G (derived at M201, but considered as F*(xIJK), i.e. M89* in other analyses). J*(xJ2) chromosomes are found in two LondonY individuals, and one each in Morpeth, Midhurst, Dorchester and the Channel Islands. JxJ2 chromosomes often comprise 30% or more of the sampled lineages in Jewish,Turkish and Arab populations from the Middle East (Rosser *et al.* 2000; Bosch *et al.* 2001; Semino *et al.* 2000; Gonçalves *et al.* 2003), where it is thought to have originated (Gonçalves *et al.* 2003). One KxPN3 chromosome is found in the BCD in an individual from Llangefni and 2 LondonY individuals. KxPN3 defines a relatively heterogenous collection of chromosomes, therefore making statements about the possible geographic origins of this clade is difficult. G chromosomes are common in the Middle East, as well as in a small isolated population in the Iberian Peninsula (Maca-Meyer *et al.* 2003).

Additionally one LondonY-chromosome has an unknown status at M89 or M201 and is underived at all of the markers tested here except SRY_{10831a}. Haplotype information was used to gain more information about this chromosome using the ystr.org database and the 5 microsatellites that are comparable between this study and the database (DYS19, DYS390, DYS391, DYS392, DYS393). A worldwide search showed that the highest frequency of the haplotype was in African populations and populations of African origin: 5.1% in London Afro-Caribbeans, 5.4% in Mozambique, 2.5% in West Africans, 3.6% in African Americans from Missouri, and 3.33% in African Americans from New York, suggesting an African origin for this haplotype. Tentatively this chromosome can be placed into hg B due to its African-specific distribution (Underhill *et al.* 2000)

4.3.4. mtDNA Comparisons with Britain

A PC plot of LondonMT and the British mtDNA comparison populations drawn using the 1st and 2nd PCs (Figure 4.6a) shows that LondonMT falls within the main group of British populations, whilst Orkney and the Western Isles/Isle of Skye both fall as outliers (at different poles of the plot). PC1 accounts for 33.7%

of the variation, and X frequencies drive the positioning of populations along this axis, with Orkney having relatively high frequencies of X, although this hg is relatively rare in Britain. PC2 explained 28.8% of the variation and H/CRS frequencies are important (the Western Isles/Isle of Skye having relatively low frequencies). As the third principal component still explained 26.5% of the variation, a plot of PCs 2 and 3 was also drawn (Figure 4.6b). This reveals that LondonMT falls as an outlier on the third PC due to the relatively high frequency of L chromosomes, which are absent from the other British populations apart from the Scottish and English GD samples. L sequences comprise up to 100% of sub-Saharan African lineages (Torroni *et al.* 1996) and are rarely seen outside Africa.

An exact test of population differentiation shows that LondonMT has a significantly different hg composition from Scotland, England/Wales and the Western Isles of the HD (Table 4.15), but is not significantly different to either Ireland or Orkney, or the Scotland and England samples of the GD. The presence of hg L lineages in Scotland and England samples and LondonMT may explain the non significant differences.

4.3.5. mtDNA Comparisons with Europe

A PC plot was drawn using the entire datasets of Al-Zahery, Helgason and González (see Appendix, Figure A.3). The Indian, Central Asian, and Saami populations fall as extreme outliers in this plot and reduce resolution of the remaining populations to the extent that most fall into one indistinct cluster. To increase the resolution within these European populations, the Indian, Central Asian, and Saami samples were removed and the plots re-drawn (Figure 4.7), which dramatically increases resolution. The first PC shows an approximate north to south gradient with northern European populations falling at the positive pole of the axis (Germany, Austria/Switzerland, Iceland) and more southerly populations at the negative extreme (North Africa, Arabia, South Portugal), PC1 explained 40.5% of the variation and hg L is important in

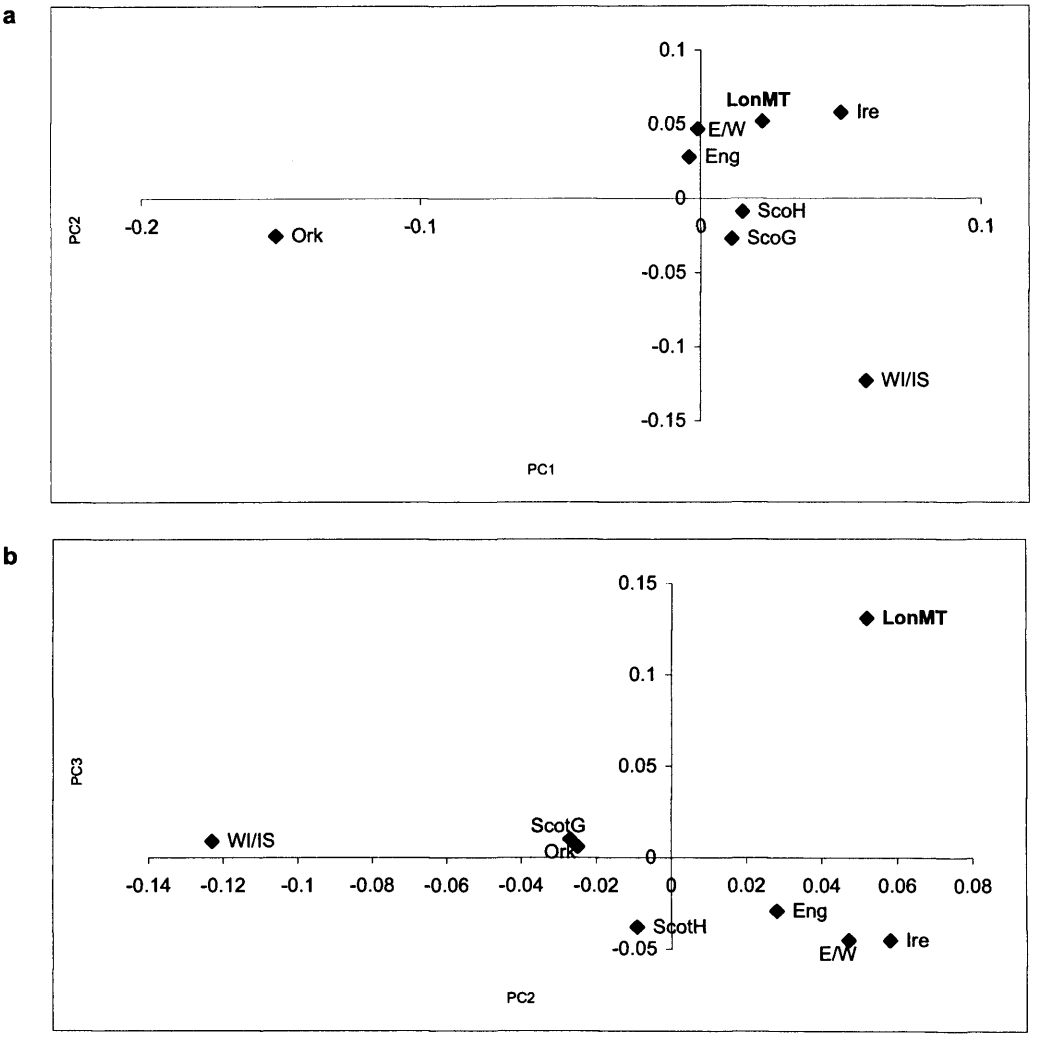


Figure 4.6. mtDNA PC Plots of LondonMT and British Comparison Populations Using Hg Frequencies. (a) PCs 1 and 2, PC1 explained 33.7% of the variation and PC2 explained 28.8%. (b) PCs 2 and 3, PC3 explained 26.5% of the variation. Abbreviations as follows: (Gonzalez *et al*. 2003 dataset) Eng = England, ScotG = Scotland; (Helgason *et al*. 2001 dataset) Ork = Orkney, WI/IS = Western Isles and Isle of Skye, ScotH = Scotland, E/W = England and Wales, Ire = Ireland.

Table 4.15. mtDNA Exact Test of Population Differentiation: LondonMT and Comparison Populations Using Hg Frequencies Ordered by Geographical Location

Region	North-western Europe										
Population	*Ice*	*Ire*	*WI/IS*	*Ork*	*Scot*	*ScotG*	*E/W*	*Eng*	*F/E*	*Scan*	*FU*
LondonMT	**0.011**	0.394	**0.020**	0.150	**0.013**	0.144	**0.025**	0.054	**0.038**	0.055	0.253

Region	North-western Europe									
Population	*Saa*	*Fin*	*Nor*	*GerA*	*NG*	*SG*	*Ger*	*A/S*	*Fra*	*F/I*
LondonMT	**0.000**	**0.013**	0.191	0.456	0.701	0.438	0.290	0.605	0.296	0.162

Region	South-western Europe					
Population	*Ita*	*NPort*	*CPort*	*SPort*	*S/P*	*Gal*
LondonMT	0.624	0.477	0.501	0.088	**0.026**	0.097

Region	Central and Eastern Europe					
Population	*E/R*	*Slav*	*Geo*	*Arm*	*Ana*	*B/T*
LondonMT	0.045	**0.032**	**0.007**	**0.039**	**0.000**	**0.043**

	Middle East					Asia		Africa
Population	*Iran*	*Ara*	*Syr*	*Pal*	*Iraq*	*C-Asia*	*Ind*	*NAf*
LondonMT	**0.000**	**0.000**	0.541	0.073	0.058	**0.000**	**0.000**	**0.000**

Bold text indicates $p<0.05$
Abbreviations as Figure 4.7

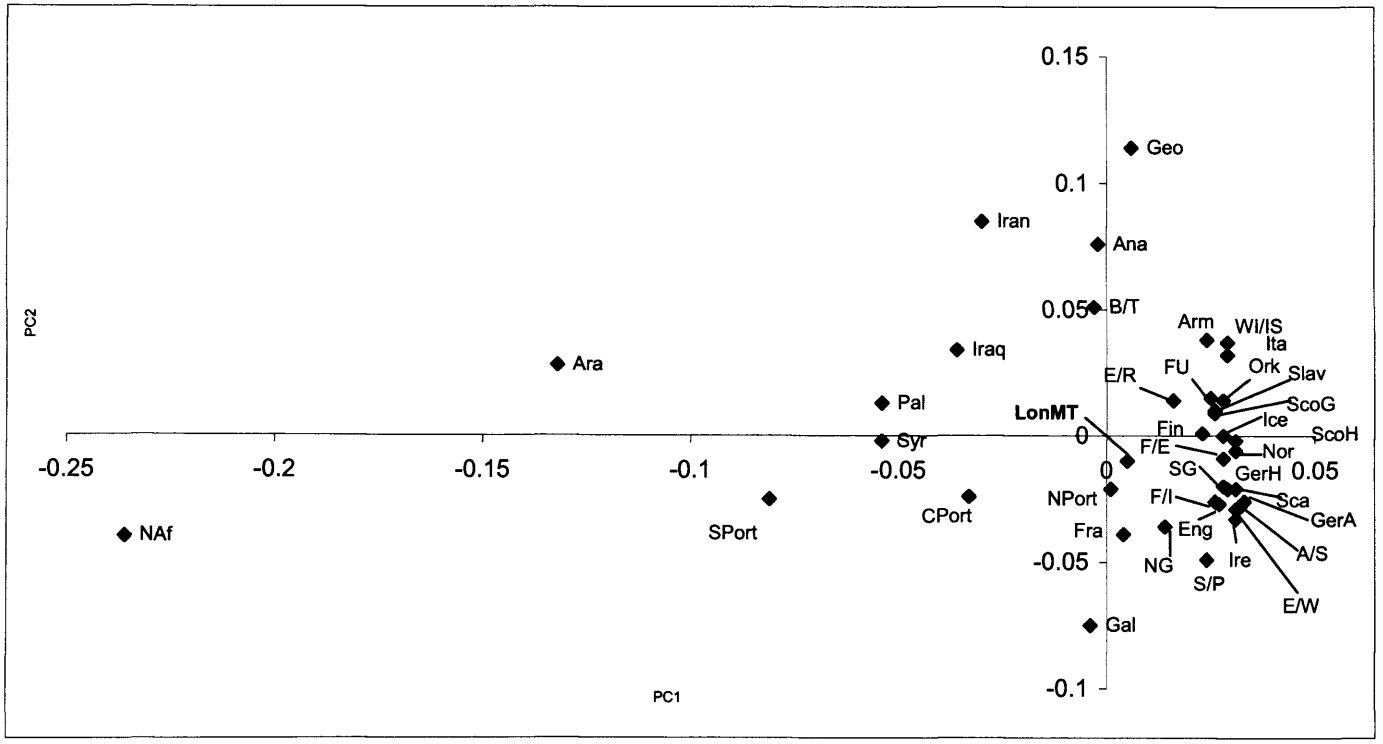


Figure 4.7. mtDNA PC Plot of LondonMT and European Comparison Populations Using Hg Frequencies. PC1 explained 40.5% of the variation and PC2 explained 20.1%. Abbreviations as in Figure 4.4, additionally: (Al-Zahery *et al.* 2003 dataset) Ara = Arabia, Syr = Syria, Pal = Palestine, Geo = Georgia, Arm = Armenia, Ana = Anatolia, Ita = Italy, Slav = Slavic, FU = Finno-Ugric speakers, GerA = Germany; (Gonzalez dataset) Fin = Finland, Nor = Norway, NG = North Germany, SG = South Germany, Fra = France, Gal = Galicia, NPort = North Portugal, CPort = North Portugal, SPort = South Portugal, NAf = North Africa; (Helgason dataset) A/S = Austria and Switzerland, E/R = European Russians, F/E = Finland and Estonia, F/I = France and Italy, GerH = Germany, Ice = Iceland, Scan = Scandinavia, B/T = Bulgaria and Turkey, S/P = Spain and Portugal, E/W = England and Wales.

separating populations, which is congruent with the relative placing of the most northern and southern populations on this axis. The position of LondonMT on this axis is noteworthy because of the presence of L chromosomes. Of the north western European populations LondonMT is the second furthest to be pulled towards North Africa, after France, with evidence for an increased L frequency compared to most of the rest of Europe. The second PC explained 20.1% of the variation; several hgs are important in the positioning of populations along this axis: H, U, and X with LondonMT being intermediately placed. An approximate east to west gradient is seen along this axis; the eastern Eurasian populations of Georgia, Iran and Anatolia are drawn towards the positive extreme of PC2 and are eastern Eurasian populations, whereas the western populations of Galicia and South Portugal are at the negative extreme. However, this geographical trend is not strict because the Western Isles/Isle of Skye sample is closer to the eastern Eurasian populations rather than the western populations.

The exact test of population differentiation reveals a trend for LondonMT to be more similar to western European populations (except Finno-Ugric speakers) than any other geographical group. LondonMT is significantly different to the two Asian populations (India and Central Asia), North Africa, the majority of the eastern Eurasian populations (Iran, Arabia, Georgia, Armenia, and Bulgaria/Turkey), and most of the Finno-Ugric speakers (Finland, Finland/Estonia, and Saami).

Three indices of diversity were calculated for LondonMT, θ_π, θ_k and H, and compared with the published results of Helgason *et al.* 2001 (Table 4.16). Regardless of which index is used, LondonMT ranks third, therefore, despite the biases of each of these indices (see section 4.2.9 above) it is a fair conclusion to make that LondonMT is amongst the most diverse European presented here.

4.3.6. mtDNA Hg Distributions

Table 4.16. mtDNA Gene Diversity (h) and Theta Parameters for LondonMT and European and African Populations Using Haplotypes

Population	*n*	*K*	*S*	*h*	θ_k	θ_π	*Rank h*	*Rank* θ_π
F/I	248	158	97	0.963	186.42	4.23	6	2
Ger	527	234	99	0.97	160.68	3.7	2	7
LonMT	117	86	72	0.97	144.9	4.15	3	3
Scand	645	243	108	0.937	141.36	3.52	11	10
E/W	429	183	91	0.934	120.18	3.35	14	14
Scot	891	250	102	0.956	115.11	3.73	8	6
S/P	352	154	95	0.935	103.85	3.26	12	15
B/T	102	71	70	0.977	102.25	4.34	1	1
A/S	187	93	70	0.958	72.84	3.55	7	9
E/R	215	90	59	0.934	57.69	3.44	15	12
W/I	197	79	53	0.968	48.43	3.75	4	5
Ice	467	114	67	0.966	47.76	3.96	5	4
Ork	152	67	55	0.946	45.24	3.37	10	13
Ire	128	61	50	0.922	45.05	2.87	16	17
F/E	202	75	59	0.949	42.74	3.49	9	11
IS	49	23	27	0.935	16.3	3.7	13	8
Saa	176	30	30	0.808	10.15	3.21	17	16

Table sorted in ascending order by thetak, values for all non-London populations are taken directly from Helgason et al. (2001). Abbreviations as Figure 4.7. n = Sample size, K = number of lineages, S = number of variable (segregating) sites

In Europe most mtDNA lineages can be placed into one of the following clusters of clades: HV, UK, TJ, and WIX (Finnilä *et al.* 2001, see also Figure 4.2), the distribution and frequencies of which were described in the Introduction. Hg frequencies in the LondonMT dataset (Table 4.7) broadly fit into the pattern of mtDNA diversity in Europe (Table 4.7) and Britain (Table 4.3). H is the commonest type and the remaining groups are found at lower frequencies. LondonMT additionally contains two L sequences (L1b and L3b) with the following sequence motifs: 93-126-187-189-223-264-270-278-311-318T and 189G-223-274-278-294-362 respectively. L1b sequences are concentrated in Western Africa with some also found in Central and Northern Africa, L3b is also common in West Africa with some occurrences in North Africa and the Middle East (Salas *et al.* 2002). One sequence has also been assigned to hg C with the sequence motif 223-249-295-298-311-325-327. C is an East-Asian specific hg (Derenko *et al.* 2003).

4.3.7. Relative Y-Chromosome and mtDNA Diversity

Table 4.17 summarises the AMOVA results. British Y-chromosomes exhibit increased differentiation between groups (1.5%) compared to mtDNA lineages (0.58%), and concomitantly the Y-chromosome has reduced within population diversity (98.49%) compared to mtDNA (99.42%) (Table 4.17a). When the AMOVA is performed using two groups (Table 4.17b) this trend is not continued however. Between group variation, i.e. the percentage of variation between LondonY or LondonMT and the rest of Britain, is approximately equal for LondonY (-0.76%) and LondonMT(-0.3%). A negative score usually indicates the absence of genetic structure (Arlequin 2.000 FAQs, http://lgb.unige.ch/arlequin/software/2.000/doc/faq/faqlist.htm).

4.4. Discussion

Table 4.17. Relative Y-Chromosome and mtDNA Hg Diversity Assessed by AMOVA

(a)

	Source of Variation	
Genetic System	*Among Populations*	*Within Populations*
Y chromosome	1.51	98.49
mtDNA	0.58	99.42

(b)

	Source of Variation		
Genetic System	*Among Groups*	*Among Populations within Groups*	*Within Populations*
Y chromosome	-0.76	1.63	99.13
mtDNA	-0.3	0.61	99.69

(a) The British Y Chromosome and mtDNA populations (including LondonY and LondonMT) have been placed into one group. (b) The British Y Chromosome (and mtDNA) populations have each been placed into "Group 1" and LondonY (and LondonMT) into "Group 2".

The aim of this chapter was to study Y-chromosome and mtDNA genetic diversity in London in relation to historical, archaeological, and Census records detailing migration to the city. This was carried out within the context of known patterns of Y-chromosome and mtDNA diversity in Britain and Europe. In summary, the findings of this chapter reveal that London Y-chromosomes and mtDNAs are similar to Britain and other western European populations. However the presence in London of several Y-chromosome and mtDNA hgs that are otherwise rare in Britain and western Europe provide the main evidence for a more diverse genetic history than other British and European populations. This is in keeping with predictions about London from historical, archaeological and Census records. Interestingly an independently collected sample of Londoners, typed for the same Y-chromosome markers used in this chapter, revealed that the two samples were significantly different from each other. In contrast to the findings for London Y-chromosomes, the metropolitan district of Liverpool showed less diversity. It was also shown that the paternal and maternal histories of Londoners were comparable. Finally issues associated with adequate sampling from ethnic minorities are considered as are some of the limitations imposed on the sampling strategy. These points will now be considered in more detail.

London and the rest of Britain have experienced a similar history of contact with continental populations, from Romans to Anglo-Saxons, Vikings and Normans (e.g. Ackroyd 1999), which must explain the lack of structure between LondonY and LondonMT and the rest of Britain. Considering the Y-chromosome first, PC analysis shows that LondonY does not fall as an outlier, compared to Orkney and Shetland for example. These results are confirmed by the exact test of population differentiation and pairwise Fst comparisons which indicate only slight structure between LondonY and the BCD; some structure is expected because of the geographic structure seen between British (Chapter 2) and European Y-chromosomes (e.g. Rosser *et al.* 2000). There is some indication of less structure with south coast populations suggesting that LondonY is typical of a south coast Y-chromosome population. The low Y-chromosome differentiation for most of the BCD, based on low Fst scores, may be a function of the small geographic distance between all British populations; it has been reported that Y-

chromosome pairwise Fst values in Europe tend to increase with increasing geographic distance (Malaspina *et al.* 2000). Due to the overall similarity between LondonY and the BCD it is not surprising that comparisons with the RD of European populations shows that LondonY is not an outlier within Europe. Indeed the analyses suggest that the genetically closest Y-chromosome populations are British and Western European populations.

Analysis of LondonMT (hg frequencies) and British populations echoes the Y-chromosome findings. This is seen in the PC plot of PCs 1 and 2 and the exact test of population differentiation showing that LondonMT is not significantly different to the Ireland and Orkney samples of the HD, and Scotland and England of the GD. However, there is some differentiation between LondonMT and Britain; the plot of PCs 2 and 3 places LondonMT as a clear outlier, and LondonMT is significantly different from Scotland and England/Wales (HD). The two L lineages, which are considered to be sub Saharan lineages (Salas *et al.* 2002), hence rare in Europe, force LondonMT as an outlier in this PC analysis and probably explain the significant difference of LondonMT and Scotland and England/Wales. This in turn provides the first genetic evidence for the known ethnic diversity in London and some indication of more mtDNA than Y-chromosome structure (both of which are considered in more detail below).

Genetic signatures of the diverse ethnic history of London are more clearly seen in the less common hgs and haplotypes that do not affect the population genetic analyses above because of their low frequency. Most studies tend to ignore such hgs and haplotypes when investigating the origins of a population because they indicate recent gene flow, thus confound attempts to investigate historical events (see for example, Brehm *et al.* 2002). In the present study these haplotypes are interesting for the very reason they are normally excluded. LondonY has several low frequency hgs that are generally rare in the BCD and Western Europe. Hg G (placed into FxIJK for analysis as described above, but here considered as G); JxJ2; and KxPN3; and one chromosome, which has been tentatively placed into B, a hg most commonly seen in Africa (Underhill *et al.* 2000). The former three hgs are all present in the BCD at low frequency. G and JxJ2 chromosomes are common in the Middle East (Maca-Meyer *et al.* 2003; Rosser *et al.* 2000; Bosch

et al. 2001; Semino *et al.* 2000; Gonçalves *et al.* 2003) and found sporadically in western Europe. The paternal heritage of these men probably lies in the Middle East, but as both hgs are found in other British and Western European populations, albeit at low frequencies, it is not possible to conclude whether the chromosomes represent recent or ancient migration to Britain and/or Europe. KxPN3 defines a relatively heterogenous collection of chromosomes, therefore making conclusions about the possible geographic origins of this clade is difficult (Capelli *et al.* 2001), beyond the statement above that KxPN3 is rare in Britain, thus the higher frequency of this clade in LondonY points to the more diverse origins of the London gene pool. The hypothesised hg B chromosome, whose haplotype certainly suggests an African origin, is more than likely the result of recent immigration from either Africa or the Caribbean, rather than a signature of the Atlantic Slave trade. Although Britain played a central part in the African slave trade from the 17^{th} to 19^{th} centuries, few African slaves actually settled in Britain (Walvin 2000). Whilst all of these rare hgs are not common enough in London to alter the general similarity LondonY has with Britain, they indicate a more diverse heritage for London Y-chromosomes compared to the rest of Britain (although note that the potential hg B chromosome was not included in any analyses involving hgs and hg+1 frequencies because its hg status is unknown).

Two LondonMT sequences belong to hgs L1b and L3b and suggest recent gene flow from the Caribbean. L1b and L3b hgs are also commonly found in African-Americans, as well as in West Africa, which supports the known importance of Western Africa in the Atlantic slave trade (Salas *et al.* 2002); the Caribbean was en route from Africa to the Americas and analysis of hg frequencies shows that the Caribbean clusters closely with Western Africa (Salas *et al.* 2002). Therefore given that few African slaves settled in Britain (Walvin 2000) and immigration from the Caribbean has been prevalent recently (McAuley 1993) the most likely conclusion is that these lineages are recent new comers to the London gene pool. As noted above these L sequences place LondonMT as an outlier on the third principal component and draw the sample towards other populations with higher frequencies of L, such as Portugal. The one hg C sequence does not have an affect on the analyses performed here. Its low frequency in the three large

datasets of European mtDNA sequences used in comparisons in this study suggest that it is the result of sporadic gene flow from East Asia. Britain has had immigrants from India since the 18th century with a recent boom in the last 60 years (McAuley, 1993), therefore, this gene flow has more than likely occurred in the last 300 years. An obvious next step would be to compare whether the (sampled) individuals who defined themselves as having a black or Asian ethnic background were the individuals whose genetic type indicated black or Asian genetic ancestry. This would have been an interesting opportunity to assess differences between self-defined ethnic group and ethnic groups based on genetic analyses. However, due to restrictions imposed on the collection of the ethnic identity questionnaires this was not possible.

As discussed above, the standard diversity index *H* is not a reliable estimator of the diversity in the present Y-chromosome dataset because it is biased by high frequency types, leading to erroneous conclusions. However, a simple count of hgs and haplotypes shows that LondonY does indeed have high frequencies of each, indeed many of the haplotypes are singletons (see Appendix, Table A.3) suggesting that the population is not closed or isolated. If this is contrasted to the Y-chromosome hgs and haplotypes in the Basques, a population which is considered isolated on the basis of many criteria (Calafell and Bertanpetit 1994; Comas *et al.* 2000; Cavalli-Sforza *et al.* 1994; Chikhi *et al.* 2002; but also see Hurles *et al.* 1999 for an opposing view), the diversity in LondonY is apparent. The diversity indices calculated for LondonMT are easier to interpret, all three indices (θ_π and θ_k, and *H*) consistently rank LondonMT 3rd, indicating that the mtDNA lineages are divergent (i.e. not closely related), and the female N_e is large. A property of some of the non-BCD populations used here is that they may be collections from more than one small town from each country or region. Therefore the effects of drift might result in country-wide samples being relatively heterogenous, so by their very nature have more diversity than the most representative sample of male or female Londoners even taking into account that most of these collections are made from small rural populations. For example, a proportion of the "England/Wales" sample of the Helgason comparison dataset (originally derived from Richards *et al.* [1996]), has been

collected from Cornwall, Clywd, Gwynedd, Dyfed, Powys, Glamorgan and Gwent.

An important question to ask of the London data is the extent to which they can be considered a representative sample of genetic diversity in London particularly as there is little statistical support for the diversity of LondonY. As Table 4.1 shows, six different ethnic groups listed in the 2001 Census for London were not included in the present sample: "Pakistani", "Bangladeshi", "White and Black African", "other Black background", "Chinese", and "Other Ethnic Group". Indeed a Kruskal-Wallis test shows that the distribution of means between the 2001 Census records and the London sample is significantly different ($p=0.0031$). Therefore the London sample is not representative of the ethnic diversity in London. This is confirmed by the fact that the two independently collected samples of London Y-chromosomes, LondonY and London2, have significantly different hg+1 frequencies. Indeed London2 has an unusual hg+1 composition that differentiates it from much of the BCD (3 YAP+ chromosomes which are not found in the BCD (see Table 2.9) and are rare in Western Europe [Semino *et al.* 2000]). This degree of differentiation is not seen in comparisons with the RD due to the hg clustering method applied to the data. What are the factors that might have caused the two samples of London Y-chromosomes to be different? There appear to be two obvious answers. First, the actual genetic diversity might be so great in London that a much larger sample than the total of 157 Y-chromosomes studied here is needed to fully capture the diversity. Despite not being fully representative of ethnic diversity in London, the LondonY, London2 and LondonMT samples collected for this study still reflect some of the diversity predicted from the 2001 Census and historical accounts.

Secondly both London samples were collected at museums in London, whilst these were random collections in the sense that no particular type of museum visitor was targeted (apart from having to be male, over 18 years of age, and resident in London), the actual subset of the London population that visits museums and wants to participate in genetic studies may be unrepresentative of the total population of London. Furthermore, from the purely anecdotal experience of the author during sample collection at the Museum of London

there did appear to be a bias in the subset of the London population that wanted to participate: predominantly people who were interested in tracing their history to Celtic British populations or Viking invaders with a lack of ethnic minorities. A correlation between ethnic group and genotype is not necessarily implied in the use of Census records to indicate the amount of genetic diversity expected in London. The degree to which biology in general, and an individual's genotype specifically, can predict their race or ethnicity is a contentious issue (Editorial, Nature Genetics 2001). In the context of variation of drug metabolising enzymes (DMEs), it appears that X-chromosome and chromosome-1 microsatellite-defined haplotypes are a better predictor of DME variation than either ethnic affiliation or geography (Wilson *et al.* 2001b), which are the usual predictors used for the purpose of analysis (McLeod 2001). However, due to the particular properties of the Y-chromosome and mtDNA, such as the lack of recombination, and small effective population size (discussed in more detail in sections 2.1.2 and 4.1.2 above) there is a degree of geographic structure associated with Y-chromosome and mtDNA hgs. Thus, individuals who describe themselves as White British in the Census are more likely than not to have either their maternal or paternal (or both) ancestors from Britain, and Black Africans from Africa or the Caribbean, Indians from India, etc. Each of these continents or regions has a set of hgs and haplotypes more commonly associated with them than with any other continent or region, therefore the ethnic composition of London can be used to crudely predict the level of genetic diversity one expects to see.

Y-chromosomes from Liverpool (C Capelli personal communication) were also analysed to assess genetic structure and diversity between Liverpool and the BCD and RD and to place the London Y-chromosome results into context with another city. Whilst Liverpool is a relatively large city (with a population size of 439,476, Source: National Statistics website: www.statistics.gov.uk) Census records show it does not has as diverse an ethnic mix as London (Table 4.1). A Kruskal-Wallis test shows that there is a significant difference in means ($p = 0.0031$) for the ethnic groups in Liverpool and Liverpool has most non-British ethnic groups at lower frequencies than London. Therefore (representative) samples collected from these two populations are predicted to exhibit less diversity in Liverpool, which is the pattern found here both with PC plots and

exact tests of population differentiation. Liverpool also has fewer haplogroups than either LondonY or London2 (n=4, 9, and 8 respectively) as well as fewer haplotypes which indicates a less diverse history for the Liverpool male population. It is possible that a larger sample size for Liverpool would however reveal more diversity.

Analysis of Y-chromosome and mtDNA lineages in the same population allow comparisons of the relative inter- and intra-population diversity of these systems (e.g. Seielstad *et al.* 1998), which has been subject to much discussion in the literature. Seielstad *et al.* 1998 asserted that mtDNA exhibited high levels of intra-population diversity and low levels of inter-population diversity (between populations within continents and between continents) with the Y-chromosome showing the opposite pattern, later confirmed by Oota *et al.* (2001). A higher female than male migration rate, through the social phenomenon of patrilocality whereby women move away from their natal home to that of their husband's, has been proposed as the main factor behind this pattern (Seielstad *et al.* 1998). This phenomenon is somewhat counter intuitive because most historical explorers, and recent migrants (Burmeister 2000), are men, but anthropological studies suggest that patrilocality occurs in around 50% of societies (Burton *et al.* 1996).

In this study, Y-chromosome and mtDNA hg diversity was calculated as haplotype information for British comparison populations was not available. Considering the British Y-chromosome and mtDNA populations as a whole, the results indicate that the Y-chromosome has increased inter-population and reduced intra-population diversity compared to mtDNA, confirming the findings of Seielstad *et al.* (1998), Oota *et al.* (2001). and Pérez-Lezaun *et al.* (1999), amongst others. The difference in the apportionment of variation between the Y-chromosomes and mtDNA sequences is however small compared to the findings of others. For example, the percentage of variation within populations has been estimated to be as high as 81.4% for mtDNA and as low as 35.5% for the Y-chromosome (Seielstad *et al.* 1998). When the samples are partitioned into two groups neither London population has high levels of between group variation (indeed, both values are negative) suggesting that neither London Y-

chromosomes or mtDNAs exhibit high levels of structuring within Britain. These results are best explained by the higher mobility of people to and from cities (Dobson and McLaughlin 2001; Vickers 1998), where a sex bias is not expected to be as pronounced, at least in recent history.

4. 5. Conclusions

This study has for the first time explicitly investigated the genetic diversity of a metropolitan district. Whilst the results showed that both the paternal and maternal histories of Londoners were broadly comparable to the rest of Britain, reflecting their shared history, evidence for increased diversity in London was found, a pattern not observed for Liverpool. The presence of more diversity in London, compared to Britain as well as Liverpool, was anticipated from historical accounts, archaeological records and recent Census records. However it is possible that, at least for the Y-chromosome, the sample analysed here was not fully representative of the genetic diversity of London, and it appears that sampling from ethnic minorities is problematic. There are also apparent limitations with the present data that are the result of restrictions imposed on sample collection.

Chapter 5. The Maternal Origins of the Lemba and Sex-Biased Admixture

5.1. Introduction

The Lemba are a Bantu-speaking group (Johnston 2003) living predominantly in Malawi, Zimbabwe and South Africa (http://www.mindspring.com/~jaypsand/lemba.htm; 26th May 2004) whose oral history claims descent from a Jewish population who came somewhere from the north (Parfitt 1997). Due in part to this unusual and somewhat enigmatic ancestral claim the Lemba have been the subject of curiosity in the ethnographic literature for at least 100 years (surveyed by Buijs 1998), and more recently researched in the scientific literature (Hughes *et al.*1978; Spurdle and Jenkins 1996; Hammer *et al.*2000; Thomas *et al.* 2000; Wilson and Goldstein 2000). The following sections will review areas relevant to this chapter, such as historical and ethnographic accounts of the Lemba, Jewish identity, the maternal lineages of Jewish, African, and Middle Eastern populations, and previous genetic and serological studies of the Lemba.

5.1.1. The Lemba

Today there are around 50,000-70,000 Lemba individuals (Parfitt 2003). Much of their tribal lore is recounted in a song (the "Ndinda song") which states that the Lemba came from Sena where people died like flies and crossed Pusela from where tribes went to Zimbabwe. The location of Sena is not known by the Lemba, except that it is somewhere to the north; places with names similar to Sena have been found in the Yemen, Judea, Egypt and Ethiopia (Parfitt 1997). Parfitt (1997) also presents a compelling case for the Sena of the Lemba tradition being in Yemen. The small town of Sena (not to be confused with the capital of Yemen, Sanaa) lies at the eastern end of the Hadramaut valley in the Yemen. Historically Sena was much larger because a dam allowed intensive irrigation and supported a bigger population, but at an unknown point in the past the dam burst and people left. Some of the tribal names found in Sena today (e.g. ba-sadik and ba-khamis) match those of Lemba clans (Sadiki and Hamisi). A valley links Sena to the port of Sayhut where the crossing to Africa is

relatively easy with the right combination of winds and currents (Parfitt 1997), although this appears to be a typical feature of the region in general (Segal 2001) and not just Sena. The Lemba's connections with the Middle East may be specifically Jewish or more simply a reflection of centuries of contact between the Middle East and eastern and sub Saharan Africa; trade and the movement of people, both to and from the Middle East and Africa have been well documented (Segal 2001).

Some early ethnographic accounts found similarities between the Lemba and Semitic peoples (Hughes *et al.*1978; Buijs 1998). Here the term Semitic is used to describe people of Arab and Jewish descent (*Semite* "a member of any of the peoples supposed to be descended from Shem, Son of Noah, including especially the Jews, Arabs, Assyrians, Babylonians, and Phoenicians", Oxford English Dictionary 1995). For example it was claimed that the Lemba had prominent (i.e. non European noses), and fair skins (detailed in Hughes *et al.*1978), but as Hughes and colleagues (1978) and Parfitt (1997) noted, they themselves could not see any phenotypic features that distinguished the Lemba from their Bantu neighbours. The Lemba speak a Bantu language, which superficially suggests that they are indeed an indigenous African population, rather than migrants from the Middle East. However, language has been identified as one of the important indicators of the extent to which an immigrant population has assimilated into the indigenous culture (Pew Hispanic Centre Report 2004). Under certain cultural and political pressures, where assimilation into the indigenous culture is encouraged and has positive effects on the emigrant population's lifestyle, the process of linguistic change from near total use of the immigrant population's mother tongue to almost exclusive use of the indigenous tongue can happen in as few as 3 generations (Pew Hispanic Centre Report 2004). Although it is unlikely to propose that the Lemba migrated to Africa as little as 3 generations ago, because of their rather vague oral history relating to their origins, it is entirely feasible that the Bantu language replaced their mother tongue.

Several Lemba traditions have also been singled out as congruent with a Jewish origin: the avoidance of pork and meat from non-cloven hoofed ruminants (Buijs

1998; Parfitt 1997), male circumcision rites and their strong endogamy (the practice of marrying within the same social group), although non-Lemba (*senzi*) women can marry Lemba men after a long ritual process of purification (Parfitt 1997). Some Lemba believe they are related to the Bene Israel or Falashas, a Jewish population in Ethiopia whose origin myth says that they came from Sennar (Parfitt 1997), which has obvious similarities with Sena of the Lemba. Buijs (1998) has recently argued however that white colonists and missionaries imposed the apparent similarities between the Lemba and Semitic populations on the community at a time when languages and tribes were being classified, rather than the notion of Jewish ancestry being "real". Later, it was argued that the Lemba propagated these Semitic links as a way of delineating their own ethnic identity in the absence of a distinct language or traditional chiefs, primarily through the Lemba Cultural Association (LCA) (Buijs 1998). A good example is that of the Lemba's flag which depicts the Star of David and an elephant. Parfitt (1997) quoted a Lemba leader who stated that the flag was an ancient Lemba symbol. It was however designed within living memory for the LCS (Buijs 1998); indeed the Star of David only started to be used as a Jewish symbol during the Middle Ages (Parfitt 1997) so it cannot be an ancient symbol carried by the Lemba from their place of origin. As Sanders (2000) has noted the story of the Lemba's origins is not unique; many tribes believe their origins began with the exile from a distant land. Whilst these peoples are not necessarily Jewish, such stories have particular resonance since the Holocaust because of the sense that Jews have been "lost" so "finding" new Jewish tribes is a form of continuity (Zoloth 2003). Recent ethnographic accounts have thus viewed the possible Jewish origins of the Lemba with some scepticism.

5.1.2. Jewish Identity

Before proceeding to review relevant genetic studies of Jewish populations, and the Lemba it is necessary to briefly describe some basic aspects of how Jewish identity is defined. Jewish populations can be separated into several different groups on the basis of caste and ancestry (Encyclopaedia Judaica 1972).

Although Jewish identity is maternally inherited, there are three male castes (Cohen, Levi and Israelite), which are determined by patrilineal descent (Encyclopaedia Judaica 1972). The Cohanim represent the Jewish high priesthood and have specific religious rights and duties, as well as restrictions, associated with their status. Levites also have some rights and duties but the restrictions are fewer. It is possible for the male descendents of converts to Judaism to be Israelites, but not Cohamin or Levites. Cohanim and Levites are thought to each comprise around 4% of the Jewish population (Behar *et al.* 2003). The further main subdivisions are into Ashkenazi and Sephardic Jews, made on the basis of ancestry. Ashkenazi Jews are descended from Jews who lived in Germany, Poland, Austria, and Eastern Europe who spoke Yiddish (Wigoder 1974), and the term Sephardi has now come to refer to descendents of Jewish communities in Spain and Portugal as well as Jews living in North Africa and the Middle East (Shamir and Shavit 1986).

5.1.3. mtDNA and Y-Chromosome Diversity in Jewish Populations

mtDNA and Y-chromosome lineages in Jewish populations have some particular characteristics which allow investigations of the Lemba to be compared and placed into an existing framework. Despite the fact that Jewish identity is maternally inherited a relatively small number of studies have examined Jewish mtDNA lineages. The results of these studies however provide a uniform picture of diversity of Jewish mtDNA lineages. A wide range of Jewish female populations (Ashkenazi Jews, Moroccan Jews, Iraqi Jews, Iranian Jews, Georgian Jews, Bukharan Jews, Yemeni Jews, Ethiopian Jews, and Indian Jews), all show evidence for reduced diversity (hgs and haplotypes) compared to their geographic hosts (Thomas *et al.* 2002; Richards *et al.* 2003; Behar *et al.* 2004a). Most of these Jewish communities also have a founding sequence type that is a) rare in their host population and other Jewish populations, and b) at higher frequency than *any* sequence in any of the host populations (Thomas *et al* .2002). These lines of evidence suggest that each of the Jewish communities were independently founded by a small number of females (Thomas *et al.*2002;

Richards *et al.* 2003; Behar *et al.* 2004a), which agrees with the matrilineal inheritance of Jewish identity. The Ashkenazi Jewish population is large (estimated to be ~8 million immediately prior to World War Two; Behar *et al.* 2004a) and appears to be differentiated (Behar *et al.* 2004a), which may explain why the Ashkenazim in the study of Thomas *et al.* (2002) did not conform to points a) and b) above. Employing a larger sample of Ashkenazi Jews, however, Behar *et al.* (2004a) found they fit into the pattern seen in other Jewish populations. The commonest mtDNA hgs in a range of Jewish populations are typically of Eurasian origin (M B Richards, unpublished results; Behar *et al.* 2004a). For example Ashkenazi Jews have a high frequency of hg K (Behar *et al.* 2004a). Indian Jews are the exception to this rule as the modal haplotype belongs to hg M which is of Asian origin (Derenko *et al.* 2003).

Even though the work in this chapter focuses on the Lemba's maternal history, two important studies of the Lemba have analysed their paternal history (Spurdle and Jenkins 1996; Thomas *et al.* 2000), hence a brief description of the characteristics of Jewish Y-chromosomes is pertinent. Several recent publications have focussed on the paternal history of Jewish populations (Skorecki *et al.* 1997; Thomas *et al.* 1998; Thomas *et al.* 2000; Hammer *et al.* 2000; Nebel *et al.* 2000; Nebel *et al.* 2001; Thomas *et al.* 2002; Lucotte and Mercier 2003; Behar *et al.* 2003; Behar *et al.* 2004b). An important finding to come to light was the presence of a haplogroup and within this a modal microsatellite-define haplotype (Cohen Modal Haplotype, CMH) at high frequency in Ashkenazi and Sephardic Cohen Jews (Skorecki *et al.* 1997; Thomas *et al.* 1998). The CMH has since been found to belong to hg JxJ2 (Nebel *et al.* 2001; Thomas *et al.* 2002), defined by the derived state at 12f2 (a marker not tested in the 1997 or 1998 studies of Thomas and colleagues), in agreement with the high frequency of hg J in many Jewish populations (see below). The frequency of the CMH and its one step neighbours was 69.4% in Ashkenazi and 61.4% in Sephardi Cohen males, but much lower in other Jewish groups (Thomas *et al.* 1998). Several studies that have considered the Ashkenazi and Sephardi populations together and not stratified the sample according to caste have traced their combined Y-chromosomes to a Middle Eastern source

population (Hammer *et al.* 2000; Nebel *et al.* 2000; Nebel *et al.* 2001; Lucotte and Mercier 2003) based on the observed hg frequencies.

5.1.4. Genetic and Serological Investigations of the Lemba

A small number of studies have assessed the Lemba's claims of Semitic ancestry. Three independent analyses of Lemba Y-chromosomes (Spurdle and Jenkins 1996; Thomas *et al.* 2000; Hammer *et al.* 2000) have concluded that there is evidence for a Semitic component in the Lemba's male ancestry. Classical markers (Hughes *et al.* 1978) and an analysis of X-linked microsatellites (Wilson and Goldstein 2000) suggest a high proportion of African influence, confirmed by the small number of Lemba mtDNA sequences analysed by Soodyall *et al.* (1996). Focussing first on the Y-chromosome data, Spurdle and Jenkins (1996) concluded on the basis of YAP, p12f2, p49a/*Taq*I and pDP31 frequencies that around 50% of the Lemba chromosomes had a Caucasoid/non-African origin. As the Jewish and Middle Eastern populations had quite similar frequencies of the Y-chromosome markers, distinguishing between a general Semitic and specific Jewish origin was difficult. The authors also found direct evidence for African male gene flow into the Lemba gene pool in the presence of an African specific p49a/*Taq*I haplotype at high frequency in the Lemba, but absent in the non-African populations. Despite Spurdle and Jenkins' study being conducted in the early stages of human Y-chromosome population studies, when a smaller number of markers were available (Hurles and Jobling 2001) the general conclusions were later confirmed by Thomas *et al.* (2000) who used 4 binary markers (YAP, SRY_{4064}, sY81, and 92r7) and 6 microsatellites (DYS388, 393, 392, 19, 390, 391). The 4 non-African populations studied by Thomas and colleagues (Ashkenazi and Sephardic Israelites, Yemen-Hadramaut and Yemen-Sena) had very high frequencies (62%-100%) of Y-chromosomes underived at all of the markers studied (termed UEP Group 1 by Thomas *et al.* 2000), as did the Lemba (65.4%), but the frequency was much lower in the Bantus (16.9%). There was also evidence for a Bantu Y-chromosome component in the frequencies of the so called UEP Group

4 chromosomes (derived at all markers except 92r7), which was at high frequency in the Bantus (80.5%) at low frequency or absent in the non-African populations, but in 30.2% of the Lemba. Due to the similarities between the Jewish and Arab populations in the frequency of UEP Group 1, a distinct Jewish input was impossible to detect, echoing the conclusions of Spurdle and Jenkins (1996) and other studies of Jewish and Middle Eastern populations (Hammer *et al.* 2000; Nebel *et al.* 2000; Nebel *et al.* 2001; Lucotte and Mercier 2003). The presence of the CMH in the Lemba (comprising 13.5% of UEP Group 1 chromosomes, 8.8% of the total Lemba Y-chromosome gene pool) however provides stronger evidence for a distinct Jewish input, particularly as the CMH was not observed in either the Bantu or Yemen Sena populations, and in only one Yemen-Hadramaut individual (Thomas *et al.* 2000). Employing a subset of the samples analysed by Spurdle and Jenkins (1996), Hammer *et al.* (2000) confirmed both African and Semitic components to their Y-chromosome gene pool.

In contrast an analysis of ABO, MNS, Rhesus, P, Duffy and Kidd blood group frequencies (Hughes *et al.* 1978) found no significant differences between the Lemba and their neighbours, the Zezuru, a Bantu speaking population. Indeed, the frequencies of the 6 blood groups were not consistent with frequencies seen in Arab populations (Hughes *et al.* 1978). Soodyall *et al.* (1996) found that the frequency of the mtDNA intergenic COII/tRNA^{Lys} 9-bp deletion in the Lemba (26.9%) was strikingly similar to many Southern African Bantu speakers and suggested that the maternal heritage of the Lemba has a significant African component Although the COII/tRNA^{Lys} 9-bp deletion was generally considered a signature of Asian populations, it has also been observed in African populations, and control region data suggests that the deletion has arisen separately in Asia and Africa (Soodyall *et al.* 1996).

Confirmation that the Lemba are indeed an admixed population comes from an analysis of linkage disequilibrium (LD) using a panel of 66 markers on the X-chromosome (Wilson and Goldstein 2000), typing Lemba, Bantu, Ashkenazi Jewish, and Ethiopian individuals from the study of Thomas *et al.* (2000). LD describes the non-random association between alleles in a population that are

more likely to be inherited together because of limited recombination between them (Jobling *et al.* 2003). The level of LD within and between populations is affected by two factors: demography, which will affect the whole genome; and genetic factors such as mutation rates and the effects of selection (Pritchard and Przeworski 2001). Of interest here are demographic factors, specifically admixture, which is known to increase the distance over which LD extends (Pritchard and Przeworski 2001). Wilson and Goldstein (2000) found that the number of marker pairs in significant LD in the Lemba was much higher than in Ashkenazi Jews, Bantus or Ethiopians (13.8%, 7.0%, 7.7%, and 6.4% respectively), all of who were used as potential parental populations for the Lemba. Additionally, the range over which LD extended was greater in the Lemba (19-24 centimorgans, cM) than the Ashkenazim and Bantu (1-6cM) and Ethiopians (0-5cM). Both of these results suggest that the Lemba have experienced more admixture than the three comparative populations. Simulated admixed populations were created, using Bantu/Ashkenzim and Bantu/Ethiopians as parental populations, which suggested that Bantus and Ashkenzim were more likely to be the parental populations rather than Bantus and Ethiopians. As the X-chromosome spends twice as much time in females than in males, it is expected to over-represent female ancestry (Jobling *et al.* 2003), hence these estimates indicate that female Bantu input has been high.

5.1.5. mtDNA Diversity in East Africa, Bantu speakers and the Middle East

The above sections show that there seems to be a clear dichotomy between accounts of the Lemba's origins (i.e. they are either African or Middle Eastern/Jewish), based on historical and ethnographic accounts (Hughes *et al.* 1978; Parfitt 1997). Genetic studies suggest inputs from both African and Middle Eastern/Jewish populations (Hughes *et al.* 1978; Spurdle and Jenkins 1996; Hammer *et al.* 2000; Thomas *et al.* 2000; Wilson and Goldstein 2000). As the maternal lineages of the Lemba form the focus of the work in this chapter, this section will review published accounts of mtDNA diversity in (relevant)

African and Middle Eastern populations; Jewish populations have been considered above.

Sub-Saharan African populations are characterised by almost exclusive presence of the superhaplogroups L1, L2 and L3 (excluding those L3 lineages that are found in Europe) (Richards *et al.* 2003; Salas *et al.* 2002; Chen *et al.* 1995; Rando *et al.* 1998; Passarino *et al.* 1998). The term L3A can be used to refer to those lineages of L3 that are not included in M or N (Rando *et al.* 1998) (i.e. non-African L3 lineages). For brevity the term L-hgs is used here to refer to L1-L3A hgs (Salas *et al.* 2002). The Sahara forms a substantial physical barrier between northern and southern Africa, which appears to have restricted population movement between northern and southern Africa. Eurasian peoples have had a history of contact with North African populations, which is reflected in the archaeological record, linguistics, and the phenotypic similarity of Northern African populations to Eurasians (Cavalli-Sforza *et al.* 1994). This history of contact is also reflected in the mitochondrial lineages found in North African populations, where Eurasian as well as African lineages can be found (Rando *et al.* 1998; Maca-Meyer *et al.* 2001; Richards *et al.* 2003). For example U6 is a marker of North African populations, although it is now thought to have originated in the Middle East around 30kya and spread to Africa where it diversified and some lineages subsequently moved back to the Middle East (Maca-Meyer *et al.* 2003). Despite being located in sub-Saharan Africa, the Ethiopian population is characterised by low frequencies of L-hgs and the presence of Eurasian hgs (Passarino *et al.* 1998; Richards *et al.* 2003), in keeping with the contention that the Ethiopian population has been greatly influenced by Eurasian populations since the Neolithic (Cavalli-Sforza *et al.* 1994). The Eurasian component of the Ethiopian mtDNA gene pool appears to consist of several Eurasian lineages, such as pre-HV1, T, J, U, and HV, none of which are at particularly high frequencies (M B Richards, unpublished results). Even so they still account for ~30% of the lineages. The hg M1, thought to be of East African origin (Quintana-Murci *et al.* 1999) comprised a further 10% of the lineages, U6 a further 3% and the remaining 55% of lineages belonged to L-hgs (Richards *et al.* 2003). Passarino *et al.* (1998) found broadly similar results,

although exact frequencies differ somewhat due to methodological differences in hg assignment.

Bantu speaking peoples comprise the single largest linguistic group in sub-Saharan Africa (1/4 of all Africans, Cavalli-Sforza *et al.* 1994). The term "Bantu" was originally used as a linguistic classification, but is now also used to define populations on the assumption that the spread of Bantu languages across sub-Saharan Africa was accompanied by the spread of peoples (Cavalli-Sforza *et al.* 1994), and it is in this latter sense that the term Bantu is used here. Evidence for an expansion of Bantu speaking peoples from homelands in the southeast of Nigeria and/or northwestern Cameroon (Van der Veen and Hombert 2001) to much of the rest of sub-Saharan Africa around 3kya comes from archaeology, linguistics (Cavalli-Sforza *et al.* 1994), and more recently Y-chromosome (Thomas *et al.* 2000; Underhill *et al.* 2001; Peirera *et al.* 2002) and mtDNA diversity (Salas *et al.* 2002), the latter of which will now be considered.

The Bantu expansions have been responsible for the spread of several lineages across much of sub-Saharan Africa, resulting in less geographic structure for some of these hgs, although older L-hgs such as L1d and L1k (Forster 2004), not explicitly associated with Bantus have retained more geographical structure. L1a and L2a are both common hgs in sub-Saharan African populations, a distribution associated with Bantu expansions (Salas *et al.* 2002). Indeed L2a now comprises the single commonest hg cluster in Africa (Torroni *et al.* 2001a; Salas *et al.* 2002); around ¼ of all African lineages belong to this group (Salas *et al.* 2002). L1a and L2a are both particularly common in South Eastern Bantu speakers, comprising 0.28 and 0.29 of the mtDNA lineages observed (Salas *et al.* 2002). The hg L3e is also common in South Eastern Bantus, where Salas *et al.* (2002) found the frequency to be 0.14, again this hg is found throughout Africa through the Bantu expansions, although it has been hypothesised to have an origin in central Africa/southern Sudan (Bandelt *et al.* 2001). Soodyall *et al.* (1996) detected an intergenic COII/$tRNA^{Lys}$ mtDNA deletion in sub-Saharan African populations, the distribution of which is also associated with Bantu expansion; note that this deletion was once thought to be Asian specific, but Soodyall *et al.* (1996) found that it also arose independently in the Africans.

Regional variation is present in the Bantu speakers of Africa, which is not surprising given their large geographical distribution. For example, L1a was not observed in two Senegalese Bantu speaking populations (Mandenkalu and Wolof) studied by Chen *et al.* (2000), whilst L2c is the single most common in the Mandenkalu (0.28) but found at very low frequency in the south eastern Bantus of Salas *et al.* (2002). Due to gene flow between Bantus and Khoisan populations, particularly in southern Africa, there are some similarities between these populations in mtDNA hg frequencies, such as the high frequency of L1a in the !Kung and Khwe (Chen *et al.* 2000) and the south eastern Bantus described by Salas *et al.* (2002). The presence, albeit at low frequency, of the Khoisan specific hg L1d in south eastern Bantus is also testament to gene flow between these populations (Salas *et al.* 2002). Many other hgs have been observed in Bantus but at much lower frequencies and as such are not considered here for brevity and the reader is directed to Salas *et al.* (2002).

A range of Eurasian mtDNA hgs are typically observed in Middle Eastern populations. Arab populations within the Middle East also show distinct evidence for gene flow from Africa in the presence of L-hgs and U6, which are typical of sub Saharan and North African populations respectively. In particular, the Hadramaut region of the Yemen has extremely high frequencies of L-hgs (Richards *et al.* 2003). This gene flow from Africa of primarily female lineages is not observed for the Y-chromosome and has been interpreted as the result of the Arab slave trade between A.D. 650 and 1900 (Richards *et al.* 2003) which saw the movement of around 2/3 more women than men (Segal 2001). Such high levels of gene flow from Africa do not typify non-Arab Middle Eastern populations however (Richards *et al.* 2003). Indeed Eurasian populations that lie west of the Indus valley are predominantly characterised by western Eurasian mtDNA hgs (Quintana-Murci *et al.* 2004). For example the highest frequency of African hgs in non-Arabian Middle Easterners was 0.04 in Kurds (Richards *et al.* 2004). Even other European populations that are geographically close to Africa, such as Spanish and Portuguese, have much lower frequencies of African hgs (<0.01 [González *et al.* 2003]). Potential traces of the Arab slave trade have been detected as far east as the Makrani population in south Pakistan (Quitanan-Murci *et al.* 2004). This movement of people to the Middle East is perhaps less

well known and studied than either the movement out of Africa of modern humans or the the Neolithic Expansion. The Arab slave trade may also account for some of the Eurasian mtDNA hgs seen in eastern African populations.

5.1.6. Aims of the Chapter

The present study aims to characterise mtDNA HVSI diversity in Lemba, Bantu and Yemen-Sena individuals to ascertain the maternal ancestry of the Lemba in the same (Lemba and Bantu) individuals typed for Y-chromosome and X-chromosome markers. In addition previously determined mtDNA HVSI sequence information for Ethiopians, Ethiopian Jews, Yemen-Hadramaut, Yemen-Jews, and Ashkenazi Jews was included in the analysis as these populations are potential contributors of maternal lineages to the Lemba.

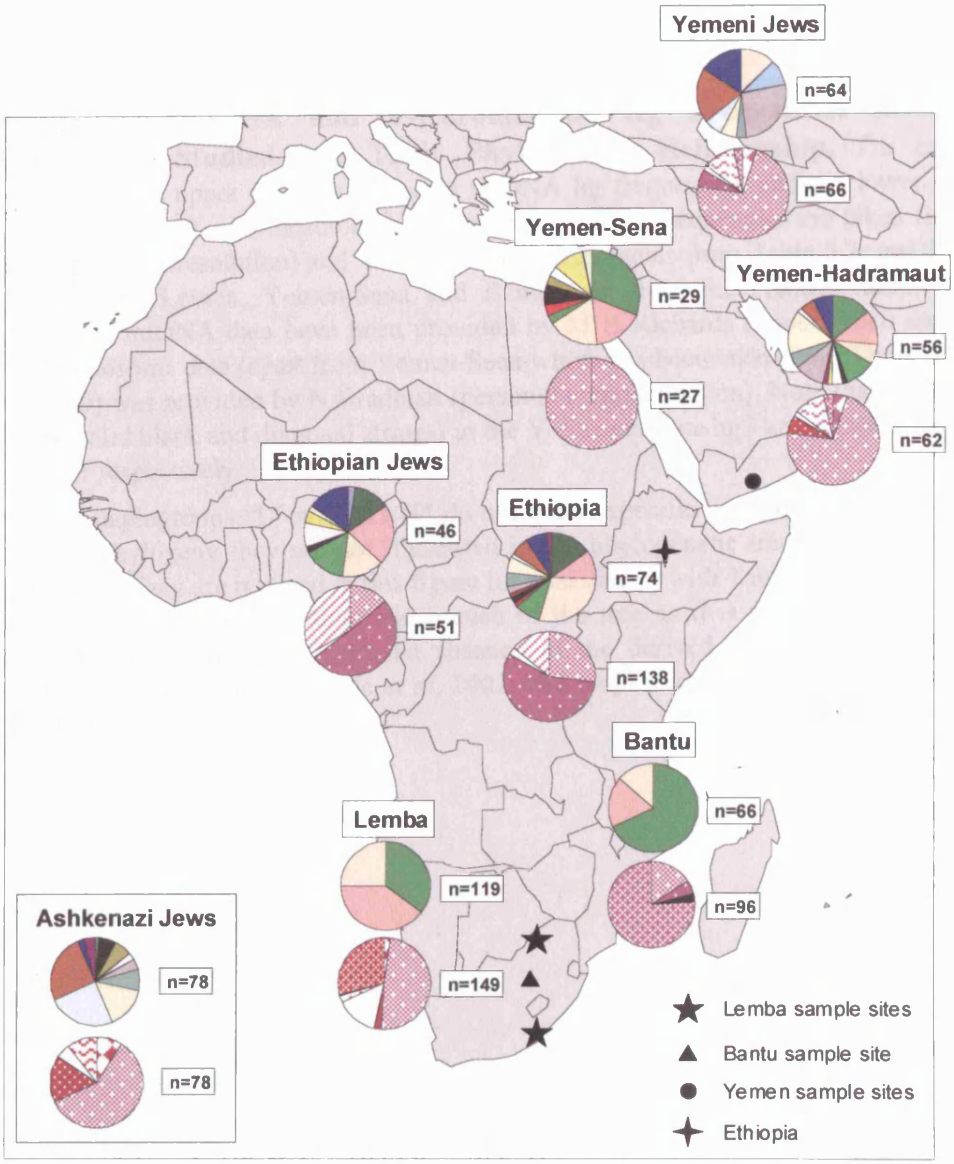
5.2. Materials and Methods

5.2.1. Study Populations

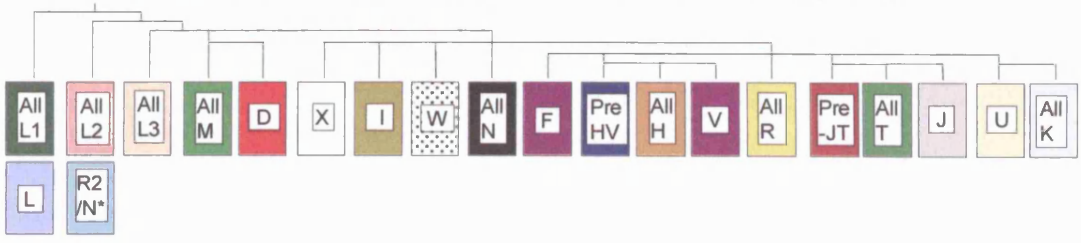
DNA samples from the Lemba, Bantu, and Yemen-Sena populations were studied for mtDNA variation. All individuals were maternally and paternally unrelated. Full details of sample collection can be found in Thomas *et al.* (2000). In brief, the Lemba samples were taken from self-designated members of the tribe from the Northern Province and Mpumalanga in South Africa, Bantus from various Bantu-speaking chieftainships in South Africa, and Yemen-Sena from the small isolated town of Sena in the Yemen (Figure 5.1).

5.2.2. mtDNA HVSI PCR Procedures

The mtDNA HVSI region was amplified using the primers conH1 and conL2 (see Appendix, Table A.2 for sequences) and a standard PCR protocol as



mtDNA Pie Chart Legend and Phylogenetic Relationship of the Observed Haplogroups



Y-Chromosome Pie Chart Legend and Phylogenetic Relationship of the Observed Haplogroups**

Figure 5.1. mtDNA and Y-Chromosome Hg Frequencies in the Populations Studied and Their Phylogenetic Relationships. Legend on following page

Figure 5.1. mtDNA and Y-Chromosome Hg Frequencies in the Populations Studied and Their Phylogenetic Relationship. For each population the upper pie chart presents mtDNA hg frequencies and the lower pie chart represents Y-chromosome hg frequencies. mtDNA frequencies are taken from Table 5.2 (low resolution) and Y chromosome frequencies from Table 5.3. mtDNA data for the Lemba, Yemen-Sena and Bantus are from the present study, the remaining mtDNA data have been provided by M B Richards (unpublished data). Y-chromosome data (apart from Yemen-Sena which has been taken from Thomas *et al.* 2000) was provided by N Bradman (personal communication). Note that the two tones (solid black and diagonal stripes) in the Y-chromosome hg7 key relate to hg7b and hg7 respectively

As the haplogroups "L" and "R2/N do not define specific lineages in the current mtDNA phylogeny they are not illustrated in the phylogenetic tree, but are listed separately. They are retained in this figure for consistency with Table 5.2.

** Hg 2 appears in more than one branch of the tree as it is determined by the derived state at SRY_{10831a} and the absence of the derived state at other typed markers (see for example Weale *et al.* 2002 who employed the same strategy as N Bradman).

described in Table 5.1. 4μl of PCR product was electrophoresed on a 1.5% (w/v TBE) agarose gel to visualise PCR products and ascertain which samples could be subsequently sequenced. All PCR products were purified using a SAP/Exonuclease I procedure and the forward strand sequenced using the primer conL2 (Table 5.1, and see Appendix, Table A.2 for sequence). A proportion of mtDNA HVSI sequences in most populations contain a T-C transversion in the poly-C stretch between np 16,184-16,193 causing the forward sequencing reaction to fail. When this occurred in the present study the reverse stand was sequenced using the primer conH3. Sequence reactions were cleaned to eliminate unincorporated nucleotides using a Sephadex™ method developed by M B Richards and electrophoresed as described (Table 5.1). The resultant sequences were aligned in SeqEd™ v1.0.3 and Sequencher™ and polymorphisms called with respect to the CRS (Anderson *et al.* 1981).

5.2.3. Assignment to lineages

Sequences were assigned to lineages using the HVSI sequence motifs described by Richards *et al.* (2000, supplementary data) and Salas *et al.* (2002). Note that the terminology of Richards *et al.* (2000) uses the hg L3a for lineages found in Africa, whilst L3a is not employed by Salas *et al.* (2002), who use L3A to refer to all non-African L3 lineages. For consistency with the comparative data of MB Richards, L3a is used for African L3 lineages. Where it was not possible to unambiguously assign a hg in this manner the samples were subjected to RFLP analysis (kindly performed by A Torroni, Università di Pavia/Università "La Sapienza", see Appendix, Table A.6). One Bantu individual that was later added to the dataset after the RFLP typing had been performed could not be assigned to a hg based on sequence motif. Hence, this sequence is excluded from analyses based on hg frequencies, but included in analyses where sequence motif alone is considered. Where relevant, polymorphisms are given as a string of nps where the mutation occurs, listing the base change for transversions (G/A-C/T) but only the np for transitions (G-A and C-T) (see Appendix, Table A.7).

Table 5.1. mtDNA HVSI PCR and Sequencing Protocol

(a) PCR Protocol

Primer Mix

Component	*Final conc. in PCR (μM)*	*Volume per reaction (μl)*
U' Primer (100μM)	0.3	0.075
R' Primer (100μM)	0.3	0.075
ddH_2O	-	1.85
Total		2.000

PCR Mix

1. Add 2μl Primer Mix to each well of an Abgene ® 1.5mM MgCI PCR optimisation pre-aliquoted plate (without added loading dye).

2. Add 1μl DNA (~5ng/μl) to each well.

Cycling Conditions

Temperature (degrees C)	*Duration*[a]	*Cycles*
94	5'	
94	1'	
55	1'	35
72	1'	
72	7'	
4	∞	

[a] Minutes ', seconds "

(b) PCR Clean Up Using SAP/Exo I

Component	*Volume per reaction (μl)*
Shrimp Alkaline Phosphatase (SAP) (1U/μl)	2
Exonuclease I (10U/μl)	1
Total	3.000

1. To degrade unincorporated primers prior to sequencing 3μl of the SAP/ExoI mix must be added to each PCR product
2. Incubate at 37°C for 1 hour and 80°C for 20 minutes

(c) Sequencing Protocol

Component	*Volume per reaction (μl)*
ddH_2O	4.49
ABI Ready Reaction Mix	2
ABI 5X Buffer	1
Sequencing Primer[b]	0.016
PCR Product	2.5
Total	10

[b] To a final concentration of 0.08μM. Forward sequencing is performed using conL2 and reverse sequencing with conH3.

continued

Table 5.1 continued

Cycling Conditions

Temperature (degrees C)	*Duration* [c]	*Cycles*
96	10"	
50	5"	25
60	4'	
4	∞	

[c] Minutes ', seconds "

Sequencing Clean Up Using Sephadex

1. Prepare a 96-well MultiScreen ® HV plate with Sephadex ™ G50 Superfine, as per manufacturers instructions. Allow to sit for 3 hours at room temperature.
2. Attach a Thermo-Fast ® Detection 96-well plate to the bottom of the MultiScreen plate as per manufacturers instructions.
3. Spin at 910g for 5 minutes at room temperature to remove water from the sephadex.
4. Remove the Thermo-Fast ® Detection 96-well plate and discard it and the contents.
5. Attach a Thermo-Fast ® Low Profile 96-well plate to the MultiScreen plate as above.
6. Pipette all of the sequencing reaction into the wells of the MultiScreen plate containing the Sephadex ™, ensuring that the Sephadex is not disturbed.
7. Spin at 910g for 5 minutes at room temperature to collect the sequences into Low Profile 96-well plate
8. Remove the Thermo-Fast ® Low Profile 96-well plate containing the sequences, incubate at 80° (for 40 mins and store at -20°C.

(d) Electrophoresis Conditions - ABI PRISM ® 377 DNA Sequencer

Time	*Filter*	*Acrylamide*	*Loading Buffer* [d]
3.5 hours	Sequencing	4.25%	3µl

[d] Consisting of Dextran blue and de-ionised formamide in the ratio 1:5

Numbering is as Anderson *et al.* (1981), and bases are given less 16,000, such that 16,223 is called 223.

5.2.4. mtDNA Comparison Populations

HVSI sequence data and hg assignment for 5 populations (Ashkenazi Jews, Ethiopians, Ethiopian Jews, Yemen-Hadramaut, Yemeni Jews) that might provide insights into the history of the Lemba was added to the dataset, kindly provided prior to publication by M Richards (personal communication). The same individuals have also been included in the mtDNA studies of Thomas *et al.* (2002) and Richards *et al.* (2003). The term Yemen-Hadramaut is used to refer to a collection of DNAs made at Seiyun, to differentiate those made from the town of Sena (i.e. Yemen-Sena), both Seiyun and Sena are in the Hadramaut region of Yemen (See Figure 5.1).

5.2.5. Y-Chromosome Comparison Populations

Y-chromosome hg and microsatellite data for several populations (Ashkenazi Cohen, Ashkenazi Levites, Ashkenazi Israelites, Ashkenazi Jews, Ethiopian Jews, Ethiopians, Yemeni Jews, Yemen-Hadramaut, Lemba and Bantu) were kindly provided by N Bradman (personal communication) to allow comparison between mtDNA and Y-chromosome data. Data for Yemen-Sena was taken directly from Thomas *et al.* (2000). Ashkenazi Israelites, Yemen-Hadramaut, Lemba, and Bantu data from N Bradman overlaps with that of Thomas *et al.* (2000); the data from the former dataset employed a larger number of UEP markers than the latter, hence the former dataset was used here, except for Yemen-Sena.

5.2.6. X-Chromosome Comparative Data

Haplotype data for 66 X-linked microsatellite loci for Lemba, Bantu, and Ashkenazi Jewish individuals were provided by J Wilson. Unfortunately X-chromosome data were not available for any Yemeni populations as the Ashkenazim were selected by Wilson and Goldstein to represent a Semitic population; direct comparisons between mtDNA, Y-chromosome and X-chromosome Bantu inputs are thus more analogous than those for the alternative parental population. Due to the high levels of LD in the Lemba, found to stretch out to around 20cM (Wilson and Goldstein 2000), loci that were separated by at least 30cM were chosen for the current analyses to provide several independent observations of the data. This resulted in data from 8 X-linked loci being used (DXS1060, 8027, 8012, 8082, 1220, 1192, 8091, 8087). Ethiopians were not chosen to be included in the present analyses as the analysis of Wilson and Goldstein (2000) indicated that Ethiopians were unlikely to be one of the parental populations of the Lemba.

5.2.7. Data Analysis

The mtDNA and Y-chromosome hg counts in Tables 5.2 and 5.3 respectively, and X-linked microsatellite markers in Table 5.4 were used in the analyses described below (note that the high frequency of the Y-chromosome hg BR*(xDE,JR) in all populations is partly a function of the low resolution of markers used to define this hg). Additionally, a lower resolution (hereafter termed low-res, and the original data as high-res) clustering of the mtDNA data was employed for the exact test of population differentiation as the high number of haplogroups observed in the dataset may make the power of discrimination too high between populations. The hgs used for the low-res mtDNA analyses are indicated in Table 5.2 by bold text.

PC plots were drawn using POPSTR (H Harpending, personal communication). Due to limitations with the POPSTR programme only 6 of the X-chromosome loci could be used (DXS1060, 8027, 8012, 8082, 1220, 1192). Exact tests of population differentiation (Arlequin 2.000, Schneider *et al.* 2000) were

Table 5.2. mtDNA Haplogroup Frequency Data For the Populations Studied

mtDNA Hg	*Lem*[a]	*Ban*[a]	*YemS*[a]	*YemH*[b]	*YemJ*[b]	*Eth*[b]	*EthJ*[b]	*AshJ*[b]
D	-	-	1 (0.034)	-	-	1 (0.014)	-	-
F	-	-	-	1 (0.018)	-	-	-	-
H	-	-	-	1 (0.018)	-	-	-	15 (0.192)
H01	-	-	-	-	-	-	-	1 (0.013)
HV*	-	-	-	-	-	1 (0.014)	-	-
HV1	-	-	-	2 (0.036)	13 (0.203)	4 (0.054)	-	4 (0.051)
All H	-	-	-	*3 (0.054)*	*13 (0.203)*	*5 (0.068)*	-	*20 (0.256)*
I	-	-	-	-	-	-	-	5 (0.064)
J*	-	-	-	2 (0.036)	6 (0.094)	-	-	3 (0.038)
J1	-	-	-	-	-	1 (0.014)	-	-
J1b	-	-	-	-	11 (0.172)	-	-	
J1b1	-	-	-	1 (0.018)	-	-	-	-
All J1	-	-	-	*1 (0.018)*	*11 (0.172)*	*1 (0.014)*	-	-
J2	-	-	-	3 (0.054)	-	1 (0.014)	-	-
K	-	-	-	1 (0.018)	4 (0.064)	1 (0.014)	-	19 (0.244)
L	-	-	-	-	-	-	1 (0.022)	-
L1	-	-	-	-	-	2 (0.027)	-	-
L1*	-	2 (0.030)	-	-	-	-	-	-
L1a	15 (0.126)	18 (0.273)	9 (0.310)	6 (0.107)	-	4 (0.054)	2 (0.043)	-
L1a1a	1 (0.008)	-	-	-	-	-	-	-
L1a2	-	1 (0.015)	-	-	-	-	-	-
L1b	1 (0.008)	-	-	-	-	5 (0.068)	-	-
L1c	3 (0.025)	5 (0.075)	-	2 (0.036)	-	-	-	-
L1c1	2 (0.017)	2 (0.030)	-	-	-	-	-	-
L1c2	4 (0.034)	-	-	-	-	-	-	-
L1c3	1 (0.008)	-	-	-	-	-	-	-
L1d	12 (0.101)	13 (0.197)	-	-	-	-	-	-
L1d1	3 (0.025)	3 (0.045)	-	-	-	-	-	-
L1d2	-	1 (0.015)	-	-	-	-	-	-
L1e	-	-	-	-	-	-	4 (0.087)	-
All L1	*42 (0.353)*	*45 (0.682)*	*9 (0.310)*	*8 (0.143)*	-	*11 (0.149)*	*6 (0.130)*	-
L2	-	-	-	6 (0.107)	-	2 (0.027)	8 (0.174)	-
L2*	4 (0.034)	2 (0.030)	-	-	-	-	-	-
L2/L3*	1 (0.008)	-	-	-	-	1 (0.014)	-	-
L2a	1 (0.008)	1 (0.015)	-	-	-	3 (0.041)	-	-
L2a1	9 (0.076)	3 (0.045)	5 (0.172)	-	-	-	-	-
L2a1b	26 (0.218)	5 (0.076)	-	-	-	-	-	-
L2b	-	-	-	-	-	2 (0.027)	2 (0.043)	-
L2c2	5 (0.0420)	1 (0.015)	-	-	-	-	-	-

contiunued

Table 5.2 continued

mtDNA Hg	*Lem*[a]	*Ban*[a]	*YemS*[a]	*YemH*[b]	*YemJ*[b]	*Eth*[b]	*EthJ*[b]	*AshJ*[b]
L2d	-	-	-	1 (0.018)	-	-	-	-
L2d1	1 (0.008)	-	-	-	-	-	-	-
All L2	*47 (0.395)*	*12 (0.182)*	*5 (0.172)*	*7 (0.125)*	-	*8 (0.108)*	*10 (0.217)*	-
L3*	2 (0.017)	-	2 (0.069)	-	5 (0.078)	12 (0.162)	5 (0.109)	-
L3a1	1 (0.008)	-	-	-	3 (0.047)	1 (0.014)	-	-
L3a1a	-	-	-	1 (0.018)	-	3 (0.041)	-	-
L3a2	-	-	-	-	-	4 (0.054)	-	-
L3b	2 (0.017)	-	-	1 (0.018)	-	-	2 (0.043)	-
L3b1	1 (0.008)	2 (0.030)	-	-	-	-	-	-
L3b2	2 (0.017)	1 (0.015)	-	-	-	-	-	-
L3d	1(0.008)	-	-	1 (0.018)	-	1 (0.014)	-	-
L3d1	8 (0.067)	-	2 (0.069)	-	-	-	-	-
L3e	-	-	-	-	-	-	-	-
L3e*	-	-	-	-	-	-	-	-
L3e1	6 (0.050)	1 (0.015)	-	-	-	-	-	-
L3e1a	3 (0.025)	2 (0.030)	-	-	-	-	-	-
L3e2b	3 (0.025)	-	-	-	-	-	-	-
L3e3	-	1 (0.015)	1 (0.034)	1 (0.018)	-	-	-	-
L3f	1 (0.008)	2 (0.030)	-	-	-	-	-	-
All L3	*30 (0.252)*	*9 (0.136)*	*5 (0.172)*	*4 (0.071)*	*8 (0.125)*	*21 (0.284)*	*7 (0.152)*	-
M*	-	-	-	4 (0.071)	-	-	-	1 (0.013)
M1	-	-	1 (0.034)	-	-	4 (0.054)	-	-
M1*	-	-	-	-	-	-	1 (0.022)	-
M1a	-	-	-	2 (0.036)	-	3 (0.041)	6 (0.130)	-
All M1	-	-	*1 (0.034)*	*2 (0.036)*	-	*7 (0.095)*	*7 (0.152)*	-
N*	-	-	2 (0.069)	-	-	-	1 (0.022)	-
N1a	-	-	-	-	-	2 (0.027)	-	-
N1b	-	-	-	1 (0.018)	-	-	-	5 (0.064)
All N1	-	-	-	*1 (0.018)*	-	*2 (0.027)*	-	*5 (0.064)*
pre-HV	-	-	-	4 (0.071)	10 (0.156)	6 (0.081)	7 (0.152)	2 (0.026)
pre-JT	-	-	-	-	-	1 (0.014)	-	-
R*	-	-	-	-	-	-	2 (0.043)	-
R2	-	-	3 (0.103)	1 (0.018)	-	-	-	-
R2/N*	-	-	-	-	6 (0.094)	-	-	-
T*	-	-	-	1 (0.018)	2 (0.031)	1 (0.014)	-	1 (0.013)
T1	-	-	-	2 (0.036)	-	3 (0.041)	-	2 (0.026)
T2	-	-	-	-	-	-	-	3 (0.038)

continued

Table 5.2. continued

mtDNA Hg	*Lem*[a]	*Ban*[a]	*YemS*[a]	*YemH*[b]	*YemJ*[b]	*Eth*[b]	*EthJ*[b]	*AshJ*[b]
U*	-	-	-	6 (0.107)	3 (0.047)	-	-	-
U1	-	-	-	-	1 (0.016)	-	-	-
U1a	-	-	-	-	-	-	-	1 (0.013)
All U1	-	-	-	-	*1 (0.016)*	-	-	*1 (0.013)*
U2	-	-	-	1 (0.018)	-	-	-	3 (0.038)
U2i	-	-	-	-	-	1 (0.014)	-	-
All U2	-	-	-	*1 (0.018)*	-	*1 (0.014)*	-	*3 (0.038)*
U3	-	-	-	-	-	-	1 (0.022)	2 (0.026)
U5a1*	-	-	-	-	-	-	-	2 (0.026)
U5a1a	-	-	1 (0.034)	2 (0.036)	-	-	-	-
All U5	-	-	*1 (0.034)*	*2 (0.036)*	-	-	-	*2 (0.026)*
U6	-	-	-	-	-	-	-	1 (0.013)
U6a*	-	-	-	-	-	-	-	2 (0.026)
U6a1	-	-	-	-	-	3 (0.041)	-	-
All U6	-	-	-	-	-	*3 (0.041)*	-	*3 (0.038)*
U7	-	-	-	-	-	-	-	1 (0.013)
V	-	-	-	-	-	1 (0.014)	-	3 (0.038)
W	-	-	-	-	-	-	4 (0.087)	-
X	-	-	1 (0.034)	2 (0.036)	-	-	-	2 (0.026)
n	*119*	*66*	*29*	*56*	*64*	*74*	*46*	*78*

Notes. Hg data is given as a count and frequency in parentheses. HVS1 sequences can be found in Appendix, Table A.7. Abbreviations as follows: Lem=Lemba, Ban=Bantu, YemS=Yemen-Sena, YemH=Yemen-Hadramaut, YemJ=Yemeni Jews, Eth=Ethiopians, EthJ=Ethiopian Jews, AshJ=Ashkenazi Jews. These abbreviations are used to refer to the same population regardless of the locus being studied, hence the abbreviation used for Lemba mtDNA and Y chromosome data will still be "Lem".

[a] Present Study

[b] From M B Richards, personal communication

calculated as in section 2.2.4, and haplotype diversity (h) was calculated for the mtDNA data as section 4.2.9. Exact tests of population differentiation for the X-linked markers were calculated separately for each locus based on the frequency of allele sizes in each population.

Admixture proportions of the relative inputs of Bantu and Yemeni mtDNA and Y-chromosome lineages, and Bantu and Ashkenazi X-chromosomes on the Lemba, were inferred using a likelihood based approach, LEA (Chikhi *et al.* 2001). Details of the LEA method have been previously described in more detail (Section 2.2.4). Briefly the admixture model assumes that two parental populations P_1 and P_2 have contributed proportion p_1 and p_2 ($p_2=p_1-p_2$) to a third hybrid population, P_h. From the moment of admixture the three populations evolve independently for T generations by drift (Figure 2.7). The mitochondrial genome and the Y-chromosome are single loci, hence the estimated admixture proportions are expected to have large associated credible intervals, leading to inaccurate point estimates (Chikhi *et al.* 2001; Chikhi *et al.* 2002), hence 95% credible intervals must also be considered. In contrast the 8 X-linked microsatellite markers, selected because they are ~30cM apart and should be unlinked in all populations, can be treated as 8 independent loci, which should greatly increase the power of the admixture estimates (Chikhi *et al.* 2001; Chikhi *et al.* 2002). Simulations were run for 100,000 iterations and the posterior pdf's for p_1 and t_1, t_2, and t_h were obtained and plotted using the locfit package for R, having removed the first 10% of the runs, the so called "burn in". Bantus were used as P_1 in all calculations, Yemen-Sena and Yemen-Hadramaut were alternately used as P_2 for mtDNA and Y-chromosome calculations, whilst for the X-chromosome Ashkenazis were employed as P_2. The hg (and X-chromosome haplotype) frequencies used in LEA calculations can be found in Tables 5.2-5.4.

5.3. Results

5.3.1. mtDNA Diversity Scores

Table 5.3. Y-Chromosome Hg Frequency Data for the Populations Studied

YCC Hg[a]	P*(xR1a)	BR*(xDE,JR)	R1a1	A3b2		E3a	E*(xE3a)	K*(xL,N3,O2b,P)	L	
Hg[b]	hg1	hg2	hg3	hg7	hg7b	hg8	hg21	hg26	hg28	
Population										n
Lem	3 (0.020)	73 (0.490)	-	-	1 (0.007)	43 (0.289)	4 (0.027)	22 (0.148)	3 (0.020)	149
Ban	-	15 (0.156)	-	-	3 (0.031)	73 (0.760)	5 (0.052)	-	-	96
YemS	-	27 (1.000)	-	-	-	-	-	-	-	27
YemH	3 (0.048)	45 (0.726)	7 (0.113)	-	-	2 (0.032)	4 (0.065)	-	1 (0.016)	62
YemJ	13 (0.197)	39 (0.591)	2 (0.030)	-	-	-	9 (0.136)	3 (0.045)	-	66
Eth	-	37 (0.268)	-	19 (0.138)	-	-	78 (0.565)	4 (0.029)	-	138
EthJ	-	7 (0.137)	-	17 (0.333)	-	-	26 (0.510)	1 (0.020)	-	51
AshC	1 (0.013)	70 (0.921)	3 (0.039)	1 (0.013)	-	1 (0.013)	-	-	-	76
AshL	7 (0.103)	14 (0.206)	8 (0.118)	1 (0.015)	-	38 (0.559)	-	-	-	68
AshI	15 (0.155)	48 (0.495)	4 (0.041)	-	-	-	22 (0.227)	8 (0.082)	-	97
AshJ	7 (0.090)	46 (0.590)	13 (0.167)	4 (0.051)	-	8 (0.103)	-	-	-	78

Notes. Shown are the counts of each hg and the frequency in parentheses. Data provided by N Bradman (personal communication).
Abbreviations as Table 5.2

[a] Hg nomenclature as per YCC (2002)

[b] Hg nomenclature used by N Bradman

Table 5.4. Counts of 8 X-Linked Microsatellite Markers in the Lemba and Two Hypothesised Parental Populations

	Allele Size	*Population*		
Locus Name	*(bp)*	*Lem*	*Ban*	*AshJ*
DXS1060	238	1 (0.011)	1 (0.014)	-
	244	1 (0.011)	5 (0.068)	1 (0.014)
	246	7 (0.080)	5 (0.068)	5 (0.070)
	248	6 (0.068)	7 (0.095)	-
	250	7 (0.080)	6 (0.081)	12 (0.169)
	252	21 (0.239)	25 (0.338)	26 (0.366)
	254	11 (0.125)	2 (0.027)	6 (0.085)
	256	26 (0.295)	16 (0.216)	13 (0.183)
	258	5 (0.057)	6 (0.081)	5 (0.070)
	260	3 (0.034)	1 (0.014)	3 (0.042)
	n	*88*	*74*	71
DXS8027	220	-	1 (0.012)	0
	226	1 (0.011)	3 (0.037)	0
	228	3 (0.033)	1 (0.012)	0
	230	2 (0.022)	3 (0.037)	1 (0.013)
	232	44 (0.478)	30 (0.366)	4 (0.052)
	234	14 (0.152)	21 (0.256)	20 (0.260)
	236	4 (0.043)	4 (0.049)	2 (0.026)
	238	8 (0.087)	5 (0.061)	15 (0.195)
	240	15 (0.163)	13 (0.159)	33 (0.429)
	242	-	1 (0.012)	1 (0.013)
	244	1 (0.011)	-	1 (0.013)
	n	*92*	*82*	*77*
DXS8012	171	37 (0.389)	30 (0.366)	36 (0.468)
	173	4 (0.042)	1 (0.012)	-
	177	22 (0.232)	17 (0.207)	19 (0.247)
	179	1 (0.011)	-	2 (0.026)
	181	1 (0.011)	3 (0.037)	-
	183	13 (0.137)	12 (0.146)	16 (0.208)
	185	11 (0.116)	10 (0.122)	1 (0.013)
	187	2 (0.021)	7 (0.085)	2 (0.026)
	189	2 (0.021)	2 (0.024)	1 (0.013)
	195	2 (0.021)	-	-
	n	*95*	*82*	*77*
DXS8082	212	2 (0.021)	4 (0.049)	2 (0.025)
	214	-	1 (0.012)	-
	216	4 (0.042)	4 (0.049)	-
	218	12 (0.125)	8 (0.096)	5 (0.063)
	220	9 (0.094)	10 (0.122)	29 (0.367)
	222	36 (0.375)	22 (0.268)	4 (0.051)
	224	9 (0.094)	5 (0.061)	2 (0.025)
	226	8 (0.083)	5 (0.061)	8 (0.101)
	228	11 (0.115)	13 (0.159)	10 (0.127)
	230	5 (0.052)	9 (0.110)	14 (0.177)
	232	-	1 (0.012)	5 (0.063)
	n	*96*	*82*	79

continued

Table 5.4. continued

Locus Name	Allele Size (bp)	Population		
		Lemba	*Bantu*	*Ashkenazi*
DXS1220	193	1 (0.011)	-	9 (0.118)
	195	-	-	5 (0.066)
	197	-	-	1 (0.013)
	207	3 (0.033)	1 (0.012)	1 (0.013)
	209	13 (0.141)	6 (0.074)	8 (0.105)
	211	10 (0.109)	15 (0.185)	3 (0.039)
	213	13 (0.141)	11 (0.136)	3 (0.039)
	215	18 (0.196)	15 (0.185)	38 (0.5)
	217	15 (0.163)	18 (0.222)	6 (0.079)
	219	9 (0.098)	10 (0.123)	2 (0.026)
	221	9 (0.098)	4 (0.049)	-
	223	1 (0.011)	1 (0.012)	-
	n	*92*	*81*	76
DXS1192	114	-	1 (0.012)	-
	116	1 (0.011)	1 (0.012)	-
	120	-	1 (0.012)	14 (0.179)
	122	23 (0.253)	25 (0.309)	12 (0.154)
	124	6 (0.066)	6 (0.074)	7 (0.090)
	126	10 (0.110)	5 (0.062)	-
	128	24 (0.264)	23 (0.284)	16 (0.205)
	130	20 (0.220)	10 (0.123)	15 (0.192)
	132	5 (0.055)	7 (0.086)	9 (0.115)
	134	2 (0.022)	1 (0.012)	4 (0.051)
	136	-	1 (0.012)	1 (0.013)
	n	*91*	*81*	*78*
DXS8091	70	-	-	2 (0.025)
	72	-	-	-
	74	27 (0.314)	21 (0.259)	13 (0.163)
	76	1 (0.012)	-	-
	78	-	6 (0.074)	-
	80	6 (0.070)	3 (0.037)	-
	82	-	-	-
	84	3 (0.035)	1 (0.012)	2 (0.025)
	86	-	6 (0.074)	19 (0.238)
	88	3 (0.035)	2 (0.025)	3 (0.038)
	90	3 (0.035)	6 (0.074)	32 (0.4)
	92	19 (0.221)	9 (0.111)	7 (0.088)
	94	17 (0.198)	12 (0.148)	2 (0.025)
	96	1 (0.012)	2 (0.025)	-
	98	3 (0.035)	4 (0.049)	-
	100	3 (0.035)	9 (0.111)	-
	n	*86*	*81*	*80*
DXS8087	279	30 (0.341)	20 (0.247)	19 (0.25)
	281	5 (0.057)	1 (0.012)	0
	283	4 (0.045)	-	2 (0.026)
	285	25 (0.284)	34 (0.420)	33 (0.434)
	287	12 (0.136)	22 (0.272)	21 (0.276)
	289	9 (0.102)	2 (0.025)	1 (0.013)
	293	2 (0.023)	2 (0.025)	0
	295	1 (0.011)	-	0
	n	*88*	*81*	76

Note. The 8 microsatellite markers are spaced approximately 30cM apart, hence assumed to be unlinked in the Lemba (see text), thus providing 8 independent X-linked loci for analysis

H scores based on haplotype frequencies for the Lemba, Bantu and Yemen-Sena were calculated and are summarised in Table 5.5. Bantus and Yemen-Sena/Yemen-Hadramaut were used as potential host populations for the Lemba using an analogous strategy as Thomas *et al.* (2002). Equivalent estimates for the comparison populations (additionally including Germans as the host population for Ashkenazi Jews, for comparison) were taken directly from the literature (Thomas *et al.* 2002). The mtDNA diversity in the Lemba (0.966) is comparable to that in the Bantus (0.964), higher than in Yemen-Sena (0.929) and slightly lower than in Yemen-Hadramaut (0.988).

5.3.2. Population Differentiation

The results of the exact test of population differentiation based on high-res and low-res mtDNA hg frequencies are summarised in Table 5.6a, results for the low-res mtDNA data are given in parentheses next to their equivalent high-res values. The high-res results show that all 6 populations are significantly differentiated from each other, confirming predictions (see methods section above); the low-res results reveal that the only non-significant comparison is between the Lemba and Bantus (p=0.262). Results for the Y-chromosome are summarised in Table 5.6b. Again, most populations are significantly differentiated from one another at the hg level. There are however some non-significant differences: Yemen-Sena and Yemen Hadramaut (p=0.120), Yemen-Sena and Ashkenazi Cohanim (p=0.845), Yemen-Hadramaut and Ashkenazi Jews (p=0.067), Yemeni-Jews and Ashkenazi Israelites (0.510) and Ashkenazi Jews (p=0.214), and Ashkenazi Israelites and Ashkenazi Jews (p=0.191). As can be seen, the Lemba are significantly different from all of the comparison populations. Table 5.7 summarises the results of the exact test of population differentiation for each of the 8 X-linked loci. Apart from one of the loci (DXS8087) Bantus and Ashkenazi Jews are significantly differentiated, and apart from the results for DXS8091 and DXS8087, the Lemba and Bantus are not significantly differentiated from each other.

Table 5.5. mtDNA Diversity (*h*) and Associated Standard Errors (SE) Within 9 Jewish Populations and Their Hosts[a]

Jewish Populations				*Host Populations*			
Population	*n*	*h*	*SE*	*Population*	*n*	*h*	*SE*
Lem[b]	119	0.966	0.0087	Ban[b]	67	0.964	0.0128
YemJ[c]	65	0.923	0.0165	YemS[b]	29	0.929	0.0264
				YemH[c]	56	0.988	0.0059
EthJ[c]	48	0.971	0.0113	Eth[c]	74	0.994	0.0076
AshJ[c]	78	0.973	0.0069	Germ[c]	174	0.988	0.0031

Notes. Abbreviations as Table 5.2, additionally Germ = German. The choice of host population for the Lemba is not straightforward (see text), however, Bantus, Yemen-Sena and Yemen-Hadramaut are all plausible hosts, therefore Lemba diversity should be compared with these 3 population. Yemen-Sena and Yemem-Hadramaut should both be considered hosts for Yemeni-Jews

[a] The Lemba have been placed with the Jewish dataset to test whether their mtDNA diversity is typical of a Jewish population in comparison to populations hypothesised to be their host

[b] Calculated by the present author based on the frequency of mtDNA sequences in Appendix Table A.7

[c] Taken directly from Thomas *et al.* (2002), Table 2. These calculations were also performed on sequence data

Table 5.6. mtDNA and Y-Chromosome Exact Test of Population Differentiation Using Hg Frequencies

a)

Population	*Lem*	*Ban*	*YemS*	*YemH*	*YemJ*	*Eth*	*EthJ*	*AshJ*
Lem	- -							
Ban	**0.068** (0.262)	- -						
YemS	**0.000 (0.000)**	**0.000 (0.000)**	- -					
YemH	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.035)**	- -				
YemJ	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.000)**	- -			
Eth	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.012)**	**0.000 (0.003)**	**0.000 (0.000)**	- -		
EthJ	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.001)**	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.011)**	- -	
AshJ	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.000)**	**0.000 (0.000)**	-

b)

Population	*Lem*	*Ban*	*YemS*	*YemH*	*YemJ*	*Eth*	*EthJ*	*AshC*	*AshL*	*AshI*	*AshJ*
Lem	-										
Ban	**0.000**	-									
YemS	**0.000**	**0.000**	-								
YemH	**0.000**	**0.000**	0.120	-							
YemJ	**0.000**	**0.000**	**0.000**	**0.008**	-						
Eth	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	-					
EthJ	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.010**	-				
AshC	**0.000**	**0.000**	0.845	**0.007**	**0.000**	**0.000**	**0.000**	-			
AshL	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	-		
AshI	**0.000**	**0.000**	**0.000**	**0.000**	0.510	**0.000**	**0.000**	**0.000**	**0.000**	-	
AshJ	**0.000**	**0.000**	**0.004**	0.067	0.214	**0.000**	**0.000**	**0.000**	**0.000**	0.191	-

Notes. Shown are the p-values. Abbreviations as Table 5.2. Bold text indicates significant comparisons, $p<0.05$. Calculations based on (a) mtDNA hg frequencies in Table 5.2, and in parentheses using the low-res mtDNA hg frequencies in Table 5.2 to increase power in differentiating populations (b) Y-chromosome hg frequencies in Table 5.3.

Table 5.7. X-Chromosome Exact Test of Population Differentiation for the Lemba, Bantu and Ashkenazi Populations Calculated Using Microsatellite Haplotype Frequencies

	Population		
Locus Name		Lem	Ban
DXS1060	Ban	0.179	
	AshJ	0.102	**0.032**
DXS8012	Ban	0.471	
	AshJ	**0.042**	**0.014**
DXS1220	Ban	0.605	
	AshJ	**0.000**	**0.000**
DXS8091	Ban	**0.009**	
	AshJ	**0.000**	**0.000**
DXS8027	Ban	0.485	
	AshJ	**0.000**	**0.000**
DXS8082	Ban	0.528	
	AshJ	**0.000**	**0.000**
DXS1192	Ban	0.629	
	AshJ	**0.001**	**0.000**
DXS8087	Ban	**0.007**	
	AshJ	**0.005**	0.665

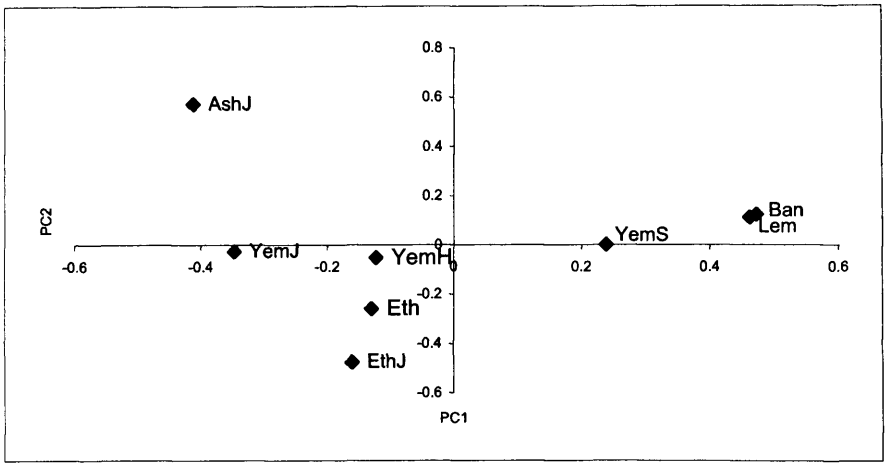
Notes. Shown are the p-values, significant comparisons ($p<0.05$) are indicated in bold). Abbreviations as Table 5.2. Calculations based on the haplotype frequencies in Table 5.4

5.3.3. Principal Components Analysis

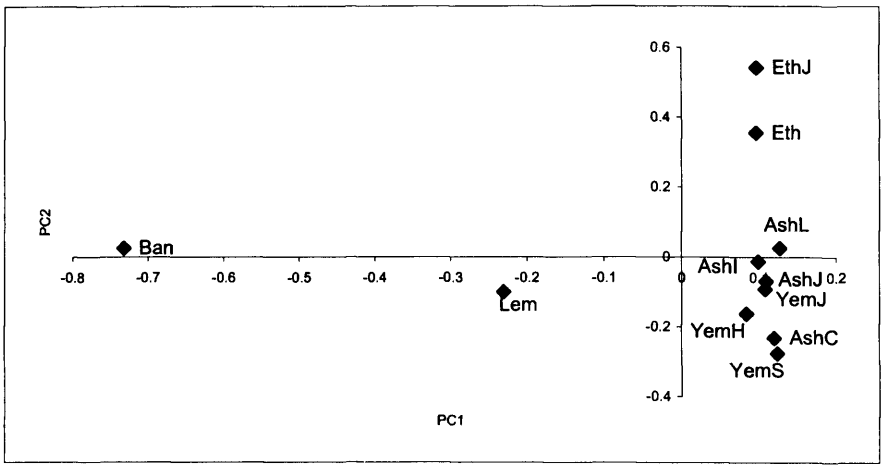
The PC plot drawn from mtDNA hg data shows two main poles towards which populations are drawn (Figure 5.2a), primarily reflecting the amount of non-African vs African sequences in each population. PCs 1 and 2 describe 44.6% of the variation. PC1 is driven by frequencies of several hgs: K (considered non-African [Torroni *et al.* 1996], and found at high frequencies in Ashkenazi Jews [Behar *et al.* 2004a]), L1a, L1d, and L21ab (i.e. L-hgs, hence predominantly African, Richards *et al.* 2003; Salas *et al.* 2002; Chen *et al.* 1995; Rando *et al.* 1998; Passarino *et al.* 1998). Ashkenazi Jews are at the negative extreme of PC1 with high frequencies of hg K, in accordance with Behar *et al.* (2004a), and lower frequencies of African hgs. Bantus and Lemba are at the opposite extreme. PC2 is primarily driven by the frequencies of hgs H and K, and L2. Bantus and the Lemba are placed very close to each other on PCs 1 and 2. The PC plot based on Y-chromosome hg frequencies is shown in Figure 5.2b, PCs 1 and 2 explain 68.7% of the variation. PC1 distinctly separates the populations; Bantus fall at the negative extreme and the remaining populations (apart from the Lemba who fall intermediately) at the other extreme; frequencies of E3a (a predominantly sub-Saharan African lineage, [Semino *et al.* 2004]) drive PC2 with the Bantus exhibiting the highest frequencies (Table 5.3). PC2 shows a trend from the top right to the bottom right of the plot and primarily reflects differences in the frequency of E*(xE3), A*(xA2) and BR*(xDE,JR), and differentiates Ethiopians/Ethiopian Jews from the Ashkenazi and Yemeni populations with the Lemba and Bantus placed intermediately on this gradient.

6 X-linked microsatellites were used to draw a PC plot (Figure 5.2c) using the 3 populations with available data (Lemba, Bantu, and Ashkenazi Jews). The first PC shows a distinct split between the Lemba and Bantus on the one hand and the Ashkenazim on the other, with strong support (PC1 explains 82.7% of thevariation). The results of PC2 are less straightforward as the Lemba and Bantu are separated at opposite poles and the Ashkenazi fall between the two, suggesting that the Lemba and Ashkenazim are more similar to each other than the Lemba are to the Bantus on this PC.

(a)



(b)



(c)

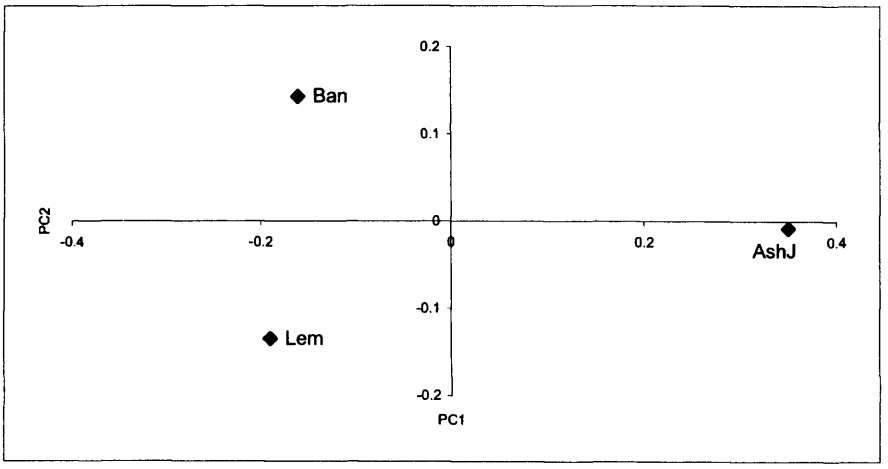


Figure 5.2. PC Plots of the Lemba and Comparison Populations for mtDNA, Y-Chromosome, and X-Chromosome Data. PC plots were drawn using the following data (a) mtDNA (high-res) hg frequencies in Table 5.2, PC1 explained 25.1% of the variation and PC2 explained 19.5%. (b) Y chromosome hg frequencies in Table 5.3, PC1 explained 36.8% of the variation and PC2 explained 21.6%. (c) the first 6 X-linked loci in Table 5.4 (DXS 1060, 8027, 8012, 8082, 1220, 1192). PC1 explained 82.7% of the variation and PC2 explained 17.3%. Abbreviations as Table 5.2

5.3.4. Admixture Analysis

Median admixture estimates and 95% credible intervals are summarised in Table 5.8 and Figure 5.3a. The median admixture estimates using the mtDNA and Y-chromosome data reveal very similar inputs for P_1 and P_2 populations for both mtDNA and Y-chromosome, with estimates ranging from 0.416 to 0.493, but as the credible intervals are large the median proportions are not precise. An exception to this general pattern is the calculation of mtDNA admixture proportions where P_1=Yemen-Hadramaut; here the credible intervals are narrow and the median input for Yemen-Hadramaut is 0.027. The increased precision of this estimate is likely to be the result of greater differentiation between the mtDNA hgs observed in Yemen-Hadramaut compared to both the Lemba and Bantus than is seen in a comparison between Yemen-Sena and the Lemba and Bantus. The median p_1 admixture proportion calculated for the multiple X-linked loci is 0.062, and the credible intervals are narrow. Median input for Bantus (i.e. p_2, where p_2=1-p_1) thus varies from ~0.50 to ~0.95 depending on which locus is used, and which population represents P_1. The posterior pdfs for the range of t_i are shown in Figure 5.3b-d and can be used to infer drift in each population since admixture. In all cases the Y-chromosome (plotted in Figure 5.3a-d as the green line when Yemen-Hadramaut is P_1 and orange line when Yemen-Sena is P_1) has experienced most drift, visualised as the extremely flat distribution of the posterior pdf for all populations. Estimates using Yemen-Sena clearly show least precision, possibly due to the small sample size. As expected from the multiple locus X-linked data (blue lines) these show increased precision. mtDNA performs relatively well in this context, as the posterior pdfs are quite similar to those for the X-linked loci.

5.3.5. mtDNA Hgs and Haplotypes

The Bantus have exclusively L-hgs, which conforms to expectations for a sub-Saharan population (Torroni *et al.* 1996; Rando *et al.* 1998; Semino *et al.* 2002;

Table 5.8. Admixture Proportions for the Lemba Calculated for mtDNA, Y-Chromosome and X-Chromosome Data

Population (locus)	*n*	*Admixture Proportion*	*Founders*	*2.5%*	*97.5%*
Lemba (mtDNA)	29	0.416	YemS	0.031	0.892
	56	0.027	YemH	0.027	0.136
Lemba (Y-chromosome)	27	0.470	YemS	0.019	0.968
	62	0.493	YemH	0.035	0.956
Lemba (X-chromosome)	-	0.062	AshJ	0.004	0.193
	-				

Notes. P_2 is consistently Bantu, P_1 populations for mtDNA and Y chromosome are Yemen-Sena and Yemen-Hadramaut alternately, and for the X chromosome p_1 is Ashkenazi Jews. No single sample size is given for the X-chromosome data as each loci has different sample sizes (see Table 5.4). LEA calculations were performed using the frequencies found in Table 5.2 (mtDNA), 5.3 (Y chromosome) and 5.4 (X chromosome).Abbreviations as Table 5.2

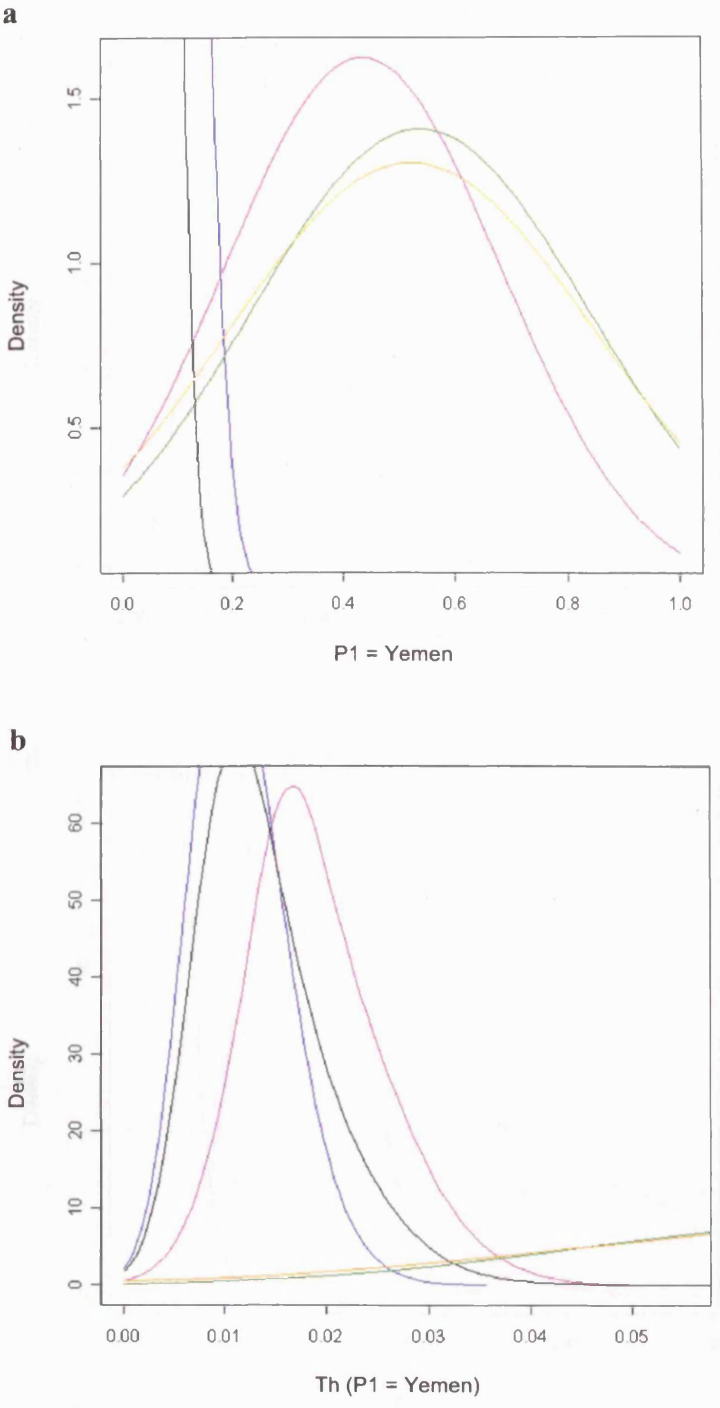


Figure 5.3. Posterior pdf's for p_1, t_h, t_1 and t_2 Calculated with LEA. (a) Posterior pdf's for p_1. Bantus are used as P_2 for all simulations and Yemen-Sena and Yemen-Hadramaut alternatively for P_1 mtDNA and Y-chromosome data. Key: purple line (mtDNA where P_1=Yemen-Sena), black line (mtDNA where P_1= Yemen-Hadramaut), orange line (Y-chromosome where P_1=Yemen-Sena), green line (Y-chromosome where P_1=Yemen-Hadramaut), and blue line (X- chromosome where P_1=Ashkenazi Jews. (b) Posterior pdf's for t_h (Lemba) based on simulations for each of the loci and P_1 populations shown in (a); colours as in (a).

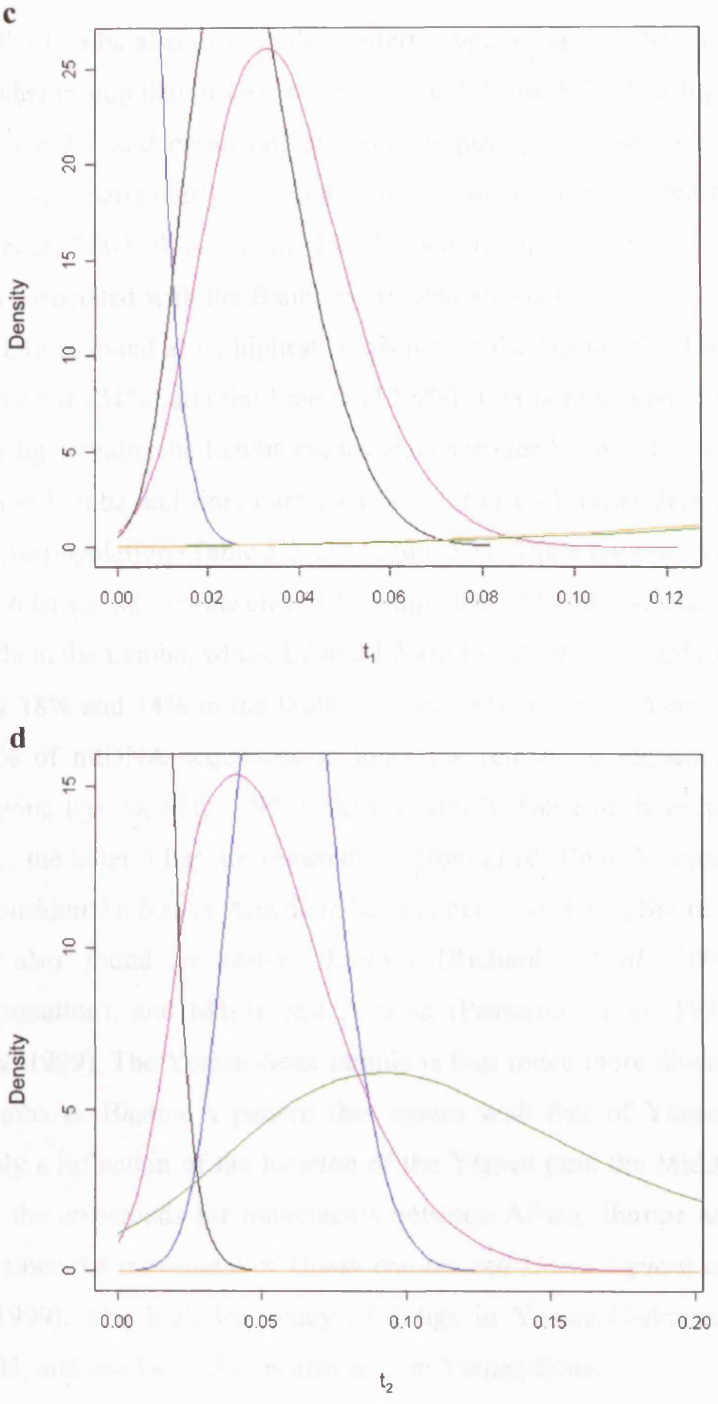


Figure 5.3. continued (c) posterior pdf's for t_1 based on simulations for each of the loci and P_1 populations shown in (a). Key: purple line (mtDNA where P_1=Yemen-Sena), black line (mtDNA where P_1= Yemen-Hadramaut), orange line (Y-chromosome where P_1=Yemen-Sena), green line (Y-chromosome where P_1=Yemen-Hadramaut), and blue line (X-chromosome where P_1=Ashkenazi Jews. (d) posterior pdf's for t_2 (Bantus) based on simulations for each of the loci and P_1 populations shown in (a), key as in (c).

Salas *et al.* 2002), the Lemba also display this pattern, suggesting that they too are a typical sub-Saharan population (see Figure 5.1 and Table 5.2). The high frequencies of L1 and L2 and comparatively low frequency of L3A in the Lemba and Bantus are particularly characteristic of south-eastern African populations (Salas *et al.* 2002; Rando *et al.* 1998). Two hgs in particular (L1a and L2a) have been associated with the Bantu expansions (Salas *et al.* 2002). In the present dataset L1a is found at its highest frequency in the Bantus (27.3%), followed by Yemen-Sena (31%) and the Lemba (12.6%). L2a is most common in the Lemba as this hg contains the Lemba modal sequence (see below). In their lack of non L-hgs the Lemba and Bantu are most similar to each other than to any other comparative population (Table 5.2 and Figure 5.1). There are however differences in the relative hg frequencies. L1 comprises 68% of the Bantu sample but only 35% in the Lemba, whilst L2 and L3 are found 39% and 25% in the Lemba but only 18% and 14% in the Bantu, respectively. L-hgs in Yemen-Sena comprise 66% of mtDNA sequences in total, the remaining sequences belong to the following hgs: D, M1, I, N*, U5a1a, R and X. None of these hgs are African specific; the latter 5 hgs are Eurasian (Torroni *et al.* 1996; Richards *et al.* 2000), D is considered a Native American hg (Forster *et al.* 1996; Smith *et al.* 1999) but is also found in eastern Eurasia (Richards *et al.* 2000, supplementary information), and M1 is east African (Passarino *et al.* 1998; Quitana-Murci *et al.* 1999). The Yemen-Sena sample is thus much more diverse than either the Lemba or Bantus, a pattern that agrees with that of Yemen-Hadramaut, probably a reflection of the location of the Yemen (and the Middle East in general) at the crossroads for movements between Africa, Europe and Asia for millennia since the movement of *Homo erectus* and *Homo sapiens* out of Africa (Klein 1999). The high frequency of L-hgs in Yemen-Hadramaut (Richards *et al.* 2003, and see Table 5.2) is also seen in Yemen-Sena.

107 different mtDNA sequences were found in the 3 populations typed here (Lemba, Bantu, Yemen-Sena), most of which are at low frequency (see Appendix, Table A.7). Including the comparison populations 2698 different sequences are observed, again mostly at low frequency. Shared sequences with a frequency of at least 5% in one population were investigated to infer possible sources of ancestry for the Lemba, summarised in Table 5.9. No sequences

Table 5.9. mtDNA Sequences Found at a Frequency of 5% or More and Shared Between at Least Two Populations

Sequence Number[a]	*Hg*	*HVSI Sequence (16,060-16,390)*	*Lem (n=119)*	*Ban (n=67)*	*YemS (n=29)*	*YemH (n=56)*	*YemJ (n=65)*	*Eth (n=79)*	*EthJ (n=48)*	*AshJ (n=78)*
1/2/3	H	0	-	-	-	7.14%	-	-	4.35%	10.26%
8	HV1	67 274	-	-	-	-	20%	-	-	1.28%
49	K	93 224 311	-	-	-	-	-	1.35%	-	6.41%
70	L3b2	124 223 278 311 362	-	1.49%	-	-	-	-	8.70%	-
73	L3d1	124 223 319	6.72%	-	6.90%	1.79%	-	-	-	-
91	pre-HV	126 304 362	-	-	-	1.79%	12.31%	-	-	-
92	pre-HV	126 305T 362	-	-	-	-	-	1.35%	13.04%	-
104	L1a	129 148 168 172 187 188G 189 223 230 278 293 311	3.36%	14.93%	17.24%	-	-	-	-	-
167	L1a	148 172 187 188G 189 223 230 311 320	7.56%	9.00%	13.79%	5.36%	-	-	-	-
206	L2a1b	182C 183C 189 223 278 290 294 309 390	12.60%	1.49%	-	3.57%	-	-	-	-
222	L2a	189 192 223 278 294 309	0.84%	-	-	-	-	-	8.70%	-
246	L2a1	223 278 286 294 309 390	2.52%	-	17.24%	-	-	-	-	-

Note. The CRS is shared between YemH, AshJ, and EthJ at a frequency of at least 5% in the former 2 populations, but as these sequences belong to different RFLP-defined hgs they are not listed here. A full list of the haplotypes observed can be found in Appendix Table A.7. Abbreviations as Table 5.2

[a] Sequence number refers to the HVS1 sequences listed in Appendix Table A.7

(>5% frequency) were shared between the Lemba and either Ashkenazi Jews, Ethiopians, or Yemeni-Jews. One sequence was shared between the Lemba and Ethiopian Jews (Sequence 22). 3 Lemba sequences (Sequences 73, 167, and 206 in Table A.7) are found in 5% or more of the individuals. Sequence 73 is a L3d1 sequence (with the motif 124-223-319) and is found in 8 Lemba individuals and 2 individuals from Yemen-Sena and 1 individual from Yemen-Hadramaut. Sequence 167 (148-172-187-188G-189-223-230-311-320) is found in 9 Lemba and belongs to L1a. It is shared with 7 Bantu, 4 Yemen-Sena individuals, and 3 individuals from Yemen-Hadramaut. Finally sequence 206, a L2a1b sequence (182C-183C-189-223-278-290-294-309-390) forms the Lemba modal sequence (LMS) being found in 16 Lemba individuals (12.6%); the type is shared with 1 Bantu and 2 Yemen-Hadramaut individuals. The modal Bantu haplotype (sequence 104) is present in both the Lemba and Yemen-Sena. Yemen-Sena has a bi-modal sequence distribution (sequence 104 and 246), one of which is the Bantu modal type, the other is not present in Bantus or Yemen-Hadramaut but is found in the Lemba in 3 individuals. The modal Yemen-Hadramaut sequence (sequence 1/2/3) is shared with Ethiopian Jews and Ashkenazi Jews.

5.4. Discussion

This study aimed to assess the maternal origins of the Lemba, a southern African Bantu-speaking population. An analysis of the mtDNA lineages presented here shows that all lineages in the Lemba are L-hgs, which suggests a wholly African maternal origin, hence ruling out a Jewish descent. This is in contrast to findings from the Y-chromosome, which concluded that most paternal lineages were either Jewish or Semitic (Spurdle and Jenkins 1996; Thomas *et al.* 2000). However, the high frequency of L-hgs in the Yemen, a hypothesised source of the Lemba, leads to problems in distinguishing whether the Lemba's origins lie in Africa or the Yemen. mtDNA diversity in the Lemba was first investigated to ascertain whether their pattern of diversity is consistent with that of several other Jewish populations, which have been shown to have reduced diversity compared to their geographic hosts (Thomas *et al.* 2002; Richards *et al.* 2003; Behar *et al.*

2004a). The choice of host population for the Lemba was not straightforward however. Bantus are an obvious first choice because the Lemba currently reside in close proximity to Bantus in southern Africa and speak a Bantu language (Johnston 2003), although this might have replaced a mother tongue through social pressures (Pew Hispanic Centre Report 2004). Yemen-Sena and Yemen-Hadramaut could also be host populations if the Lemba recently migrated from the Middle East (Parfitt 1997; Spurdle and Jenkins 1996; Thomas *et al.* 2000; Hammer *et al.* 2000). In comparison with Bantus and Yemen-Sena, the pattern of mtDNA diversity in the Lemba is atypical of a Jewish population as the Lemba have levels of sequence diversity equal to, or greater than, Bantus and Yemen-Sena. The frequency of the LMH is slightly lower compared to the modal sequences observed in the Bantus and Yemen-Sena, again contradicting the pattern seen in other Jewish populations. Compared to Yemen-Hadramaut in contrast, the Lemba do have reduced haplotype diversity, and a high frequency modal sequence. Using the same dataset as that analysed here, Richards *et al.* (2003) noted that Yemen-Hadramaut had experienced high levels of primarily African gene flow compared to other Near Eastern populations, which may explain why the Lemba have reduced diversity compared to Yemen-Hadramaut. Indeed the high levels of Yemen-Hadramaut diversity can be clearly visualised in Figure 5.1. Hence, depending on the choice of host population, the pattern of Lemba mtDNA diversity argues both for and against potential Jewish origins, confounded by the lack of historical records relating to the Lemba's origins.

To clarify matters, three other lines of evidence were used to infer the association, or otherwise, between the Lemba and the comparative Jewish populations, all of which indicate that the Lemba do not have an mtDNA hg composition similar to that of the Jewish populations. First, mtDNA PC plots show that the Lemba do not cluster with any of the Jewish populations on either PC1 or PC2. This can be contrasted with the PC plot of Y-chromosome hg frequencies, where PC1 shows the Lemba drawn towards the Jewish and Middle Eastern populations, in accord with the high frequencies of Jewish and Middle Eastern Y-chromosome types in the Lemba, both in this dataset (Thomas *et al.* 2000) and other datasets (Spurdle and Jenkins 1996; Hammer *et al.* 2000). Second, the Lemba are significantly different from all of the Jewish comparison

populations. Finally, only one sequence (either found at a frequency of ≥5% or at any frequency) is shared between the Lemba and any Jewish population (Ethiopian Jews) and this only appears as a singleton in the Lemba dataset suggesting a single recent introgression event or a mutation in the Lemba sequence. It is thus unlikely that the Lemba are related to Ethiopian Jews, as has been suggested (Parfitt 1997). That the Lemba's Jewish identity is male, rather than female mediated is unusual given the usual matrilineal inheritance of Jewish identity, apart from the Cohen and Levite male-inherited castes (Encyclopaedia Judaica 1972). Intriguingly this does correlate with the presence of the male-inherited CMH in the Lemba (Thomas *et al.* 2000). Drift may of course have eradicated any low frequency Jewish lineages, as low frequency alleles are more prone to being lost by drift (e.g. Tishkoff and Verrelli 2003). It is thus very unlikely that all or even a high proportion of Lemba mtDNA sequences were of Jewish origin, and subsequently lost by drift.

The next question to address therefore, is which population represents the most likely maternal parental source of the Lemba. Potential sources are Bantus, Yemen-Sena, and Yemen-Hadramaut, based on the elimination of several Jewish populations above, and suggestions from genetics (Soodyall *et al.* 1996), serology (Hughes *et al.* 1978) and historical and ethnographic accounts (detailed by Hughes *et al.* 1978; Buijs 1998; Parfitt 1997). Detecting the relative influence of different parental populations on a given hybrid population depends on the extent to which the parental populations are differentiated (Bertorelle and Excoffier 1998). It is difficult to differentiate these Yemeni populations from the Lemba and Bantus as 65% of the Yemen-Sena mtDNA sequences and around 35% of Yemen-Hadramaut sequences are L-hgs. The Arab slave trade, which is hypothesised to be the reason for the high frequency of L-hgs in Yemen-Hadramaut (Richards *et al.* 2003), must also explain the even higher frequency of L-hgs in Yemen-Sena, a population that has strong ties with Africa today (Parfitt 1997). Full sequencing of the mitochondrial genome may help differentiate the L-hgs in Yemen-Sena and Bantus, and therefore disentangle their relative influences on the Lemba.

PC analysis and the exact test of population differentiation based on mtDNA hg frequencies (the latter using low resolution hg frequencies) suggests that Bantus have contributed a large proportion of mtDNA lineages to the Lemba, as these two populations are virtually indistinguishable on both PC1 and PC2 and they are not significantly differentiated from each other at the hg level (p=0.262). This can be contrasted to the PC of Y-chromosome hg frequencies where the Lemba fall midway between the Bantus and the Semitic and Ethiopian populations. The high proportion of mtDNA L-hgs in Yemen-Sena is also evident in the PC analysis because of all the Semitic populations, including Yemen-Hadramaut, Yemen-Sena falls closest to Bantus and the Lemba. PC plots therefore indicate that after Bantus, Yemen-Sena is more likely than Yemen-Hadramaut to be a maternal source population for the Lemba. The PC plot drawn for the X-linked microsatellites also confirms the close relationship between Bantus and the Lemba on PC1, conversely PC2 presents a more complex picture.

An analysis of shared modal or rare sequences between populations can be an informative way to assess finer details of similarity between populations, hence shared ancestry. However, the extent of sequence sharing between Bantus, Yemen-Sena, Yemen-Hadramaut and the Lemba is such that the relative inputs of the latter three populations on the Lemba is difficult to distinguish in much the same way that was found for the L-hgs. The LMH (sequence 206) is found in both Bantus and Yemen-Hadramaut at low frequency (0.015, 0.034; 1 and 2 individuals respectively), which does not aid in distinguishing Bantu from Yemeni inputs. The LMH belongs to hg L2a1b, which appears to be a south eastern African lineage associated with the Bantu expansions (Salas *et al.* 2002). In 3 Mozambican Bantu speaking populations from the south east and south west of Mozambique it comprised the commonest L21a1b sequence as well as the single most common sequence in the entire dataset found in 18-20% of the sampled individuals (Salas *et al.* 2002), whilst Yemen-Hadramaut is the only Middle Eastern population where the LMH has been found (M B Richards, unpublished data). The high frequency of the LMH in the Mozambican Bantu speakers and the contrasting low frequency of the type in the Yemen and Middle East in general, suggest that the LMH was more likely contributed to the Lemba

by Bantus, possibly from Mozambique, rather than Yemeni women. The modal haplotype in the Bantus is found in 4 Lemba individuals as well as 5 Yemen-Sena individuals (where it forms one of the modal types in Yemen-Sena), therefore it is again difficult to determine which of these populations contributed the type to the Lemba. One final interesting point to make regarding haplotype sharing is that whilst sequences are shared between the Lemba, Bantus and one or both of the Yemeni populations, one sequence is shared exclusively between the Lemba and Yemen-Sena, and the sequence comprises the 2nd Yemen-Sena modal type (sequence 246). This again hints at a closer relationship between Lemba and Yemen-Sena than with Yemen-Hadramaut.

Admixture proportions for the Lemba were thus estimated using Yemen-Sena, Yemen-Hadramaut, and Bantus as parental populations and employing a likelihood based approach (LEA). An advantage of LEA over methods such as PC analysis and haplotype sharing, for assessing the origins of the Lemba is that the admixture model explicitly takes into account the effects of drift since the admixture event, sampling variance, and uncertainty on the estimation of ancestral allele frequencies (Chikhi *et al.* 2001). All of these factors have the potential to affect the present day distribution of genetic types in the potential parental populations and the hybrid. A drawback of LEA estimates, as well as those from other estimates of admixture proportions (e.g. Bertorelle and Excoffier 1998), are the associated wide credible intervals (Chikhi *et al.* 2001; Chikhi *et al.* 2002). In particular, single locus mtDNA and Y-chromosome data are expected to have wide credible intervals and reduced reliability (Chikhi *et al.* 2001; Chikhi *et al.* 2002), which was typically seen with the present data, except in the mtDNA calculation that involved a large number of alleles (when P_1=Yemen-Hadramaut, see results section above). The wide credible intervals observed here, and in other work, for mtDNA and Y-chromosome data may also be a realistic reflection of the paucity of information contained in the data, although coming to this conclusion leads to a more general debate regarding the value of mtDNA and Y-chromosome data in studying human history. A detailed discussion of this issue is not the focus of the present work, rather the fact that these loci have proved useful in answering questions of human history (see for example Cavalli-Sforza and Feldman 2003).

The wide credible intervals encountered here might explain why the admixture proportions for mtDNA and Y-chromosome data (apart from mtDNA where P_1=Yemen-Hadramaut) do not reflect PC plots drawn for the two loci. Median $p1$ admixture proportions for both loci are ~0.5 making it difficult to differentiate which Yemeni population is the most likely P_1, hence obscuring the clear sex biased gene flow previously discussed. In particular, the pdf's for t_i indicate that the Y-chromosome data has experienced most drift since admixture rendering these estimates the least reliable. The two reliable admixture estimates (mtDNA where P_1=Yemen-Hadramaut, and X-chromosome) show a strong trend for female Bantu input being extremely high, and approaching 100%. As the X-chromosome is skewed towards reflecting female input in admixture events (Jobling *et al.* 2003) it can be reliably used to accompany estimates from mtDNA. It should be noted however that the two admixture estimates indicating the highest Bantu input use Yemen-Hadramaut (mtDNA) and Ashkenazi Jews (X-chromosome) as P_1 rather than Yemen-Sena, which will strongly affect the estimates if these populations are less likely parental populations than Yemen-Sena. Ashkenazi Jews have indeed been ruled out as a maternal parental population based on several analyses above.

Considering the results discussed above, it seems that whilst the maternal lineages of the Lemba are entirely African in origin, they may have been contributed to the Lemba by African (Bantus) as well as non-African (Yemeni) women, particularly Yemen-Sena. If, on the basis of the Y-chromosome data, it is assumed that the majority of Lemba men have a Semitic or Jewish ancestry outside Africa it is thus feasible that some Lemba women from their ancestral home migrated with the Lemba men, traces of which can be seen in the mtDNA sequences shared exclusively with the Lemba and Yemen-Sena populations. This latter finding does corroborate Parfitt's (1997) assertion that Yemen-Sena is the source location of the Lemba. Under this scenario gene flow from Bantu women into the Lemba, once they arrived in Africa, introduced other mtDNA sequences, such as the LMH. If some of the maternal lineages of the Lemba are thus derived from the Yemen it is interesting to note that none of the Eurasian mtDNA sequences observed in the Yemen are found in the Lemba. This could be the result of either inadequate sampling of the Lemba, or drift eradicating

such lineages. The Lemba do not have the reduced genetic diversity associated with extensive drift, as observed in the mtDNA lineages of many Jewish populations (Thomas *et al.* 2002; Richards *et al.* 2003; Behar *et al.* 2004a) as well as other populations known to have experienced drift (Gresham *et al.* 2001; Soodyall *et al.* 1997). However, high levels of gene flow from the Bantu could mask earlier drift and indeed explain the degree of similarity between the Lemba and Bantus. The large genetic distance over which LD in the Lemba extends (≤21 cM) suggests that admixture has been very recent, as only 3 generations are required to reduce the amount of LD at 20cM by one half (Wilson and Goldstein 2001); this obviously reduces the timescale for drift. If one equates the Lemba admixture event with their arrival in Africa (which may or may not be a valid assumption) such a recent admixture event in the Lemba is slightly questionable on ethnographic grounds; given the paucity of details they have about their origins one might expect admixture to have occurred many generations ago. Substructure within the Lemba may thus better explain the degree of LD, although this was not thought to be an important factor (Wilson and Goldstein 2001).

Despite the conclusions of several authors (Seielstad *et al.* 1998; Perez-Lezuan *et al.* 1999; Oota *et al.* 2001), that in human history females have tended to migrate more than men because of the practice of patrilocality, it is possible to find strong evidence in the literature of male migration events which have introduced non-indigenous Y-chromosomes into the local gene pool (Hurles *et al.* 1998; Carvajal-Carmona *et al.* 2000). In these two examples the non-indigenous Y-chromosomes have been of European origin and brought to Polynesia and Colombia, respectively, by male European colonizers. In both cases mtDNA analysis indicates that in contrast most female lineages are indigenous. Such evidence is congruent with the general perception that it is usually men who travel far from their natal home to conquer and colonise new lands. The Lemba appear to fit into this category with greater evidence for male than female migration from a Jewish/Semitic population. Such data are of course still compatible with the findings of Seielstad *et al.* (1998), Perez-Lezuan *et al.* (1999), Oota *et al.* (2001) if one assumes that the effect of these one-off

migration events on the overall pattern of mtDNA and Y-chromosome diversity within and between human populations has been slight compared to the cumulative effect of patrilocality over successive generations.

5.5. Conclusions

Based on the present distribution of mtDNA sequences it has been possible to exclude a Jewish population as a maternal source for the Lemba, contrasting with findings for the Y-chromosome. Lemba mtDNA sequences are exclusively of African origin, initially suggesting that Bantus have had the biggest impact on the Lemba mtDNA gene pool. Due to the similarity of Yemeni and Bantu mtDNA lineages it is difficult to differentiate between the influences of Yemen-Sena and Yemen-Hadramaut and Bantu however this is confounded by the lack of contemporary written records relating to the Lemba's history. The data that is available points to both Bantu and Yemeni populations contributing mtDNA lineages, with Yemen-Sena having a greater input than Yemen-Hadramaut.

Chapter 6. Discussion

6.1 General overview

Genetic approaches in human evolution have been successfully used to study the history of *H. sapiens* (see for example the recent review by Cavalli-Sforza and Feldman 2003). There has been a tendency to study ancient events in our history, such as the migrations of humans out of Africa (Hammer 1995; Underhill *et al.* 2000; Underhill *et al.* 2001; Penny *et al.* 1995; Chen *et al.* 1995; Harpending *et al.* 1998; Ingman *et al.* 2000; Goldstein *et al.* 1995a; Antunez-de-Mayolo *et al.* 2002), and the dispersal of Bantu-speaking peoples within Africa (Salas *et al.* 2020). The reasons for this focus on ancient history have been discussed in detail in the Introduction (section 1.3). Initially classical markers were used in the study of genetic history (see for example Cavalli-Sforza *et al.* 1994), as detailed in the Introduction (section 1.3). However as increasing numbers of polymorphic markers on the Y-chromosome and mtDNA were identified and their geographic distribution assessed these two loci started to be applied to questions of human history. As many studies have shown, the Y-chromosome and mtDNA have had considerable success in the field of genetic history (see for example Jobling and Tyler-Smith 2003 for a summary for the Y-chromosome and Maca-Meya *et al.* 2001 for mtDNA). The increased number of polymorphisms available on these loci means that questions relating to recent human history (i.e. within the scope of written records and oral history) are more easily tackled, although there are some instances of classical markers being successfully used to study recent history (see for example Cavalli-Sforza et al 2004).

Studies that focus on recent events fall into two categories, those that investigate events localised to a small geographic region, and those that study events involving geographically disparate populations (see Wilson *et al.* 2001a and Carvajal-Carmona *et al.* 2000 as examples of these two extremes). These two scenarios have different expectations about the (genetic) similarities or differences of the populations involved. In the former case the populations might be expected to be typically more similar to each other than in the latter example, because homogenising events such as migration between populations are more

likely to occur over smaller geographic distances (Cavalli-Sforza *et al.* 1994). There are some instances where neighbouring populations can be isolated from each other, hence genetically distinct. The Basques are a much cited example in Europe (Calafell and Bertanpetit 1994; Comas *et al.* 2000; Cavalli-Sforza *et al.* 1994; Chikhi *et al.* 2002; but also see Hurles *et al.* 1999). However, if one makes the assumption that in many cases geographically close populations are more similar to each other than geographically separated populations then the latter of the two scenarios described at the start of the paragraph will be intuitively easier to investigate genetically. In contrast, in the former example the geographic proximity of the populations under investigation means they are expected to be genetically less distinct (Cavalli-Sforza *et al.* 1994). A small number of studies have focussed on the former class of events, such as the impact of Anglo-Saxons on the male British gene pool (Weale *et al.* 2002), and the genetic relationship of men with the Sykes surname (Sykes and Irven 2000). However, due to the relatively small number of such studies it is not apparent whether recent events in human history that involve geographically close populations can, in general, be successfully analysed using existing Y-chromosome and mtDNA polymorphisms, or if the level of differentiation between the populations involved is too small to make meaningful inferences.

Recent events in the history of the British male and female populations, and the oral history of a Bantu-speaking African population were analysed in this thesis. The gene pool of British men, sampled from small towns across the British Isles, was found to be differentially affected by the invasion of Anglo-Saxon/Danish and Norwegian men. For many of the sample locations the genetic results corroborated historical predictions about the amount of influence of these invading populations. However for some locations there were clear discrepancies, such as for the Isle of Man and Rush (used as a proxy for Dublin). Nonetheless these results showed that it was possible to identify these different influences. Such local patterns of variation had been suggested by other studies (Wilson *et al.* 2001a; Weale *et al.* 2002) but were only fully apparent with the novel, comprehensive, sampling strategy implemented to cover the British Isles (Chapter 2).

With this thorough pattern of British Y-chromosome diversity in place, two further questions that required an in-depth knowledge of British Y-chromosomes were addressed. The first of these looked at the question of whether the analysis of men with the same surname suggested that surnames were simply random samples from the British population or if each name had a discrete history. A variety of surnames were considered to allow a more generalised approach to the study of surnames than the eponymous investigation of Sykes (Sykes and Irven 2000). The analysis revealed that it was possible to identify evidence for random versus non-random adoption and some general trends were apparent (Chapter 3). The second question focussed on one small but historically diverse part of Britain, London, to assess the level of diversity observed, in comparison to that seen in the rest of Britain and that predicted from historical sources and Census records. Cities are normally ignored when studying the history of a country or particular region because they are thought to harbour too much diversity to be informative about historical events (Cavalli-Sforza *et al.* 1994). It was this assumed diversity in London that was explicitly studied. Y-chromosome and mtDNA lineages were considered in London, both of which revealed the presence of low frequency lineages not usually observed in British or European populations, confirming expectations about the history of immigration experienced by London. However the majority of sampled Y-chromosome and mtDNA lineages fell within hgs normally expected to be seen in British and European populations.

Finally, mtDNA diversity in the Lemba, who had previously been shown to have a Semitic male component (Spurdle and Jenkins 1996; Hammer *et al.* 2000; and Thomas *et al.* 2000), confirming aspects of their oral tradition, was studied. Previous evidence (Hughes *et al.* 1978; Soodyall *et al.* 1996), taken in conjunction with that of the Y-chromosome data, suggests that the Semitic component was only male mediated. Analysis showed that all Lemba mtDNA sequences were of African origin, however, the high frequency of L-hgs in the Yemen confounded attempts to distinguish an African from a Yemeni component. This last study provided an interesting chance for comparison between a European and an African population, the former of which was only informed by historical and archaeological findings, whilst the latter population

had little conclusive written history but suggestive evidence from studies of the Y-chromosome and oral tradition.

In the Introduction (section 1.2) it was argued that despite recent events being better documented (because of the existence of historical records, oral history, as well as the better preservation of archaeological and fossil remains), such evidence is rarely conclusive and the fidelity of written records cannot be assumed *a priori*. Therefore genetics is just as valuable a resource for recent history as it is for ancient events. This is clearly indicated by the fact that each of the chapters presented here addresses questions in recent history that could not be answered using any other lines of evidence. For example the history of surnames has fascinated people for generations with countless surname societies being formed to determine whether particular names have a common origin. Until now there has virtually been no progress despite the intense interest and historical research being carried out. Genetics, and in particular Y-chromosome polymorphisms, has provided a direct means of testing this.

It is important to note that all of the studies presented here have been able to answer, at least in part, the questions that were posed, hence testifying to the fact that Y-chromosome and mtDNA diversity can be informative for events that have happened recently in human history. This conclusion is particularly pertinent because of one of the potential problems associated with using the neutral markers on the Y-chromosome and mtDNA genome for studies of recent history. If one assumes that the frequencies of Y-chromosome and mtDNA lineages have not been differentially affected by selection (but see for example Krausz *et al.* 2001 for the Y-chromosome and Mishmar *et al.* 2003 for mtDNA) differences between populations only accrue through drift. As drift occurs at a rate that is related to time, for populations are only recently separated the time for drift to have effect is smaller. The populations may also exchange migrants which may lead to the homogenisation of genetic types (Cavalli-Sforza *et al.* 1994). Such migrations are more likely to occur if the populations are geographically close. The former of these issues was addressed by using quickly mutating systems on the Y-chromosome and mtDNA, rather than simply relying

on the more stable but less diverse range of available markers, and the latter by the sample design (which is considered in more detail below).

These considerations meant that in most instances neither of these factors seemed to be a problem, and only in two cases was such a lack of differentiation observed. North German and Danish Y-chromosomes could not be differentiated, which is highly likely to be the result of their geographical proximity, allowing migration, and the fact that they might have shared a recent common ancestor (Chapter 2). Yemeni and Bantu mtDNA lineages also shared a surprising number of sequences (Chapter 5), as the result of female migration associated with the Arab slave trade (Richards *et al.* 2003). In both of these cases the lack of differentiation hindered analysis because the influence of various potential source populations could not be differentiated. Employing further Y-linked microsatellites, from the ≥139 that exist (Kayser *et al.* 2004) and full mtDNA genome sequencing, which has recently been shown to be informative (Ingman *et al.* 2000; Finnilä *et al.* 2001; Torroni *et al.* 2001a; Maca-Meyer *et al.* 2003), will likely aid in distinguishing these populations.

As discussed in more detail in Chapters 2-5, the results of the genetic analyses presented both agree with and contradict predictions from other disciplines. Where all lines of evidence agree on a particular event, one can be fairly certain that it is "real". For example, the fact that historical, archaeological, and linguistic evidence suggested that Norwegian Vikings had a considerable influence on Shetland and Orkney has been confirmed by evidence from the Y-chromosome which showed an enrichment of Norwegian Y-chromosome types in samples from Orkney and Shetland (Wilson *et al.* 2001a; Chapter 2 and Capelli *et al.* 2003). It was also possible to show that the MacLeod surname has more than likely had a single origin, as proposed by Clan history (Morrison 1986, and see Chapter 3), and that London Y-chromosome and mtDNA lineages exhibit higher levels of diversity than available comparison populations, as suggested by historical (Inwood 1998; Ackroyd 2000) and Census records (Source: National Statistics website: www.statistics.gov.uk, Crown copyright

material is reproduced with the permission of the Controller of HMSO; Dobson and McLaughlan 2001).

Instances of disagreement are arguably more noteworthy than those of agreement and could indicate important events that need to be investigated further (Hurles and Jobling 2001). Indeed such examples validate the need for genetic analyses of recent events. The examples of incongruence found here seem to be indicative of one or more of the following scenarios: a "real" difference in the various accounts; recent migration which has obscured patterns of genetic diversity; or inaccurate choice of populations to sample, leading to erroneous results. The issue of sampling is important in many contexts, particularly for recent events, as discussed in the Introduction (Section 1.5). For this reason sampling will be considered next, before returning to examples of incongruence between different sources of data.

6.2 Sampling

In the study of British Y-chromosomes (Chapter 2), the effect on the gene pool of clearly demarked populations was studied (Anglo-Saxon, Norwegian, and Danish). The fact that these 3 potential source populations are well defined from historical records (Hill 1981; Richards 1991; Welch 1992; and Davies 1999) meant it was possible to sample with some precision. However, the choice of population to represent Anglo-Saxons was complicated by uncertainties over their most likely source; Schleswig-Holstein was used here, but Weale *et al.* (2002) chose Frisians. Conclusions are also confounded by the fact that the North German and Danish samples analysed in this thesis could not be distinguished, hence analyses were performed using the combined North German/Danish sample. The present work (Chapter 2, Capelli *et al.* 2003) and Weale *et al.* (2002) draw different conclusions about the amount of gene flow from Anglo-Saxons into the British male gene pool, despite Schleswig-Holstein and Frisia not being significantly different from each other at the hg level ($p=0.3$). However the Frisian sample is more similar to English Y-chromosomes (analysed both in Chapter 2/Capelli *et al.* 2003, and Weale *et al.* 2003) than is

North Germany/Denmark. Hence the conclusions of Weale *et al.* (2003) are conceivably influenced by this finding. Discrepancies in the conclusions of the two studies must also be related to the different methodologies employed. First, hg2 (in the terminology of Jobling and Tyler-Smith 2000) is defined by a higher resolution marker in the present work than by Weale *et al.* (2003), which may be important, as discussed in Chapter 2 (section 2.4.1). Second, the methods of analysis were different. For example the present work inferred admixture proportions using LEA (Chikhi *et al.* 2001) whilst Weale and colleagues investigated the question from the perspective of population splitting using the BATWING programme. The different ways in which these two programmes address the question, and the assumptions that are integral to each of the models, may have also influenced how the data have been interpreted.

A further potential error introduced by sampling is the use of Basques in Chapter 2 to represent the Y-chromosomes of indigenous Europeans in the British gene pool, following well established norms (Calafell and Bertanpetit 1994; Comas *et al.* 2000; Cavalli-Sforza *et al.* 1994; Chikhi *et al.* 2002). However there is evidence that Basques are not as genetically isolated as often assumed (Hurles *et al.* 1999), hence compromising their use as a proxy for the European Palaeolithic population. If this is the case then estimates of the degree of indigenous versus other European types in the British male gene pool might be inaccurate. At present however, Basques still appear to be the best population to use in this context.

Within Britain, sample locations and sample donors were chosen to minimise the effect of recent migration from other British regions and further afield (Section 2.2.1). Small towns, which experience less migration than metropolitan districts (Cohen 2004), were selected to sample from, and donors had to be able to trace their paternal grandfather back to the same region. It would be more reliable to insist that sample donors could trace their male ancestry back even further, however it would have undoubtedly reduced the number of men who could have participated in the study. Therefore a workable compromise, between the number and accuracy of the samples collected, was reached. This sample strategy was rigorous enough to be able to detect small-scale patterns of

variation within Britain in the genetic influence of the European populations on the British male gene pool (Chapter 2) using the designed grid system. Only two instances of recent migration appear to have been a problem, and are discussed in more detail below. Although several aspects of the history of the male genetic history of the Channel Islands could be tackled with the European Y-chromosomes sampled here, it became apparent that samples from Normandy might have been beneficial to attempt to differentiate the effects of gene flow from North German/Danish men and Normans. Given that Normans were descended from Danish Vikings (Davies 1999) however, it was not clear that there would have been considerable differentiation between these populations using the range of Y-chromosome markers employed here. Hence Normandy was not initially sampled from. Subsequent time constraints did not allow these samples to be collected.

The presence of the database of Y-chromosomes from Britain and several European countries meant that Y-chromosome comparison populations for the surnames and Londoners studied here were already in place. Given the patchy and sometimes contradictory information regarding the history of some of the surnames (see Chapter 3), and the hypothesised diversity of London Y-chromosomes (see Chapter 4), it was particularly important that British Y-chromosome be well characterised before the surnames were investigated. However, if conclusions made about the British dataset (the extent of indigenous, Anglo Saxon or Norwegian influence on a particular region for instance) are flawed because of an incorrect sampling strategy, it will consequently affect comparisons of these data with the surnames and London Y-chromosomes. The problems of obtaining an unbiased sample from a particular region were illustrated in the analysis of London Y-chromosomes. The availability of an independently collected set of London DNAs typed for the same Y-chromosome markers as those used here showed that the samples were significantly different from each other at the hg level (exact test of populations differentiation, $p=0.002$). This is probably for a combination of reasons: (i) the amount of diversity in the London gene pool is too great to be accurately captured in a total of 157 Y-chromosomes; (ii) the samples were collected from museums in London, and the subset of the London population that visits

museums and wants to participate in genetics studies is not a representative cross-section of the true London population.

6.3 Historical and Genetic Disparities

As stated above several factors appear to cause disparities between the results obtained from genetic analyses and other data sources, such as archaeology, history and oral tradition: (i) "real" differences, (ii) recent migration, (iii) sampling problems. There are several instances of what appear to be "real" disparity between historical and genetic findings. First, the Isle of Man and the Western Isles do not have evidence for a great enrichment of Norwegian Y-chromosomes (Chapter 2). This is despite the fact that both islands are on the route thought to have been taken by Norwegian Vikings along the west coast of Britain (Richards 1991; Davies 1999), the Isle of Man has an enduring legacy of ties with its Viking heritage (Richards 1991) and that other islands along the same route (Shetland and Orkney) clearly have a genetic enrichment of Norwegian Y-chromosomes (Chapter 2 and Wilson *et al.* 2001a). Very recent immigration to the Isle of Man may explain these findings, as the Isle has favourable tax breaks, however the eligibility of volunteers taking part in the study should deal with this by ensuring that only men who could trace their paternal grandfather to the island participated. Furthermore, available census information suggests that immigration from England was not great before ~1965 (Isle of Man Census, 2001). Immigration to the Western Isles is low, hence it is less likely to be a problem, indeed the isle has recently experienced a net loss of inhabitants ((C) Crown copyright. Data supplied by General Register Office for Scotland). Therefore it is possible that the influence of Vikings on both of these islands did not extend to gene flow with the indigenous inhabitants, perhaps the rule of the Vikings on these Islands was more a matter of elite dominance, therefore.

It was concluded that the surname Whytock could have had an independent origin to the name Whittock, despite evidence suggesting that they were spelling

variants of the same name (Reaney 1997). Despite the small sample size of Whytocks, the fact that all Y-chromosomes belonged to the same, extremely rare, hg (E3b) (which was not observed in the Whittock sample or the Whytock's geographical neighbours at all) and had identical haplotypes strongly supports the conclusion that they have had a single origin, independent of the Whittocks (given the caveat of a fortuitous associated between one of these spelling variants and a non-paternity event). For the Lemba it was possible to show that their mtDNA lineages did not have any evidence for Jewish origins (Chapter 5), which contradicts their oral tradition (Parfitt 1997), although Buijs (1998) has argued that the Lemba's claims to a Jewish heritage is a recent construction. Nonetheless, the lack of identifiable Jewish maternal lineages contradicts findings for the Y-chromosome (Thomas *et al.* 2000).

A lack of Norwegian enrichment was also found at the site of Rush in Ireland. Rush is ~25km north of the city of Dublin, for which it was used as a proxy to control for recent migrations to and from Dublin, a city founded by Norwegian Vikings (Davies 1999). Note however the existence of wars between Danish and Norwegian Vikings in the vicinity of Dublin, although Norwegian Vikings seemed to have won these battles and maintained a longer presence in this region (Davies 1999). As Rush was a proxy, it might explain why the Y-chromosomes that were sampled were not characterised by a Norwegian signature but by high frequencies of R1*(xR1a1), which typifies the sample analysed from Castlerea (Chapter 2) and other samples from Ireland (Wilson *et al.* 2001a; Hill *et al.* 2000) typed for the analogous hgs P*(xR1a1) and P, respectively. It may also be an example of elite dominance of the Norwegians, with potential genetic input from Danish Vikings, although there is not an apparent enrichment of North German/Danish input in Ireland inferred from the admixture analysis. However it is apparent that it may not be possible to differentiate these scenarios; given Dublin's status as the capital of Ireland, and a port, means that it has probably experienced much immigration and migration since the Viking period. Therefore any genetic legacy of the Vikings may have been erased from Dublin and may not be visible in other parts of Ireland. Hence it is not clear whether in this instance the lack of congruence is "real" or not.

In contrast, several documented events of migration appear to have affected some of the findings. The migration of English people to the sample site of Llanidloes in Wales, and mainland Scotland (Davies 1999) seem to be the best explanation for why these locations both have an enrichment of North German/Danish Y-chromosomes, despite limited, or no evidence, to suggest that either the Anglo Saxons or Danish Vikings invaded Wales or mainland Scotland (Davies 1999). As noted above, for the Channel Islands it is impossible to know whether the enrichment of North German/Danish Y-chromosomes seen on Jersey in particular is the result of Danish Vikings, or the better documented period of Norman rule. An additional confounding influence is the German occupation of the Channel Islands during World War II, a period which saw illegitimacy rates raise to around ¼ of births (Briggs 1995). Whilst migration is thus a potential confounder in these contexts, especially when it involves the relatively small movement of people within the same region, it is of course an important topic in its own right. Indeed, the documented long history of immigration to London, with the associated increased levels of genetic diversity, was the starting point for the analysis of Y-chromosome and mtDNA lineages in Londoners. Furthermore, the migration of African women to Arabia (Segal 2001; Richards *et al.* 2003) and even further afield (Quintana-Murci *et al.* 2004) as a result of the Arab slave trade has resulted in Yemeni mitochondrial lineages having high frequencies of African L-hgs, hence confounding attempts to disentangle the possible migration of the Lemba from the Middle East to Africa.

Finally, issues around sampling appear to have directly influenced some of the findings presented in this thesis. Results for two of the surnames, Folland and Speechley, were surprising in finding that the names were random draws from their geographic neighbours given that family history research strongly suggested single origins from specific locations for both of these names, corroborated by highly localised geographic distributions of these names in the British Telecom Telephone Directory (Chapter 3). The Y-chromosome analysis showed that both of these surnames had a modal haplotype that could potentially be interpreted as a founding type, however both haplotypes belonged to the common hg R1*(xR1a1) (and the haplotype in the Speechley sample was AMH+1). Both Folland and Speechley also have relatively small sample sizes, a

result of the rareness of the names in Britain. Therefore, the combination of the rare names (small sample size) and common hg means that it is difficult to interpret the results meaningfully. In this case, typing further microsatellites would also prove useful. The case of Rush, may also be due to problems with sampling methodology, as noted above.

6.4 Future Directions

Y-chromosome and mtDNA studies of human populations have an important future in genetics research for as long as there are questions to be addressed about the history of populations. Furthermore these loci should prove to be extremely useful in studying recent events in human history for populations that are geographically closely related, as shown above. As more markers become available and finer resolution analysis is simpler and more cost effective their use will become commonplace allowing finer dissection of recent human history. Of course one has to acknowledge problems with the field, which would benefit from a better understanding and modelling of the mutational dynamics of DNA sequences, whether they be stretches of HVSI, Y-chromosome microsatellites, or RFLP sites on the Y-chromosome and in the mitochondrial genome. Integral to this is further research on whether the distribution of Y-chromosome and mtDNA lineages is and has been affected by selection, which will have profound effects on our understanding and interpretation of human history. It must also be remembered that the Y-chromosome and mtDNA provide only two observations of the underlying genealogies. Hence there can be considerable "noise", or lack of resolution, in the data, therefore not all questions can be effectively answered using only these two loci. For example, credible intervals for results obtained with certain population genetic analyses can be so wide when single loci data are used that distinguishing between competing hypotheses can be difficult. This problem was encountered with some of the LEA analyses performed in this thesis (Chapters 2 and 5). The inclusion of several unlinked microsatellites on the X chromosome in Chapter 5 highlighted the increased clarity obtained from using multiple loci, a point

which is often made by researchers (see for example Chikhi *et al.* 2001), but which is perhaps not fully appreciated until one has access to such data.

Furthermore the apparent conclusion that parts of the autosomal genome contain blocks of LD creates the potential that multiple autosomal sites can be used to provide high resolution haplotype systems (Stumpf and Goldstein, 2003). If these blocks of LD have persisted over long enough periods during which time polymorphic sites appear then they can be used as multiple, independent, observations of a genealogy, complementing information from the Y-chromosome and mtDNA (Stumpf and Goldstein 2003). However, Y-chromosome and mtDNA studies should not be thought of as obsolete as they uniquely allow the study of the paternal and maternal history of human populations, which, as Seielstad *et al.* (1998) showed, can have quite disparate histories.

References

Abbotts, J., Williams, R., Smith, G. D., (1999) Association of Medical, Physiological, Behavioural and Socio-Economic Factors with Elevated Mortality in Men of Irish Heritage in West Scotland. *J. Public Health Med.* **21**: 46-54.

Ackroyd, P. (2000). London, The Biography. Vintage, (London UK).

Al-Zahery, N., Semino, G., Benuzzi, G., Magri, C., Passarino, G., Torroni, A., and Santachiara-Benerecettia, A. S (2003) Y-chromosome and mtDNA Polymorphisms in Iraq, a Crossroad of the Early Human Dispersal and of Post-Neolithic Migrations. *Mol. Phylogenet Evol.* **28:** 458–472.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., Young, I. G. (1981). Sequence and Organization of the Human Mitochondrial Genome. *Nature.* **290:** 457-65.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., and Howell, N. (1999). Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA *Nat. Genet.* **23:** 147.

Anthony, D. W. (2000) Comment on Burmeister, S. (2000). Approaches to an Archaeological Proof of Migration. *Curr. Anthropol.* **41:** 539-567.

Antunez-de-Mayolo, G., Antunez-de-Mayolo, A., Antunez-de-Mayolo, P., Papiha, S. S., Hammer, M., Yunis, J. J., Yunis, E. J., Damodaran, C., Martinez de Pancorbo, M., Caeiro, J. L., Puzyrev, V. P., Herrera, R. J. (2002). Phylogenetics of Worldwide Human Populations as Determined by Polymorphic *Alu* Insertions. *Electrophoresis* **23**: 3346-3356.

Árnason, E. (2003) Genetic Heterogeneity of Icelanders. *Ann. Hum. Genet.* **67:** 5-16.

Arredi, B., Poloni, E. S., Paracchini, S., Zerjal, T., Fathallah, D. M., Makrelouf, M., Pascali, V. L., Novelletto, A., and Tyler-Smith, C. (2004). A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa. *Am. J. Hum. Genet.* **75:** 338–345.

Awadalla, P., Eyre-Walker, A., Maynard Smith, J. (1999) Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA. *Science* **286:** 2524-2525.

Awadalla, P. (2003) Does mtDNA Recombine? In: *Human Evolutionary Genetics: Origins, Peoples and Disease.* (Eds.) Jobling, M. A, Hurles, M. E, Tyler-Smith, C. Garland Science (New York, USA).

Bandelt, H.-J., Alves-Silva, J., Guimarães, P. E. M., Santos, M. S., Brehm, A., Pereira, L., Coppa, A., Larruga, J. M., Rengo, C., Scozzari, R., Torroni, A., Prata, M. J., Amorim, A., Prado, V. F., and Pena, S. D. J. 2001. Phylogeography

of the Human Mitochondrial Haplogroup L3e: A Snapshot of African Prehistory and Atlantic Slave Trade. *Ann. Hum. Genet.* **65:** 549-563.

Barrai, I. Rodriguez-Larralde, A., Mamolini E., Manni F., Scapoli, C. (2000). Elements of Surname Structure in Austria. *Ann. Hum. Biol.* **27:** 607-622.

Behar, D. M., Thomas, M. G., Skorecki, K., Hammer, M.F., Bulygina, E., Rosengarten, D., Jones, A.L., Held, K., Moses, V., Goldstein, D. B., Bradman, N., Weale, M. E. (2003). Multiple Origins of Ashkenazi Levites: Y Chromosome Evidence for Both Near Eastern and European Ancestries. *Am. J. Hum. Genet.* **73:** (768-79).

Behar, D. M., Hammer, M. F., Garrigan, D., Villems, R., Bonne-Tamir, B., Richards, M., Gurwitz, D., Rosengarten, D., Kaplan, M., Pergola, S. D., Quintana-Murci, L., and Skorecki, K. (2004)a. MtDNA Evidence for a Genetic Bottleneck in the Early History of the Ashkenazi Jewish population. *Eur. J. Hum. Genet.* **12:** 355–364.

Behar, D. M., Garrigan, D., Kaplan, M. E., Mobasher, Z., Rosengarten, D., Karafet, T. N., Quintana-Murci, L., Ostrer, H., Skorecki, K., and Hammer, M. F. (2004)b. Contrasting Patterns of Y Chromosome Variation in Ashkenazi Jewish and Host non-Jewish European Populations. *Hum. Genet.* **114:** 354-365.

Bender, B., with Caillaud, R. (1986). The Archaeology of Brittany, Normandy, and the Channel Islands: An Introduction and Guide. Faber and Faber Ltd. (London, UK).

Bertorelle, G. and Excoffier, L. (1998) Inferring Admixture Proportions from Molecular Data. *Mol. Biol. Evol.* **15:** 1298–1311.

Bosch, E., Calafell, F., Santos, F. R., Pérez-Lezaun, A., Comas, D., Benchemsi, N., Tyler-Smith, C., and Bertranpetit, J. (1999). Variation in Short Tandem Repeats Is Deeply Structured by Genetic Background on the Human Y Chromosome. *Am. J. Hum. Genet.* **65:** 1623–1638.

Bosch, E., Calafell, F., Comas, D., Oefner, P. J., Underhill, P. A., and Bertranpetit, J. (2001). High-Resolution Analysis of Human Y-Chromosome Variation Shows a Sharp Discontinuity and Limited Gene Flow Between Northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.* **68:** 1019–1029.

Bowie, N. and Jackson, G. W. L. (2003). Surnames in Scotland Over the Last 140 Years. General Register Office for Scotland. Occasional Paper 9.

Bowler, J. M., Johnston, H., Olley, J. M., Prescott, J. R., Roberts, R. G., Shawcross, W., and Spooner, N. A., (2003). New Ages for Human Occupation and Climatic Change at Lake Mungo, Australia. *Nature* **421:** 837-840.

Brehm, A., Pereira, L., Bandelt, H.-J., Prata, M. J., and Amorim, A. (2002). Mitochondrial Portrait of the Cabo Verde Archipelago: The Senegambian Outpost of Atlantic Slave Trade. *Ann. Hum. Genet.* **66**: 49-60.

Briggs, A. (1995) The Channel Islands, Occupation and Liberation 1940-1945. BT Batsford Ltd (London, UK).

Brown, W. M., George, M., Wilson, A. C. (1979). Rapid Evolution of Animal Mitochondrial DNA. *Proc. Natl. Acad. Sci. USA.* **76:** 1967-1971.

Brook, C. G. D. and Marshall, N. J (Eds.) (1996). Essential Endocrinology. Blackwell Scientific (Oxford, UK).

Budowle, B., Allard, M. W., Wilson, M. R., and Chakraborty, R. (2003). Forensics and Mitochondral DNA: Applications, Debates, and Foundations. *Annu. Rev. Genomics Hum. Genet.* **4:** 119–141.

Bryson, B. (1990) Mother Tongue. The English Language. Hamish Hamilton (London, UK).

Buijs, G. (1998). Black Jews in the Northern Province: A Study of Ethnic Identity in South Africa. *Ethnic and Racial Studies* **21:** 661-682.

Burmeister, S. (2000). Approaches to an Archaeological Proof of Migration. *Curr. Anthropol.* **41:** 539-567.

Burton, M. L., Moore, C. C., Whiting, J. W. M., Romney, A.K. (1996). Regions Based on Social Structure. *Curr. Anthropol.* **37:** 87-123.

Calafell, F. and Bertanpetit, J. (1994). Principal Component Analysis of Gene Frequencies and the Origin of the Basques. *Am. J. Phys. Anthropol.* **93:** 201-215.

Cann, R. L. (2002). Tangled Genetic Routes. *Nature* **416**: 32-33.

Cann, R. L., Stoneking, M. and Wilson, A. C. (1987). Mitochondrial DNA and Human Evolution. *Nature.* **325:** 31-36.

Capelli, C., Wilson, J. F., Richards, M., Stumpf, M. P. H., Gratrix, F., Oppenheimer, S., Underhill, P., Pascali, V. L., Ko, T.-M., Goldstein, D. B. (2001). A Predominantly Indigenous Paternal Heritage for the Austronesian-Speaking Peoples of Insular Southeast Asia and Oceania. *Am. J. Hum. Genet.* **68:** 432–443.

Capelli, C., Redhead, N., Abernethy, J. K., Gratrix, F., Wilson, J. F., Moen, T., Hervig, T., Richards, M., Stumpf, M. P. H., Underhill, P. A., Bradshaw, P., Shaha, A., Thomas, M. G., Bradman, N., and Goldstein, D. B. (2003) A Y Chromosome Census of the British Isles. *Curr. Biol.* **13:** 979-984.

Carvajal-Carmona, L. G., Soto, I. D., Pineda, N., Ortíz-Barrientos, D., Duque, C., Ospina-Duque, J., McCarthy, M., Montoya, P., Alvarez, V. M., Bedoya, G., and Ruiz-Linares, A. (2000). Strong Amerind/White Sex Bias and a Possible Sephardic Contribution Among the Founders of a Population in Northwest Colombia. *Am. J. Hum. Genet.* **67:** 1287–1295.

Casallotti, R., Simoni, L., Belledi, M., Barbujani, G. (1999). Y-chromosome Polymorphisms and the Origins of the European gene pool. *Proc. R. Soc. Lond. B.* **266:** 1959-1965.

Casanova, M., Leroy, P., Boucekkine, C, Weissenbach, J., Bishop, C., Fellous, M., Purrello, M., Fiori, G., Siniscalco, M. (1985). A Human Y-Linked DNA Polymorphism and its Potential for Estimating Genetic and Evolutionary Distance. *Science* **230:** 1403-1406.

Cavalli-Sforza, L. L., Menozzi, P., Piazza, A. (1994). The History and Geography of Human Genes. Princeton University Press. (Princeton, NJ).

Cavalli-Sforza, L. L. and Feldman, M. W. (2003). The Application of Molecular Genetic Approaches to the Study of Human Evolution. *Nat. Genet. Supp.* **33:** 266-275.

Cavalli-Sforza, L. L., Moroni, A., and Zei, G. (2004). Consanguinity, Inbreeding and Genetic Drift in Italy. Princeton Univeristy Press (Princeton, NJ).

Cerda-Flores, R. M., Barton, S. A., Marty-González, L. F., Rivas, F., Chakraborty, R., (1999). Estimation of Nonpaternity in the Mexican Population of Nueva leon. *Am. J. Phys. Anthropol.* 109: 281-93.

Chen, Y.-S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S., Wallace, D. C. (1995). Analysis of mtDNA Variation in African Populations Reveals the Most Ancient of all Human Continent-Specific Haplogroups. *Am. J. Hum. Genet.* **57:** 133-149.

Chen, Y.-S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K., and Wallace, D. C. (2000). mtDNA Variation in the South African Kung and Khwe—and Their Genetic Relationships to Other African Populations. *Am. J. Hum. Genet.* **66:** 1362–1383.

Chikhi, L., Bruford, M. W., and Beaumont, M. A. (2001). Estimation of Admixture Proportions: A Likelihood-Based Approach Using Markov Chain Monte Carlo. *Genetics* **158:** 1347–1362.

Chikhi, L., Nichols, R. A., Barbujani, G., and Beaumont, M. A. (2002). Y Genetic Data Support the Neolithic Demic Diffusion Model. *Proc. Natl. Acad. Sci. USA.* **99:** 11008–11013.

Clark, J. D., Beyene, Y., Wolde Gabriel, G., Hart, W. K., Renne, P. R., Gilbert, H., Defleur, A., Suwa, G., Katoh, S., Ludwig, K. R., Boisserie, J. R., Asfaw, B., White, T. D. (2003). Stratigraphic, Chronological and Behavioural Contexts of

Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423:** 747-52.

Cohen, J. E. (2004). Human Population: The Next Half Century. *Science* **302:** 1172-1175.

Comas, D., Calafell, F., Benchemsi, N., Helal, A., Lefranc, G., Stoneking, M., Batzer, M. A., Bertranpetit, J., and Sajantila, A. (2000). *Alu* Insertion Polymorphisms in NW Africa and the Iberian Peninsula: Evidence for a Strong Genetic Boundary Through the Gibraltar Straits. *Hum. Genet.* **107:** 312–319.

Cotton, J. and White, B. (1998). Ancient Bodies: The Lives of Prehistoric Londoners. In *London Bodies. The Changing Shape of Londoners from Prehistoric Times to the Present Day* (Compiled by Werner, A.) Museum of London (London, UK).

Cruciani, F., Santolamazza, P., Shen, P., 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. *Am. J. Hum. Genet.* **70:** 1197–1214.

Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Béraud-Colomb, E., Zaharova, B., Lavinha, J., Vona, G., Aman, R., Calí, F., Akar, N., Richards, M., Torroni, A., Novelletto, A., and Scozzari, R. (2004). Phylogeographic Analysis of Haplogroup E3b (E-M215) Y Chromosomes Reveals Multiple Migratory Events Within and Out Of Africa. *Am. J. Hum. Genet.* **74:** 1014-1022.

Davies, N. (1999). The Isles A History. MacMillan (London, UK).

Degioanni, A. and Darlu, P. (2001). A Bayesian Approach to Infer Geographical Origins of Migrants Through Surnames. *Ann. Hum. Biol.* **28:** 537-545.

de Knijff, P. (2000). Messages Through Bottlenecks: On the Combined Use of Slow and Fast Evolving Polymorphic Markers on the Human Y Chromosome. *Am. J. Hum. Genet.* **67:** 1055-1061.

Derenko, M. V., Grzybowski, T., Malyarchuk, B. A., Dambueva, I. K., Denisova, G. A., Czarny, J., Dorzhu, C. M., Kakpakov, V. T., Miścicka-Śliwka, D., Woźniak, M., Zakharov, I. A. (2003). Diversity of Mitochondrial DNA Lineages in South Siberia. *Ann. Hum. Genet.* **67:** 391-411.

Di Benedetto, G., Ergüven, A., Stenico, M., Castrí, L., Bertorelle, G., Togan, I., and Barbujani, G. (2001). DNA Diversity and Population Admixture in Anatolia. *Am. J. Phys. Anthropol.* **115:** 144–156

Dieringer, D. and Schlötterer, C. (2003). Two Distinct Modes of Microsatellite Mutation Processes: Evidence From the Complete Genomic Sequences of Nine Species. *Genome Res.* **13:** 2242-2251.

Dobson, J. and McLaughlan, G. (2001). International Migration to and From the United Kingdom, 1975-1999: Consistency, Change and Implications for the Labour Market. *Populations Trends 106.* Office for National Statistics.

Dorward, D. (2000). Scottish Surnames. HarperCollins Publishers (Glasgow, UK).

Dubut, V., Chollet, L., Murail, P., Cartault, F., Béraud-Colomb, E., Serre, M., and Mogentale-Profizi, N. (2004). mtDNA Polymorphisms in Five French Groups: Importance of Regional Sampling. *Eur. J. Human Genet.* **12:** 293–300.

Editorial (2001). Genes, Drugs and Race. *Nat. Genet.* **29:** 239-240.

Elson, J. L., Andrews, R. M., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., and Howell, N. (2001). Analysis of European mtDNAs for Recombination. *Am. J. Hum. Genet.* **68:** 145-153.

Encyclopaedia Judaica. (1972). Keter Publishing (Jerusalem, Israel).

Eyre-Walker, A., Smith, N. H., Smith, J. M. (1999). How Clonal are Human Mitochondria? *Proc. R. Soc. Lond. B.* **266:** 477-483.

Eyre-Walker, A., and Awadalla, P. (2001). Does Human mtDNA Recombine? *J. Mol. Evol.* **53:** 430-435.

Finnilä, S., Lehtonen, M. S., Majamaa, K. (2001). Phylogenetic Network for European mtDNA. *Am. J. Hum. Genet.* **68:** 1475-1484.

Fix, A. G. (1996). Gene Frequency Clines in Europe: Demic Diffusion or Natural Selection? *J. Roy. Anthropol. Inst.* **2:** 625-643.

Fliss, M. S., Usadel, H., Caballero, O. L., Wu, L., Buta, M. R., Eleff, S. M., Jen, J., and Sidransky, D. (2000). Facile Detection of Mitochondrial DNA Mutations in Tumors and Bodily Fluids. *Science* **287:** 2017-2019.

Foote, S., Vollrath, D., Hilton, A., Page, D. C. (1992). The Human Y Chromosome: Overlapping DNA Clones Spanning the Euchromatic Region. *Science* **258:** 60-66.

Forster, P. (1995). Einwanderungsgeschichte Noeddeutschlands (Immigration History of North Germany). In: *North-Western European Language Evolution.* (Eds.) Falting, V. F., Walker, A. G. H., and Wilts, O. Univeristy of Odense, (Odense, Denmark).

Forster, P., Harding, R., Torroni, A., and Bandelt, H.-J. (1996). Origin and Evolution of Native American mtDNA variation: a Reappraisal. *Am. J. Hum. Genet.* **59:** 935-945.

Forster, P., Röhl, A., Lünnemann, P., Brinkmann, C., Zerjal, T., Tyler-Smith, C., Brinkmann, B. (2000). A Short Tandem Repeat-Based Phylogeny for the Human Y Chromosome. *Am. J. Hum. Genet.* **67:** 182-196.

Forster, P. (2004). Ice Ages and the Mitochondrial DNA Chronology of Human Dispersals: A Review. *Phil. Trans. R. Soc. Lond. B* **359:** 255–264.

Francalacci, P., Morelli, L., Underhill, P. A., Lillie, A. S., Passarino, G., Useli, A., Madeddu, R., Paoli, G., Tofanelli, S., Calò, C. M., Ghiani, M. E., Varesi, L., Memmi, M., Vona, G., Lin, A. A., Oefner, P., Cavalli-Sforza, L. L. (2003). Peopling of Three Mediterranean Islands (Corsia, Sardinia, and Sicily) Inferred by Y-Chromosome Biallelic Variability. *Am. J. Phys. Anthropol.* **121:** 270-279.

Gamble, C., Davies, W., Pettitt, P., and Richards, M. (2004). Climate Change and Evolving Human Diversity in Europe During the Last Glacial. *Phil. Trans. R. Soc. Lond. B.* **359**: 243–254.

Gerber, A. S., Loggins, R., Kumar, S., and Dowling, T. E. (2001). Does Nonneutral Evolution Shape Observed Patterns of DNA Variation in Animal Mitochondrial Genomes? *Annu. Rev. Genet.* **35:** 539-66.

Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L., Feldman, M. W. (1995)a. Genetic Absolute Dating Based on Microsatellites and the Origin of Modern Humans. *Proc. Natl. Acad. Sci., USA* **92:** 6723-6727.

Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L., and Feldman, M. W. (1995)b. An Evaluation of Genetic Distances for Use With Microsatellite Loci. *Genetics* **139:** 463-471.

Goldstein, D. B. and Chikhi, L. (2002). Human Migrations and Population Structure: What We Know and Why it Matters. *Annu. Rev. Genomics Hum. Genet.* **3:** 129–52.

Gonçalves, R., Rosa, A., Freitas, A., Fernandes, A., Kivisild, T., Villems, R., Brehm, A. (2003). Y-Chromosome Lineages in Cabo Verde Islands Witness the Diverse Geographic Origin of its First Male Settlers. *Hum. Genet.* **113:** 467-472.

González, A. M., Brehm, A., Pérez, J. A., Maca-Meyer, N., Flores, C., and Cabrera, V. M. (2003). Mitochondrial DNA Affinities at the Atlantic Fringe of Europe. *Am. J. Phys. Anthropol.* **120:** 391–404.

Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A., Excoffier, L. (1995). Evolutionary Correlation between Control Region Sequence and Restriction Polymorphisms in the Mitochondrial Genome of a Large Senegalese Mandenka Sample. *Mol. Biol. Evol.* **12:** 334-345.

Graham-Campbell, J. and Batey, C. (1998). Vikings in Scotland. Edinburgh University Press (Edinburgh, UK).

Gresham, D., Morar, B., Underhill, P. A., Passarino, G., Lin, A. A., Wise, C., Angelicheva, D., Calafell, F., Oefner, P. J., Shen, P., Tournev, I., de Pablo, R., Kučinskas, V., Perez-Lezaun, A., Marushiakova, E., Popov, V., and Kalaydjieva, L. (2001). Origins and Divergence of the Roma (Gypsies) *Am. J. Hum. Genet.* **69:** 1314–1331.

Gusmão, L., Sánchez-Diz, P., Alves, C., Beleza, S., Lopes, A., Carracedo, A., and Amorim, A. (2003). Grouping of Y-STR Haplotypes Discloses European Geographic Clines. *Forensic Sci. Int.* **134:** 172–179.

Hagelberg, E., Goldman, N., Lio, P., Whelan, S., Schiefenhovel, W., Clegg, J.B., Bowden, D.K. (1999). Proc. R. Soc. Lond. B Biol. Sci. (1999) 7: 485-92

Hagelberg, E., Goldman, N., Lio, P., Whelan, S., Schiefenhovel, W., Clegg, J.B., Bowden, D.K. (2000). Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Erratum in: Proc. R. Soc. Lond. B. Biol. Sci. 7:1595-6.

Hall, J. and Conheeney, J. (1998). Roman Bodies: The Stresses and Strains of Life in Roman London. In: *London Bodies. The Changing Shape of Londoners from Prehistoric Times to the Present Day.* (Compiled by Werner, A.) Museum of London (London, UK).

Hammer, M. F. (1994). A Recent Insertion of an *Alu* Element on the Y Chromosome Is a Useful Marker for Human Population Studies. *Mol. Biol. Evol.* **11:** 749-761.

Hammer, M. F. (1995) A Recent Common Ancestry for Human Y Chromosomes. *Nature* **378:** 376-378.

Hammer, M. F., Karafet, T., Rasanayagam, A., Wood, E. T., Altheide, T. K., Jenkins, T., Griffiths, R. C., Templeton, A. R., Zegura, S. L. (1998). Out of Africa and Back Again: Neststed Cladistic Analysis of Human Y Chromosome Variation. *Mol. Biol. Evol.* **15:** 427-441.

Hammer, M. F., Redd, A. J., Wood, E. T., Bonner, M. R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, S., Oppenheim, A., Jobling, M. A., Jenkins, T., Ostrer, H., Bonné-Tamir, B. (2000). Jewish and Middle Eastern non-Jewish Populations Share a Common Pool of Y-chromosome Biallelic Haplotypes. *Proc. Natl. Acad. Sci. USA* **97:** 6769–6774.

Hammer, M. F. and Zegura, S. L. (2002). The Human Y Chromosome Haplogroup Tree: Nomenclature and Phylogeography of Its Major Divisions. *Annu. Rev. Anthropol.* **31:** 303-321.

Harpending, H. C., Batzer, M. A., Gurven, M., Jorde, L. B., Rogers, A. R., and Sherry, S. T. (1998). Genetic Traces of Ancient Demography. *Proc. Natl. Acad. Sci. USA* **95:** 1961-1967.

Hawkes, J. (1937). The Archaeology of the Channel Islands. Volume 1. Société Jersiaise. (Jersey, UK).

Helgason, A., Sigurðardóttir, S., Gulcher, J. R., Ward, R., and Stefánsson, K. (2000). mtDNA and the Origin of the Icelanders: Deciphering Signals of Recent Population History. *Am. J. Hum. Genet.* **66:** 999–1016.

Helgason, A., Hickey, E., Goodacre, S., Bosnes, V., Stefánsson, K., Ward, R., and Sykes, B. (2001). MtDNA and the Islands of the North Atlantic: Estimating the Proportions of Norse and Gaelic Ancestry. *Am. J. Hum. Genet.* **68:** 723–737.

Helgason, A., Hrafnkelsson, B., Gulcher, J. R., Ward, R., Stefánsson, K. (2003). A Populationwide Coalescent Analysis of Icelandic Matrilineal and Patrilineal Genealogies: Evidence for a Faster Evolutionary Rate of mtDNA Lineages than Y Chromosomes. *Am. J. Hum. Genet.* **72:** 1370-1388.

Heyer, E., Zietkiewicz, E., Rochowski, A., Yotova, V., Puymirat, J., and Labuda, D. (2001). Phylogenetic and Familial Estimates of Mitochondrial Substitution Rates: Study of Control Region Mutations in Deep-Rooting Pedigrees. *Am. J. Hum. Genet.* **69:** 1113–1126.

Hill, D. (1981). An Atlas of Anglo Saxon England. Blackwell (Oxford, UK).

Hill, E.W., Jobling, M. A., and Bradley, D. G. (2000). Y-Chromosome Variation and Irish Origins. *Nature* **404:** 351-352.

Holmes, C. (1997). Cosmopolitan London. In: *London the Promised Land?: The Migrant Experience in a Capital City.* (Ed.) Kershen, A. J. Brookfield, Vt (Hants, UK).

Hughes, A. J. B., Lowe, R. F., Gadd, K. G., Ellis, B. P. B. (1978). The Sero-Anthropology of the Rhodesian Lemba. *Hum. Hered.* **28:** 261-269.

Hurles, M. E., Irven, C., Nicholson, J., Taylor, P. G., Santos, F. R., Loughlin, J., Jobling, M. A., and Sykes, B. (1998). European Y-Chromosomal Lineages in Polynesians: A Contrast to the Population Structure Revealed by mtDNA. *Am. J. Hum. Genet.* **63:** 1793–1806.

Hurles, M. E., Veitia, R., Arroyo, E., Armenteros, M., Bertranpetit, J., Pérez-Lezaun, A., Bosch, E., Shlumukova, M., Cambon-Thomsen, A., McElreavey, K., López de Munain, A., Röhl, A., Wilson, I. J., Singh, L., Pandya, A., Santos, F. R., Tyler-Smith, C., and Jobling, M. A., (1999). Recent Male-Mediated Gene Flow over a Linguistic Barrier in Iberia, Suggested by Analysis of a Y-Chromosomal DNA Polymorphism. *Am. J. Hum. Genet.* **65:** 1437–1448.

Hurles, M. E. and Jobling, M. A. (2001). Haploid Chromosomes in Molecular Ecology: Lessons from the Human Y. *Mol. Ecol.* **10:** 1599-1613.

Ingman. M., Kaessmann. H., Pääbo, S., Gyllensten, U. (2000) Mitochondrial Genome Variation and the Origin of Modern Humans. *Nature* **408:** 708-712.
The International HapMap Consortium (2003). The International HapMap Project. *Nature* **426:** 789-796.

The International Human Genome Sequencing Consortium (2001). Initial Sequencing and Analysis of the Human Genome. *Nature* **409:** 860-921.

The International SNP Map Working Group (2001). A Map of Human Genome Sequence Variation Containing 1.42 Million Single Nucleotide Polymorphisms. *Nature* **409:** 928-933.

Inwood, S. (1998). A History of London. MacMillan (London, UK).

Jobling, M. A. and Tyler-Smith, C. (1995). Fathers and Sons: The Y Chromosome and Human Evolution. *Trends Genet.* **11:** 449-456.

Jobling, M. A., Bouzekri, N., and Taylor, P. G. (1998). Hypervariable Digital DNA Codes for Human Paternal Lineages: MVR-PCR at the Y-Specific Minisatellite, MSY1 (DYF155S1). *Hum. Mol. Genet.* **7:** 643-653.

Jobling, M. A. and Tyler-Smith, C. (2000). New Uses for New Haplotypes the Human Y Chromosome, Disease and Selection. *Trends Genet.* **16:** 356-362.

Jobling, M. A. (2001). In the Name of the Father: Surnames and Genetics. *Trends Genet.* **17:** 353-357.

Jobling, M. A., Hurles, M. E., Tyler-Smith, C. (2003). Human Evolutionary Genetics: Origins, Peoples and Disease. Garland Science (New York, USA).

Jobling, M. A. and Tyler-Smith, C. (2003). The Human Y Chromosome: An Evolutionary Marker Comes of Age. *Nat. Rev. Genet.* **4:** 598-612.

Jones, G. (1984). A History of the Vikings, 2nd edn. Oxford University Press. (Oxford, UK).

Johnston, J. (2003). Case Study: The Lemba. *Developing World Bioethics.* **3:** 109-111.

Jorde, L. B. and Bamshad, M. (2000). Questioning Evidence for Recombination in Human Mitochondrial DNA. *Science* **288:** 1931a-1932a.

Kayser. M., Rower. L., Hedman. M., Henke, L., Henker, J., Brauer, S., Krüger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., Stoneking, M., Sajantila, A. (2000). Characteristics and Frequency of Germline Mutatations in Microsatellite Loci from the Human Y Chromosome, as Revealed by Direct Observation in Father/Son Pairs. *Am. J. Hum. Genet.* **66:** 1580-1588.

Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A. C., Mohyuddin, A., Qasim Mehdi, S., Rosser, Z., Stoneking, M., Jobling, M. A., Sajantila, A., and Tyler-Smith, C. (2004). A Comprehensive Survey of Human Y-Chromosomal Microsatellites. *Am. J. Hum. Genet.* **74:** 1183-1197.

Kershen, A. J. Introduction. In: *London the Promised Land?: The Migrant Experience in a Capital City.* (Ed.). Kershen, A. J. Brookfield, Vt. (Hants , UK).

Kimmel, M. and Chakraborty, R. (1996). Measure of Variation at DNA Repeat Loci Under a General Stepwise Mutation Rate. *Theor. Pop. Biol.* **50:** 345-367.
Kivisild, T. and Villems, R. (2000). Questioning Evidence for Recombination in Human Mitochondrial DNA. *Science* **288:** 1931a.

Klein, R. (1999). The Human Career: Human Biological and Cultural Origins, 2nd edition, University of Chicago Press (Chicago, USA).

Krausz, C., Quintana-Murci, L., Rajpert-De Meyts, E., Jørgensen, N., Jobling, M. A., Rosser, Z. H., Skakkebaek, N. E., McElreavey, K. (2001). Identification of a Y Chromosome Haplogroup Associated with Reduced Sperm Counts. *Hum. Mol. Genet.* **10:** 1873-1877.

Lahr, M. M. and Foley, R. A (1998). Towards a Phylogeography of Modern Human Origins: Geography, Demography and Diversity in Recent Human Evolution. *Am. J. Phys. Anthropol. Suppl.* **27:** 137-76.

Lehmann, R. P. M. (1991). Ogham: The Ancient Script of the Celts. In: *The Origins of Writing.* (Ed.) Senner, W. M. University of Nebraska Press. (Nebraska, USA).

Lempriere, R. (1976). Customs, Ceremonies and Traditions of the Channel Islands. Hale (London, UK).

Lasker, G.W. (1985). Surnames and Genetic Structure. Cambridge University Press (Cambridge, UK).

Lucotte, G. and Mercier, G. (2003). Y-Chromosome DNA Haplotypes in Jews: Comparisons with Lebanese and Palestinians. *Genet. Testing* **7:** 67-71.

Maca-Meyer, M., González, A.-M., Larruga, J. M., Flores, C., and Cabrera, V. M. (2001). Major Genomic Mitochondrial Lineages Delineate Early Human Expansions. *BMC Genetics* **2:** 13-20.

Maca-Meyer, N., Sánchez-Velasco, P., Flores, C., Larruga, J.-M., González, A.-M., Oterino, A., Leyva-Cobián F. (2003). Y Chromosome and Mitchondrial DNA Characterization of Pasiegos, a Human Isolate from Cantabria (Spain). *Ann. Hum. Genet.* **67:** 329-339.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B., and Torroni, A. (1999). The Emerging

Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs. *Am. J. Hum. Genet.* **64:** 232-249.

Malaspina, P., Cruciani, F., Santolamazza, P., Torroni, A., Pangrazio, A., Akar, N., Bakalli, V., Brdicka, R., Jaruzelska, J., Kozlov, A., Malyarchuk, B., Mehdi, S. Q., Michalodimitrakis, E., Varesi, L., Memmi, M. M., Vona, G., Villems, R., Parik, J., Romano, V., Stefan, M., Stenico, M., Terrenato, L., Novelletto, A., and Scozzari, R. (2000). Patterns of Male-Specifc Inter-Population Divergence in Europe, West Asia and North Africa. *Ann. Hum. Genet.* **64:** 395-412.

McBrearty, S. and Brooks, A. S. (2000). The Revolution That Wasn't: A New Interpretation of the Origin of Modern Human Behavior. *J. Hum. Evol.* **39:** 453–563.

McAuley, I. (1993). Guide to Ethnic London. IMMEL Publishing (London, UK).

McLeod, H. L. (2001). Pharmacogenetics: More Than Skin Deep. *Nat. Genet.* **29:** 247-248.

Mishmar, D., Ruiz-Pesinia, E., Golikb, P., Macaulay, V., Clarke, A. G., Hosseini, S., Brandon, M., Easley, K., Cheng, E., Brown, M. D., Sukernik, R. I., Olckers, A., and Wallace, D. C. (2003). Natural Selection Shaped Regional MtDNA Variation in Humans. *Proc. Natl. Acad. Sci. USA* **100:** 171–176.

Morrison, A. (1986). The Chiefs of Clan MacLeod. The Associated Clan MacLeod Societies. (Edinburgh, UK).

Mullis, K. B., Faloona, F., Scharf, S. J., Saiki, R. K., Horn, G. T., and Erlick, H. A. (1986). Specific Enzymatic Amplification of DNA *in vitro*: The Polymerase Chain Reaction. *Cold Spring Harbor Symp. Quant. Biol.* **51:** 263-273.

Myhill, H. (1964). Introducing the Channel Islands. Faber and Faber (London, UK).

Nebel, A., Filon, D., Weiss, D. A., Weale, M., Faerman, M., Oppenheim, A., and Thomas, M. G. (2000). High-Resolution Y Chromosome Haplotypes of Israeli and Palestinian Arabs Reveal Geographic Substructure and Substantial Overlap with Haplotypes of Jews. *Hum. Genet.* **107:** 630–641.

Nebel, A., Filon, D., Brinkmann, B., Majumder, P. P., Faerman, M., Oppenheim, A. (2001). The Y Chromosome Pool of Jews as Part of the Genetic Landscape of the Middle East. *Am. J. Hum. Genet.* **69:** 1095–1112.

Nicholson, G. J., Tomiuk, J., Czarnetzki, A., Bachmann, L., Pusch, C. M. (2002). Detection of Bone Glue Treatment as a Major Source of Contamination in Ancient DNA Analyses. *Am. J. Phys. Anthropol.* **118:** 117-120.

Nicolle, E. T. (1935). A chronology of Events and Occurrences with Regard to the Island of Jersey. Augmented and revised by Ralph Mollet. Chroniques de Jersey (Jersey, UK).

Nordborg, M. (2000). Coalescent Theory. In: *Handbook of Statistical Genetics* (Eds.) Balding, D., Bishop, M., and Cannings, C. Wiley (Chichester, UK).

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., and Stoneking, M. (2001). Human MtDNA and Y-Chromosome Variation is Correlated with Matrilocal Versus Patrilocal Residence. *Nat. Genet.* **29:** 20-21.

The Concise Oxford English Dictionary (1995). 9th Edition. Oxford University Press. (Oxford, UK).

Parfitt, T. (1997). Journey to the Vanished Land. The Search for a Lost Tribe of Israel. Phoenix (London, UK).

Parfitt, T. (2003). Constructing Black Jews:Genetic Tests and the Lemba – the 'Black Jews of South Africa'. *Developing World Bioethics* **3:** 112-118.

Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M., and Santachiara-Benerecetti, A. S. (1998). Different Genetic Components in the Ethiopian Population, Identified by MtDNA and Y-Chromosome Polymorphisms. *Am. J. Hum. Genet.* **62:** 420–434.

Passarino, G., Cavalleri, G. L., Lin, A. A., Cavalli-Sforza, L. L., Børresen-Dale, A.-L., Underhill, P. A. (2002). Different Genetic Components in the Norwegian Population Revealed by the Analysis of MtDNA and Y Chromosome Polymorphisms. *Eur. J. Hum. Genet.* **10**: 521-529.

Peirera, L., Gusmão, L., Alves, C., Amorim, A., and Prata, M. J. (2002). Bantu and European Y-Lineages in Sub-Saharan Africa. *Ann. Hum. Genet.* **66:** 369-378.

Penny, D., Steel, M., Waddell, P. J., Hendy, M. D. (1995). Improved Analyses of Human mtDNA Sequences Support a Recent African Origin for *Homo sapiens*. *Mol. Biol. Evol.* **12:** 863-882.

Pérez-Lezaun, A., Calafell, F., Comas, D., Mateu, E., Bosch, E., Martínez-Arias, R., Clarimón, J., Fiori, G., Luiselli, D., Facchini, F., Pettener, D., and Bertranpetit, J. (1999). Sex-Specific Migration Patterns in Central Asian Populations, Revealed by Analysis of Y-Chromosome Short Tandem Repeats and mtDNA. *Am. J. Hum. Genet.* **65:** 208–219.

Pew Hispanic Centre Report (2004). Assimilation and Language. Survey Brief 2004. Pew Hispanic Centre. (Washington DC, USA).

Pritchard, J. K. and Przeworski, M. (2001). Linkage Disequilibrium in Humans: Models and Data. *Am. J. Hum. Genet.* **69:** 1–14.

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. *Genetics* **23:** 437-441.

Quintana-Murci, L., Weale, M., Thomas, M. G., Erdei, E., Bradman, N., Shanks, J. H., Krausz, C., McElreavey, K. (2003). Y Chromosome Haplotypes and Testicular Cancer in the English Population. *J. Med. Genet.* **40:** 1-5.

Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A. S., Coppa, A., Ayub, Q., Mohyuddin, A., Tyler-Smith, C., Mehdi, Q., Torroni, A., and McElreavey, K. (2004). Where West Meets East: The Complex mtDNA Landscape of the Southwest and Central Asian Corridor. *Am. J. Hum. Genet.* **74:** 827–845.

Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M., and Bandelt, H.-J. (1998) Mitochondrial DNA Analysis of Northwest African Populations Reveals Genetic Exchanges with European, Near-Eastern, and sub-Saharan Populations. *Ann. Hum. Genet.* **62:** 531-550.

Reaney, P. H. (1927). A Grammar and Dialect of Penrith (Cumberland): Descriptive and Historical, with Specimens and Glossary. The University Press. (Manchester, UK).

Reaney, P.H. (1997). A Dictionary of English Surnames. Third Edition (Revised by Wilson, R.M.). Oxford University Press (Oxford, UK).

Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S., and Altshuler, D. (2003). Human Genome Sequence Variation and the Influence of Gene History, Mutation and Recombination. *Nat. Genet.* **32:** 135-142.

Renfrew, C. (2000). Archaeogenetics: Towards a Population Prehistory of Europe. In: *Archaeogeneitcs: DNA and the population history of Europe.* (Eds.) Renfrew, C. and Boyle, B. McDonald Institute for Archaeological Research (Cambridge, UK).

Renwick, A., Davison, L., Spratt, H., King, J. P., and Kimmel, M. (2001). DNA Dinucleotide Evolution in Humans: Fitting Theory to Facts. *Genetics* **159:** 737–747.

Richards, J. D. (1991). Viking Age England. B. T. Batsford Ltd./English Heritage (London, UK).

Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J., Sykes, B. (1996). Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool. *Am. J. Hum. Genet.* **59:** 185-203.

Richards, M. and Macaulay, V. (2000). Genetic Data and the Colonisation of Europe: Genealogies and Founders. In: *Archaeogeneitcs: DNA and the Population History of Europe.* (Eds.) Renfrew, C. and Boyle, B. McDonald Institute for Archaeological Research (Cambridge, UK).

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

Richards, M. and Macauly, V. (2001). The Mitochondrial Gene Tree Comes of Age. *Am. J. Hum. Genet.* **68:** 1315-1320.

Richards, M., Macaulay, V., Torroni, A., and Bandelt, H.-J. (2002). In Search of Geographical Patterns in European Mitochondrial DNA. *Am. J. Hum. Genet.* **71:** 1168-1174.

Richards, M., Rengo, C., Cruciani, F., Gratrix, F., Wilson, J. F., Scozzari, R., Macaulay, V., and Torroni, A. (2003). Extensive Female-Mediated Gene Flow from Sub-Saharan Africa into Near Eastern Arab Populations. *Am. J. Hum. Genet.* **72:** 1058–1064.

Roewer, L., Kayser, M., de Knijff, P., Anslinger, K., Betz, A., Caglià, A., Corach, D., Füredi, S., Henke, L., Hidding, M., Kärgel, H. J., Lessig, R., Nagy, M., Pascali, V. L., Parson, W., Rolf, B., Schmitt, C., Szibor, R., Teifel-Greding, J., Krawczak, M. (2000). A New Method for the Evaluation of Matches in Non-Recombining Genomes: Application to Y-Chromosomal Short Tandem Repeat (STR) Haplotypes in European Males. *Forensic Sci. Int.* **114:** 31-43.

Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., Caglia, A., Carracedo, A., Corach, D., Dekairelle, A. F., Dobosz, T., Dupuy, B. M., Furedi, S., Gehrig, C., Gusmão, L., Henke, J., Henke, L., Hidding, M., Hohoff, C., Hoste, B., Jobling, M. A., Kargel, H. J., de Knijff, P., Lessig, R., Liebeherr, E., Lorente, M., Martinez-Jarreta, B., Nievas, P., Nowak, M., Parson, W., Pascali, V. L., Penacino, G., Ploski, R., Rolf, B., Sala, A., Schmidt, U., Schmitt, C., Schneider, P. M., Szibor, R., Teifel-Greding, J., Kayser, M. (2001). Online Reference Database of European Y-Chromosomal Short Tandem Repeat (STR) Haplotypes. *Forensic Sci. Int.* **118:** 106-13.

Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barać, L., Peric¡ić, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L. A., Battaglia, V., Achilli, A., Al-Zahery, N., Parik, J., King, R., Cinniogȼlu, C., Khusnutdinova, E., Rudan, P., Balanovska, E., Scheffrahn, W., Simonescu, M., Brehm, A., Gonçalves, R., Rosa, A., Moisan, J.-P., Chaventre, A., Ferak, V., Füredi, S., Oefner, P. J., Shen, P., Beckman, L.,

Mikerezi, I., Terzić, R., Primorac, D., Cambon-Thomsen, A., Krumina, A., Torroni, A., Underhill, P. A., Santachiara-Benerecetti, A. S., Villems, R., and Semino, O. (2004). Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe *Am. J. Hum. Genet.* **75:** 128–137.

Rosser, R. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D. G., Brede, G., Cooper, G., Côrte-Real, H. B. S. M., de Knijff, P., Decorte, R., Dubrova, Y. E., Evgrafov, O., Gilissen, A., Glisic, S., Gölge, M., Hill, E. W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S. A., Krumina, A., Kučinskas, V., Lavinha, J., Livshits, L. A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T. A., Mikelsaar, A.-V., Mitchell, R. J., Nafa, K., Nicholson, J., Nørby, S., Pandya, A., Parik, J., Patsalis, P. C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M. J., Previderé, C., Roewer, L., Rootsi, S., Rubinsztein, D. C., Saillard, J., Santos, F. R., Stefanescu, G., Sykes, B. C., Tolun, A., Villems, R., Tyler-Smith, C., and Jobling, M. A. (2000). Y-Chromosomal Diversity in Europe is Clinal and Influenced Primarily by Geography, Rather Than by Language. *Am. J. Hum. Genet.* **67:** 1526-43.

Rowley, T. (1997). Norman England. B. T. Batsford/English Heritage (London, UK).

Salas, A., Richards, A., De la Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V., and Carracedo, A. (2002). The Making of the African mtDNA Landscape. *Am. J. Hum. Genet.* **71:** 1082–1111.

Sanders, S. (2000). Invisible Races. *Transition* **10:** 76-97

Schlötterer, C. (2000). Evolutionary Dynamics of Microsatellite DNA. *Chromosoma* **109:** 365–371.

Schlötterer, C. (2001). Genealogical Inference of Closely Related Species Based on Microsatellites. *Genet. Res., Camb.* **78:** 209-212.

Schneider, S., Kueffer, J.-M., Roessli, D., and Excoffier, L. (1997). Arlequin ver 1.1. A Population Genetic Data Analysis Programme. Genetics and Biometry Laboratory, University of Geneva. (Geneva, Switzerland).

Schneider, S., Roessli, D., and Excoffier, L. (2000). Arlequin ver. 2000: A Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva. (Geneva, Switzerland).

Schurr, T. G. and Sherry, S. T. (2004). Mitochondrial DNA and Y Chromosome Diversity and the Peopling of the Americas: Evolutionary and Demographic Evidence. *Am. J. Hum. Biol.* **16:** 420–439.

Segal, R. (2001). Islam's Black Slaves. The History of Africa's other Black Diaspora. Atlantic Books (UK).

Seielstad, M. T., Minch, E., and Cavalli-Sforza, L. L. (1998). Genetic Evidence for a Higher Female Migration Rate in Humans. *Nat. Genet.* **20:** 278-280.

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., Underhill, P. A. (2000). The Genetic Legacy of Paleolithic *Homo sapiens sapiens* in Extant Europeans: A Y Chromosome Perspective. *Science* **290:** 1155-1159.

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P. J., Zhivotovsky, L. A., King, R., Torroni, A., Cavalli-Sforza, L L., Underhill, P. A., and Santachiara-Benerecetti, A. S. (2004). Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. *Am. J. Hum. Genet.* **74:** 1023–1034.

Senner, W. M. (1991). Theories and Myths on the Origins of Writing: A Historical Overview. In: *The Origins of Writing.* (Ed.) Senner, W. M. University of Nebraska Press. (Nebraska, USA).

Shamir, I. and Shavit, S. (Eds.) (1986). Encylopedia of Jewish Histroy. Events and Eras of the Jewish People. Massada Publishers (Israel).

Sigurðardóttir, S., Helgason, A., Gulcher, J. R., Stefansson, K., and Donnelly, P. (2000). The Mutation Rate in the Human MtDNA Control Region. *Am. J. Hum. Genet.* **66:** 1599–1609.

Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J., and Barbujani, G. (2000). Geographic Patterns of mtDNA Diversity in Europe. *Am. J. Hum. Genet.* **66:** 262-278.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S. F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S. P., Waterston, R. H., Wilson, R. K., Rozen, S., Page, D. C. (2003). The Male-Specific Region of the Human Y Chromosome is a Mosaic of Discrete Sequence Classes. *Nature* **423:** 825-37.

Skorecki, K., Seli, S., Blazer, S., Bradman, R., Bradman, N., Waburton, P. J., Ismajiowicz, M., Hammer, M. F. (1997). Y Chromosomes of Jewish Priests. *Nature* **385:** 32.

Smith, D. G., Malhi, R. S., Eshleman, J., Lorenz, J. G., and Kaestle, F. A. (1999). Distribution of mtDNA Haplogroup X Among Native North Americans. *Am. J. Phys. Anthropol.* **110:** 271-284.

Sokal, R. R., Harding, R. M., Lasker, G. W., Mascie-Taylor, C. G. (1992). A Spatial Analysis of 100 Surnames in England and Wales. *Ann. Hum. Biol.* **19:** 445-76.

Soodyall, H., Vigilant, L., Hill, A. V., Stoneking, M., and Jenkins, T. (1996). MtDNA Control-Region Sequence Variation Suggests Multiple Independent Origins of an "Asian-Specific" 9-bp Deletion in Sub-Saharan Africans. *Am. J. Hum. Genet.* **58:** 595-608.

Soodyall, H., Jenkins, T., Mukherjee, A., Du Toit, E., Roberts, D. F. and Stoneking, M. (1997). The Founding Mitochondrial DNA Lineages of Tristan da Cunha Islanders. *Am. J. Phys. Anthropol.* **104:** 157–166.

Soodyal, H., Nebel, A., Morar, B., and Jenkins, T. (2003). Genealogy and Genes: Tracing the Founding Father of Tristan da Cunha. *Eur. J. Hum. Genet.* **11:** 705-709.

Spurdle, A. B. and Jenkins, T. (1996). The Origins of the Lemba "Black Jews" of Southern Africa: Evidence from p12F2 and Other Y-Chromosome Markers. *Am. J. Hum. Genet.* **59:** 1126-1133.

Storz, J. F., Ramakrishnan, U., and Alberts, S. C (2001). Determinants of Effective Population Size for Loci with Different Modes of Inheritance. *J. Hered.* **92:** 497-502.

Strachan, T. and Read, A. P. (1999). Human Molecular Genetics 2. BIOS (Oxford, UK).

Stevenson, J. (1998). Saxon Bodies. In: *London Bodies. The Changing Shape of Londoners from Prehistoric Times to the Present Day.* (Compiled by Werner, A.) Museum of London (London, UK).

Stevenson, W. The Middle Ages 1000-1500. In Jamieson, A. G. (Ed.). (1986). *A People of the Sea: The Maritime History of the Channel Islands.* Methuen (London, UK).

Stringer, C. (2000). Coasting Out of Africa. *Nature* **405:** 24-26.

Stringer, C. (2003). Out of Ethiopia. *Nature* **423:** 692-695.

Stumpf, M. P. H. and Goldstein D. B. (2003). Demography, Recombination Hotspot Intensity, and the Block Structure of Linkage Disequilibrium. *Curr. Biol.* **13:** 1-8.

Schwartz, M., Vissing, J. (2002). Paternal Inheritance of Mitochondria. *N. Engl. J. Med.*22:576-80.

Schwartz, M., Vissing, J. (2004). No Evidence For Paternal Inheritance of mtDNA in Patients with Sporadic mtDNA Mutations. *J. Neurol. Sci.* **15;** 99-101.

Sykes, B. and Irven, C. (2000). Surnames and the Y chromosome. *Am. J. Hum. Genet.* **66:** 1417-1419.

Thomas, M. G., Skorecki, K., Ben-Ami, H., Parfitt, T., Bradman N., Goldstein D. B. (1998). Origins of Old Testament Priests. *Nature* **394:** 138-140.

Thomas, M. G., Bradman, N., Flinn, H. M., (1999). High Throughput Analysis of 10 Microsatellite and 11 Diallelic Polymorphisms on the Human Y-Chromosome. *Hum. Genet.* **105:** 577-581.

Thomas, M G., Parfitt, T., Weiss, D. A., Skorecki, K., Wilson, J. F., le Roux, M., Bradman, N., and Goldstein, D. B. (2000). Y Chromosomes Traveling South: The Cohen Modal Haplotype and the Origins of the Lemba—the "Black Jews of Southern Africa". *Am. J. Hum. Genet.* **66:** 674–686.

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., and Goldstein, D. B. (2002). Founding Mothers of Jewish Communities: Geographically Separated Jewish Groups Were Independently Founded by Very Few Female Ancestors *Am. J. Hum. Genet.* **70:** 1411–1420.

Tishkoff, S. A. and Verrelli, B. C. (2003). Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease. *Annu. Rev. Genomics Hum. Genet.* **4:** 293–340.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.-L., Wallace, D. C. (1996). Classification of European mtDNAs From an Analysis of Three European Populations. *Genetics* **144:** 1835-1850.

Torroni, T., Rengo, C., Guida, V., Cruciani, F., Sellitto, D., Coppa, A., Calderon, F. L., Simionati, B., Valle, G., Richards, M., Macaulay, V., Scozzari, R. (2001a). Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates? *Am. J. Hum. Genet.* **69:** 1348–1356

Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H.V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M.-L., Huoponen, K., Laitinen, V., Koivumäki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A. S., Semino, O., and Scozzari, R. (2001)b. A Signal, from Human mtDNA, of Postglacial Recolonization in Europe. *Am. J. Hum. Genet.* **69:** 844-852.

Tyler-Smith, C. and McVean, G. (2003). The Comings and Goings of a Y Polymorphism. *Nat. Genet.* **35:** 201-202.

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonné-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W., Cavalli-Sforza, L. L., and Oefner, P. J. (2000). Y Chromosome Sequence Variation and the History of Human Populations. *Nat. Genet.* **26:** 358-361.

Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Lahr, M. M., Foley, R. A., Oefner, P. J., and Cavalli-Sforza, L. L. (2001). The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations. *Ann. Hum. Genet.* **65:** 43-62.

Van der Veen, L. J. and Hombert, J.-M. (2001). On the Origin and Diffusion of Bantu: a Multidisciplinary Approach. *Proceedings of the 32nd Annual Conference on African Languages (ACAL 32).* Berkeley.

Vernesi, C., Caramelli, D., Dupanloup, I., Bertorelle, G., Lari, M., Cappellini, E., Moggi-Cecchi, J., Chiarelli, B., Castrì, L., Casoli, A., Mallegni, F., Lalueza-Fox, C., and Barbujani, G. (2004). The Etruscans: A Population-Genetic Study. *Am. J. Hum. Genet.* **74:** 694–704.
Vickers, L. (1998). Trends in Migration in the UK. *Population Trends* **94:** 25-34.

Vollrath, D., Foote, S., Hilton, A., Brown, L. G., Beer-Romero, P., Bogan, J. S., and Page, D. C. (1992). The Human Y Chromosome: A 43-Interval Map Based on Naturally Occurring Deletions. *Science* **258:** 52-59.

Walvin, J. (2000). Britain's Slave Empire. Tempus (Glos, UK).

Weale, M. E., Yepiskoposyan, L., Jager, R. F., Hovhannisyan, N., Khudoyan, A., Burbage-Hall, O., Bradman, N., Thomas, M. G. (2001). Armenian Y Chromosome Haplotypes Reveal Strong Regional Structure Within a Single Ethno-National Group. *Hum. Genet.* **109:** 659-674.

Weale, M.E., Weiss D. A., Jager, R. F., Bradman, N., and Thomas, M. G. (2002). Y Chromosome Evidence for Anglo-Saxon Mass Migration. *Mol. Biol. Evol.* **19:** 1008-1021.

Welch, M. (1992). Anglo Saxon England. B. T. Batsford Ltd./English Heritage (London, UK).

Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N. M., Zerjal, T., Webster, M. T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A., and Bodmer, W. F. (2001). The Eurasian Heartland: A Continental Perspective On Y-Chromosome Diversity. *Proc. Natl. Acad. Sci. USA.* **98:** 10244-10249.

Wigoder, G. (Ed.) (1974). Encylopedia Dictionary of Judaica. Leon Amiel Publisher (New York, USA).

Wildman, D. E., Uddin, M., Liu, G., Grossman, L. I., and Goodman, M. (2003). Implications of Natural Selection in Shaping 99.4% Nonsynonymous DNA Identity Between Humans and Chimpanzees: Enlarging Genus *Homo*. *Proc. Natl. Acad. Sci. USA.* **100:** 7181-7188.

Wilson, J. F. and Goldstein, D. B. (2000). Consistent Long-Range Linkage Disequilibrium Generated by Admixture in a Bantu-Semitic Hybrid Population. *Am. J. Hum. Genet.* **67:** 926–935.

Wilson, J. F., Weiss, D. A., Richards, M., Thomas, M. G., Bradman, N., and Goldstein, D. G. (2001a). Genetic Evidence for Different Male and Female Roles During Cultural Transitions in the British Isles. *Proc. Natl. Acad. Sci. USA*. **98:** 5078-83.

Wilson, J. F., Weale, M. E., Smith, A. C., Gratrix, F., Fletcher, B., Thomas, M. G., Bradman, N., and Goldstein, D. B. (2001)b. Population Genetic Structure of Variable Drug Response. *Nat. Genet.* **29:** 265-269.

Wolpoff, M. H., Hawks, J., and Caspari, R. (2000). Multiregional, Not Multiple Origins. *Am. J. Phys. Anthropol.* **112:** 129-136.

Wormald, P. (1991). The Ninth Century. In: *The Anglo-Saxons.* (Ed.) Campbell, J. Penguin (London).
Y Chromosome Consortium (2002). A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Res.* **12:** 339-48.

Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A.-V., Krumina, A., Kučinskas, V., Hurles, M. E., Tyler-Smith, C. (2001). Geographical, Linguistic, and Cultural Influences on Genetic Diversity: Y-Chromosomal Distribution in Northern European Populations. *Mol. Biol. Evol.* **18:** 1077-1087.

Zerjal, T., Wells, R. S., Yuldasheva. N., Ruzibakiev, R., and Tyler-Smith, C. (2002). A Genetic Landscape Reshaped by Recent Events: Y-Chromosomal Insights into Central Asia. *Am. J. Hum. Genet.* **71**: 466-482.

Zoloth, L. (2003). Yearning for the Long Lost Home: The Lemba and the Jewish Narrative of Genetic Return. *Developing World Bioethics* **3:** 127-132.

Appendices

Appendix. Table A.1. List of Suppliers

ABgene®

Thermo-Fast® Detection Plate, 96-well

Thermo-Fast® Low Profile Plate, 96-well

Pre-Aliquoted PCR optimisation plate (1.5mM MgCI and without loading dye)

Amersham Biosciences, UK (including Pharmacia Biotech)

Shrimp alkaline phosphatase

Exonuclease I

dNTP set

Applied Biosystems, USA (including any products from companies now subsumed within Applied Biosystems)

Oligonucleotides labelled with NED™ fluorescent dye

BigDye® Terminator v1.1 Cycle Sequencing Kit (including 5X Sequencing Buffer and Ready Reaction Mix)

BigDye® Terminator v1.1/v3.1 Sequencing Buffer (5X)

GeneAmp® PCR System 9700

GeneAmp® PCR System 2700

ABI PRISM ® 377 DNA Sequencer (and appropriate reagents)

ABI PRISM ® 3700 DNA Sequencer (and appropriate reagents)

SeqEd™ v1.0.3

ABI PRISM® GeneScan® v3.1 for Macs

ABI PRISM® GeneScan® v3.7 for PCs

TAMRA™ 350

Dextran Blue

Clontech, USA

TaqStart™ Antibody

Gene Codes, USA

Sequencher ™

HT Biotechnology Ltd., UK

Super-Taq Polymerase

10X PCR Buffer

Jencons, UK

ALC PK120 CWS plate centrifuge

Mikro 20 microfuge

Microzone, UK

MicroCLEAN

Millipore, USA

MultiScreen ® HV plate, 96-well

MWG, Germany

All labelled and unlabelled oligonucleotides, expect those labelled with NED™ fluorescent dye (see Applied Biosystems above)

New England Biolabs Inc, USA

The following enzymes:

HinFI, AflIII, BclI, BsrGI, NlaIII, BsrI, DraIII

The above enzymes are supplied with the appropriate NEB Buffer (2, 3, 4), and Bovine Serum Albumin (BSA).

Promega

Promega Wizard ® Genomic DNA Purification Kit

Qiagen

HotStarTaq DNA Polymerase (which includes 10x PCR Buffer and 25mM $MgCl_2$)

Sarstedt, Germany

Transportation Swab, 101x16.5mm

Sigma-Aldrich

Sigma Magnesium Chloride, 1.00 M

Sephadex ™ G50 Superfine

All other products and reagents were purchased from lab suppliers such as VWR International and Sigma-Aldrich.

Appendix. Table A.2. Sequences of the Primers Used in This Thesis

Mulitplex Kit	*Primer*	*Fluorescent Label (5')*	*Sequence 5'-3'*
YSTR1	DYS19L	TET	CTA CTG AGT TTC TGT TAT AGT
YSTR1	DYS19R	-	ATG GCA TGT AGT GAG GAC A
YSTR1	DYS388L	TET	GTG AGT TAG CCG TTT AGC GA
YSTR1	DYS388R	-	CAG ATC GCA ACC ACT GCG
YSTR1	DYS390L	-	TAT ATT TTA CAC ATT TTT GGG CC
YSTR1	DYS390r	FAM	TGA CAG TAA AAT GAA CAC ATT GC
YSTR1	DYS391L	FAM	CTA TTC ATT CAA TCA TAC ACC CAT AT
YSTR1	DYS391r	-	ACA TAG CCA AAT ATC TCC TGG G
YSTR1	DYS392L	-	AAA AGC CAA GAA GGA AAA CAA A
YSTR1	DYS392R	HEX	CAG TCA AAG TGG AAA GTA GTC TGG
YSTR1	DYS393L	-	GTG GTC TTC TAC TTG TGT CAA TAC
YSTR1	DYS393R	HEX	AAC TCA AGT CCA AAA AAT GAG G
EURO1	M9 long F	-	CAT TGA ACG TTT GAA CAT GTC
EURO1	M9 long R	TET	TGC AGC ATA TAA AAC TTT CAG G
EURO1	92R7 U	HEX	TCA GAA AGA TAG TAA GAG GAA CAC TTC
EURO1	92R7 R	-	GCA TTG TTA AAT ATG ACC AGC A
EURO1	M17 F	-	GTG GTT GCT GGT TGT TAC GT
EURO1	M17 R	TET	AGC TGA CCA CAA ACT GAT GTA GA
EURO1	M173 F	-	ACA ATT CAA GGG CAT TTT GTG C
EURO1	M173 R	FAM	CTT ACT CAG TAT GGG TAA AAG AAA TGC
EURO1	M170 F	-	TTA CTA TTT TAT TTA CTT AAA AAT CAT TGA TC
EURO1	M170 R	HEX	CCA ATT ACT TTC AAC ATT TAA GAC C
EURO1	M172 F	TET	TTA GCC AGA TGA CCA GGA TGC
EURO1	M172 R	-	GAA AAT AAT AAT TGA AGA CCT TTT GAG T
-	M26 F	-	CAA TTT CTT TCT GAA TTA GAA TGA TC
-	M26 R	HEX	CCA TAC ACA AGG ATG CAG CAC
-	M89 F	HEX	GAA AGT GGG GCC CAC AG
-	M89 R	-	AAC TCA GGC AAA GTG AGA CAT G
-	TAT F	-	GAC TCT GAG TGT AGA CTT GTG A
-	TAT R	FAM	GAA GGT GCC GTA AAA GTG TGA A
-	12f2 D	HEX	CTG ACT GAT CAA AAT GCT TAC AGA TC
-	12f2 G	-	GGA TCC CTT CCT TAC ACC TTA TAC
-	M35II F	FAM	GAA ACT GAG AGG GCA AGG TC
-	M35II R	-	GGA GCT TCT GCC TGT TGC
-	SRY10831L	FAM	TCA TTC AGT ATC TGG CCT CTT G
-	SRY10831R	-	CAC CAC ATA GGT GAA CCT TGA A
-	conH1	-	CCT GAA GTA GGA ACC AGA TG
-	conL2	-	CAC CAT TAG CAC CCA AAG CT
-	conH3	-	CGG AGC GAG GAG AGT AGC

Suppliers for the primers can be found in Appendix, Table A.1

Appendix. Table A.3. Y-Chromosome Microsatellite Haplotype and UEP Information for the British and European Populations

Haplotype Number	Microsatellite Locus [a]								Count in Population																												
	388	393	392	19	390	391	HG [b]	Total	Sket	Ork	Dur	Wis	Sth	Ptl	Oban	Mpt	Pnt	IoM	York	Saw	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chip	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht1	10	12	14	15	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht2	10	13	11	15	24	10	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht3	10	13	13	14	23	11	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht4	11	13	13	14	23	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-
ht5	11	13	13	15	24	10	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht6	12	11	13	14	23	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht7	12	11	13	14	24	11	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht8	12	12	11	14	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht9	12	12	13	13	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht10	12	12	13	13	25	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht11	12	12	13	14	23	10	R1*(xR1a1)	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	2	-	-	-	-	-	-	-	-	-	-
ht12	12	12	13	14	23	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht13	12	12	13	14	24	10	R1*(xR1a1)	4	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-
ht14	12	12	13	14	24	11	R1*(xR1a1)	11	1	1	-	-	-	-	-	-	-	-	-	1	-	-	1	2	1	1	-	-	-	-	1	1	-	-	-	-	1
ht15	12	12	13	14	25	10	R1*(xR1a1)	6	-	-	-	-	-	2	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht16	12	12	13	14	25	11	R1*(xR1a1)	6	-	-	-	-	-	-	-	-	-	-	-	1	3	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-
ht17	12	12	13	15	23	10	R1*(xR1a1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht18	12	12	13	15	23	11	R1*(xR1a1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht19	12	12	13	15	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht20	12	12	13	16	23	12	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht21	12	12	14	14	22	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht22	12	12	14	14	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht23	12	12	14	14	25	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-
ht24	12	13	11	14	23	10	R1*(xR1a1)	2	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht25	12	13	11	14	24	10	R1*(xR1a1)	2	-	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht26	12	13	11	14	24	11	R1*(xR1a1)	6	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2	-	-	-	-	-	-	-	-	-	-	-	3	-	-
ht27	12	13	11	14	25	11	R1*(xR1a1)	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht28	12	13	11	15	23	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht29	12	13	12	13	23	10	R1*(xR1a1)	2	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht30	12	13	12	14	22	10	R1*(xR1a1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht31	12	13	12	14	22	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht32	12	13	12	14	23	10	R1*(xR1a1)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-
ht33	12	13	12	14	23	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht34	12	13	12	14	24	9	R1*(xR1a1)	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht35	12	13	12	14	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht36	12	13	12	14	24	11	R1*(xR1a1)	13	-	-	1	-	-	-	1	-	-	-	-	1	1	2	-	-	1	2	2	-	-	-	-	-	-	-	-	1	-	1
ht37	12	13	12	14	25	11	R1*(xR1a1)	7	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	2	-	-	-	-	-	-	-	-	-	-	-	1
ht38	12	13	12	15	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht39	12	13	12	15	25	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht40	12	13	12	15	26	12	R1*(xR1a1)	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht41	12	13	12	17	24	10	R1*(xR1a1)	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht42	12	13	13	11	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht43	12	13	13	12	23	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht44	12	13	13	12	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht45	12	13	13	12	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht46	12	13	13	13	23	10	R1*(xR1a1)	2	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht47	12	13	13	13	23	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht48	12	13	13	13	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht49	12	13	13	13	24	11	R1*(xR1a1)	12	-	-	-	2	-	-	-	1	-	4	-	-	-	-	-	1	-	1	-	-	-	-	3	-	-	-	-	-	-	-
ht50	12	13	13	13	24	12	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht51	12	13	13	13	25	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht52	12	13	13	13	26	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht53	12	13	13	14	21	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.3. continued

									Shet	Ork	Dur	Wis	Sth	Ptl	Oban	Mpt	Pnt	IoM	York	Sow	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chip	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht 54	12	13	13	14	22	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 55	12	13	13	14	22	11	R1*(xR1a1)	8	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	2	-	-	-	-	1	1	-	1	1	-
ht 56	12	13	13	14	23	10	R1*(xR1a1)	78	-	6	14	1	3	1	2	2	2	-	2	1	3	2	2	2	1	4	1	1	4	4	2	2	2	4	1	2	4	3
ht 57	12	13	13	14	23	11	R1*(xR1a1)	190	5	8	1	5	3	4	1	8	12	9	5	6	11	5	5	6	-	11	7	7	5	15	11	8	4	1	1	15	3	8
ht 58	12	13	13	14	23	12	R1*(xR1a1)	5	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1	-	-	-	-	-	1	1	-	-	-	-	-	-	-
ht 59	12	13	13	14	23	13	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht 60	12	13	13	14	24	9	R1*(xR1a1)	5	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-	1	-	-	-	1
ht 61	12	13	13	14	24	10	R1*(xR1a1)	224	8	16	8	8	10	11	9	15	5	7	5	9	9	4	2	9	4	5	4	5	3	8	5	5	7	6	8	9	13	7
ht 62	12	13	13	14	24	11	R1(xR1a1)	381	10	15	6	19	3	7	8	19	18	12	4	13	12	12	22	8	10	22	14	10	15	16	12	9	13	8	13	17	20	14
ht 63	12	13	13	14	24	12	R1*(xR1a1)	14	-	1	-	-	-	-	-	-	1	-	-	-	-	-	2	1	1	-	3	-	-	-	-	-	1	2	1	-	1	-
ht 64	12	13	13	14	25	10	R1*(xR1a1)	35	2	-	-	2	2	1	1	-	2	1	1	-	1	-	1	2	-	5	-	-	2	1	-	-	4	1	-	5	-	1
ht 65	12	13	13	14	25	11	R1*(xR1a1)	78	4	4	1	7	3	1	2	4	3	1	1	1	2	3	9	1	3	2	2	2	2	-	1	4	2	-	-	2	6	5
ht 66	12	13	13	14	25	12	R1*(xR1a1)	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-
ht 67	12	13	13	14	26	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
ht 68	12	13	13	14	26	11	R1*(xR1a1)	6	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	1	-	1	-	-	-	-	1	-	1
ht 69	12	13	13	15	23	10	R1*(xR1a1)	7	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	2	-	1
ht 70	12	13	13	15	23	11	R1*(xR1a1)	21	1	-	-	-	-	-	-	-	-	-	-	2	3	-	-	1	-	1	-	-	-	3	-	-	3	-	-	2	1	4
ht 71	12	13	13	15	23	12	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht 72	12	13	13	15	24	10	R1*(xR1a1)	17	-	-	-	1	-	1	1	2	1	-	-	-	1	-	-	1	-	1	1	-	-	1	1	-	-	2	-	2	-	1
ht 73	12	13	13	15	24	11	R1*(xR1a1)	42	-	4	-	1	-	-	-	1	7	-	2	-	3	-	5	1	-	1	1	1	2	-	3	-	-	1	2	4	2	1
ht 74	12	13	13	15	25	10	R1*(xR1a1)	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
ht 75	12	13	13	15	25	11	R1*(xR1a1)	9	-	-	-	2	-	-	-	-	-	-	-	-	-	-	3	2	-	-	-	-	1	-	-	-	-	-	-	-	1
ht 76	12	13	13	15	26	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht 77	12	13	13	16	24	10	R1*(xR1a1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 78	12	13	13	16	24	11	R1*(xR1a1)	7	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-	1	2	-	-	-	-	-	-
ht 79	12	13	13	17	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht 80	12	13	14	13	23	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht 81	12	13	14	13	23	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 82	12	13	14	13	24	10	R1*(xR1a1)	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 83	12	13	14	14	22	10	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 84	12	13	14	14	23	10	R1*(xR1a1)	3	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-
ht 85	12	13	14	14	23	11	R1*(xR1a1)	9	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	1	1	1	-	2	-	-
ht 86	12	13	14	14	24	9	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht 87	12	13	14	14	24	10	R1*(xR1a1)	21	1	8	2	-	2	-	-	-	-	-	2	-	2	-	-	1	-	1	-	-	-	-	1	-	-	-	-	-	-	1
ht 88	12	13	14	14	24	11	R1*(xR1a1)	22	-	-	-	1	-	-	4	-	-	1	-	-	-	-	1	2	1	-	3	-	1	1	-	-	-	3	-	-	-	4
ht 89	12	13	14	14	24	12	R1*(xR1a1)	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 90	12	13	14	14	25	10	R1*(xR1a1)	12	-	-	-	1	1	-	-	1	-	-	-	-	-	-	-	4	2	-	-	-	-	1	-	-	-	-	-	-	1	1
ht 91	12	13	14	14	25	11	R1*(xR1a1)	44	-	-	1	5	2	3	3	3	-	3	-	-	-	-	1	5	7	2	3	-	-	-	-	-	-	-	-	1	3	2
ht 92	12	13	14	14	26	11	R1*(xR1a1)	2	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht 93	12	13	14	15	23	11	R1*(xR1a1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 94	12	13	14	15	24	10	R1*(xR1a1)	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	1	-
ht 95	12	13	14	15	24	11	R1*(xR1a1)	4	-	1	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht 96	12	13	14	15	25	10	R1*(xR1a1)	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 97	12	13	14	15	25	11	R1*(xR1a1)	3	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 98	12	13	14	16	24	10	R1*(xR1a1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 99	12	13	14	16	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ht 100	12	13	15	14	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht 101	12	13	15	14	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 102	12	13	15	14	25	11	R1*(xR1a1)	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 103	12	13	15	15	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 104	12	13	16	14	23	11	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 105	12	14	11	16	25	11	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 106	12	14	12	14	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht 107	12	14	12	15	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.3. continued

									Shet	Ork	Dur	Wis	Sth	Ptl	Oban	Mpt	Pnt	IoM	York	Sow	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chp	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht 108	12	14	13	13	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 109	12	14	13	13	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ht 110	12	14	13	14	22	10	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 111	12	14	13	14	23	10	R1*(xR1a1)	6	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	1	1	-	1	-	-	-
ht 112	12	14	13	14	23	11	R1*(xR1a1)	15	-	-	-	-	-	-	-	-	-	1	-	1	-	1	-	-	-	1	1	-	-	-	1	3	1	1	2	1	1	-
ht 113	12	14	13	14	23	12	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht 114	12	14	13	14	24	10	R1*(xR1a1)	8	-	-	-	-	-	1	-	1	-	-	-	-	1	1	-	-	-	-	-	-	-	-	2	-	1	1	-	-	-	-
ht 115	12	14	13	14	24	11	R1*(xR1a1)	26	4	1	-	-	1	-	-	1	1	-	-	-	-	1	7	-	2	1	2	-	-	1	-	1	1	1	-	1	-	-
ht 116	12	14	13	14	24	12	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 117	12	14	13	14	25	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 118	12	14	13	14	25	11	R1*(xR1a1)	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	2	-	1	-	-	-	1	-	-	1	-
ht 119	12	14	13	15	23	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht 120	12	14	13	15	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht 121	12	14	13	15	24	11	R1*(xR1a1)	3	-	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht 122	12	14	13	16	25	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht 123	12	14	14	14	23	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht 124	12	14	14	14	24	11	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 125	12	14	14	14	25	11	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 126	12	14	14	14	26	11	R1*(xR1a1)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 127	12	14	15	14	25	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 128	12	15	13	14	23	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht 129	12	15	13	14	24	10	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 130	12	15	13	14	24	11	R1*(xR1a1)	5	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	1
ht 131	12	15	13	14	25	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht 132	12	15	13	14	25	11	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 133	12	15	14	15	24	10	R1*(xR1a1)	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 134	12	17	13	15	24	11	R1*(xR1a1)	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 135	13	13	9	16	26	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht 136	13	13	11	15	23	10	R1*(xR1a1)	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 137	13	13	12	14	23	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 138	13	13	12	14	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 139	13	13	12	15	23	10	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 140	13	13	13	14	23	10	R1*(xR1a1)	2	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht 141	13	13	13	14	23	11	R1*(xR1a1)	6	-	1	-	-	-	-	-	-	-	-	-	-	1	1	2	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht 142	13	13	13	14	23	12	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 143	13	13	13	14	24	10	R1*(xR1a1)	2	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht 144	13	13	13	14	24	11	R1*(xR1a1)	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	1	-	-
ht 145	13	13	13	14	25	10	R1*(xR1a1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 146	13	13	13	15	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 147	13	13	13	16	23	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 148	13	13	13	16	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht 149	13	13	14	14	25	12	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht 150	13	14	13	14	22	12	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 151	13	14	13	14	24	11	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 152	13	14	13	15	23	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht 153	13	14	13	15	27	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 154	14	13	11	13	22	10	R1*(xR1a1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 155	14	13	13	14	24	10	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 156	14	13	13	14	24	11	R1*(xR1a1)	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht 157	14	13	13	16	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht 158	14	14	13	14	23	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht 159	17	13	13	14	24	11	R1*(xR1a1)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ht 160	10	13	11	15	23	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.3. continued

									Shet	Ork	Dur	Wis	Sth	Ptl	Oban	Mpl	Pnt	IoM	York	Sow	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chp	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht161	10	13	11	15	24	10	R1a1	5	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	1	-	-	2	-	-	-	-	-	-	-
ht162	10	13	11	15	25	10	R1a1	3	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-
ht163	10	13	11	16	24	10	R1a1	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht164	10	13	11	16	25	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht165	10	13	11	16	26	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht166	10	13	11	17	24	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht167	10	13	11	17	25	10	R1a1	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-
ht168	10	13	12	16	25	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht169	10	14	11	15	25	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht170	11	13	11	16	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht171	12	12	11	15	22	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht172	12	12	11	15	23	10	R1a1	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht173	12	12	11	15	25	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht174	12	12	11	16	25	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht175	12	12	13	14	24	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht176	12	13	11	13	25	11	R1a1	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht177	12	13	11	14	24	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht178	12	13	11	14	25	10	R1a1	3	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht179	12	13	11	14	25	11	R1a1	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht180	12	13	11	14	26	10	R1a1	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht181	12	13	11	15	23	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht182	12	13	11	15	24	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht183	12	13	11	15	24	11	R1a1	10	1	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	2	-	-	-	3	-
ht184	12	13	11	15	25	10	R1a1	14	1	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	2	-	-	-	-	1	-	3	4	-
ht185	12	13	11	15	25	11	R1a1	43	4	3	-	1	-	-	1	-	5	1	1	-	-	-	-	1	-	1	-	1	-	-	-	-	-	-	-	1	22	1
ht186	12	13	11	15	25	12	R1a1	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht187	12	13	11	15	26	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht188	12	13	11	15	26	11	R1a1	4	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	2	-
ht189	12	13	11	16	24	10	R1a1	6	-	1	-	-	-	-	-	-	-	3	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht190	12	13	11	16	23	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht191	12	13	11	16	24	11	R1a1	3	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-
ht192	12	13	11	16	25	10	R1a1	21	3	7	-	-	-	-	-	-	-	1	-	-	1	-	-	-	-	1	-	-	-	-	-	-	1	-	1	6	-
ht193	12	13	11	16	25	11	R1a1	37	3	3	1	4	-	-	-	1	-	1	-	1	1	-	-	-	-	-	-	-	1	-	-	1	-	-	-	6	13	1
ht194	12	13	11	16	26	10	R1a1	3	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1	-	-	-	-	-	-
ht195	12	13	11	16	26	11	R1a1	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht196	12	13	11	16	26	12	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht197	12	13	11	17	24	10	R1a1	2	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht198	12	13	11	17	24	11	R1a1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-
ht199	12	13	11	17	25	10	R1a1	3	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-
ht200	12	13	11	17	25	11	R1a1	3	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-
ht201	12	13	11	17	26	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht202	12	13	12	13	23	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht203	12	13	12	15	25	11	R1a1	2	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht204	12	13	13	14	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht205	12	13	13	15	25	11	R1a1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-
ht206	12	13	14	15	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht207	12	14	11	15	25	11	R1a1	4	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2	-
ht208	12	14	11	16	25	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht209	12	14	11	16	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht210	12	14	13	13	24	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht211	12	14	13	14	24	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht212	13	13	11	15	25	11	R1a1	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht213	13	13	11	16	24	10	R1a1	2	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht214	13	14	11	14	22	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht215	13	14	11	15	25	10	R1a1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-

continued

Appendix. Table A.3. continued

									Shet	Ork	Dur	Wis	Sth	Ptl	Oban	Mpt	Pnt	IoM	York	Sow	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chip	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht216	14	13	11	15	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht217	14	14	11	16	25	10	R1a1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht218	10	14	12	16	23	11	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht219	11	15	12	14	23	10	I*(xI1b2)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht220	12	13	11	13	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht221	12	13	11	14	22	10	I*(xI1b2)	2	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht222	12	13	11	14	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht223	12	13	11	15	23	10	I*(xI1b2)	2	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht224	12	13	11	15	25	10	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht225	12	13	11	15	25	11	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht226	12	13	13	14	24	11	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht227	12	13	13	14	25	11	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht228	12	13	14	13	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht229	12	14	11	13	26	9	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht230	12	14	11	14	22	10	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht231	12	14	11	14	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht232	12	14	11	15	21	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht233	12	14	11	15	25	10	I*(xI1b2)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht234	12	14	11	15	25	11	I*(xI1b2)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht235	12	14	11	17	26	11	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht236	12	14	12	14	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht237	12	14	12	15	24	11	I*(xI1b2)	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht238	12	14	13	14	24	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht239	13	12	11	15	25	10	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht240	13	12	12	15	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht241	13	13	11	14	22	10	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht242	13	13	11	14	23	10	I*(xI1b2)	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht243	13	13	11	14	25	11	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht244	13	13	11	15	22	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht245	13	13	11	15	24	11	I*(xI1b2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-
ht246	13	13	11	15	25	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht247	13	13	11	15	25	11	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht248	13	13	11	16	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht249	13	13	11	16	24	10	I*(xI1b2)	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	1	-	-
ht250	13	13	11	16	24	11	I*(xI1b2)	3	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht251	13	13	11	17	25	10	I*(xI1b2)	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht252	13	13	11	17	25	11	I*(xI1b2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-
ht253	13	13	11	17	25	13	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht254	13	13	11	17	26	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht255	13	13	12	15	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht256	13	13	12	15	23	11	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht257	13	13	12	15	24	10	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht258	13	13	12	15	24	11	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht259	13	13	12	16	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht260	13	13	12	16	24	11	I*(xI1b2)	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht261	13	13	13	16	23	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht266	13	14	11	16	25	11	I*(xI1b2)	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht267	13	14	11	16	26	11	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht268	13	14	11	17	24	11	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht269	13	14	11	17	25	11	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht270	13	14	11	17	25	12	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht271	13	14	11	17	26	10	I*(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht262	13	14	11	15	22	10	I*(xI1b2)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.3. continued

									Shet	Ork	Dur	Wis	Sth	Ptl	Oban	Mpt	Pnt	IoM	York	Sow	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chip	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht263	13	14	11	15	23	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht264	13	14	11	15	26	12	I²(x11b2)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht265	13	14	11	16	22	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht272	13	14	11	17	26	11	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht273	13	14	12	14	23	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht274	13	14	12	15	21	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht275	13	14	12	15	22	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht276	13	14	12	15	23	9	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht277	13	14	12	15	23	10	I²(x11b2)	16	1	1	-	-	-	-	-	-	-	-	3	1	-	-	-	-	-	2	-	-	-	1	-	-	-	-	-	6	1	-
ht278	13	14	12	15	23	11	I²(x11b2)	2	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht279	13	14	12	15	24	10	I²(x11b2)	2	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht280	13	14	12	16	22	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht281	13	14	12	16	23	10	I²(x11b2)	4	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-
ht282	13	14	12	16	23	11	I²(x11b2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1
ht283	13	14	12	16	24	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht284	13	14	12	17	23	11	I²(x11b2)	2	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht285	13	14	13	15	22	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht286	13	14	13	15	23	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht287	13	14	13	16	23	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht288	13	14	13	16	24	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht289	13	15	12	14	24	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht290	13	15	12	15	22	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht291	13	15	12	15	23	10	I²(x11b2)	10	1	-	-	-	-	-	-	-	1	1	-	2	-	-	-	-	-	1	-	-	1	1	-	-	-	-	-	1	1	-
ht292	13	15	12	15	23	11	I²(x11b2)	2	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht293	13	15	12	15	23	13	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht294	13	15	12	15	24	10	I²(x11b2)	4	-	-	-	2	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht295	13	15	12	15	25	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht296	13	15	12	15	24	11	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht297	13	15	12	16	23	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht298	13	15	13	15	22	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht299	13	15	13	15	24	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht300	13	15	13	16	23	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht301	13	15	13	17	23	10	I²(x11b2)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht302	13	15	14	15	23	11	I²(x11b2)	2	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht303	13	16	12	15	23	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht304	14	9	11	14	22	10	I²(x11b2)	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht305	14	11	11	14	23	10	I²(x11b2)	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht306	14	12	11	14	22	10	I²(x11b2)	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	2	-	-
ht307	14	12	11	14	22	11	I²(x11b2)	2	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht313	14	13	10	14	22	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht314	14	13	10	14	23	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht315	14	13	11	13	22	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht316	14	13	11	13	25	11	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht317	14	13	11	14	21	9	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht318	14	13	11	14	21	10	I²(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht308	14	12	11	14	23	10	I²(x11b2)	3	-	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht309	14	12	11	15	22	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht310	14	12	11	15	23	10	I²(x11b2)	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht311	14	12	12	14	22	10	I²(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht312	14	12	12	15	23	10	I²(x11b2)	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht319	14	13	11	14	22	10	I²(x11b2)	73	1	4	-	-	1	-	-	1	4	1	1	2	6	-	-	-	-	8	-	4	2	2	3	2	5	-	-	16	6	4
ht320	14	13	11	14	22	11	I²(x11b2)	9	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	2	1	1
ht321	14	13	11	14	22	12	I²(x11b2)	2	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht322	14	13	11	14	23	10	I²(x11b2)	81	2	1	4	6	2	-	1	2	5	3	5	1	2	3	-	-	-	7	-	2	1	-	1	2	5	1	-	10	13	2
ht323	14	13	11	14	23	11	I²(x11b2)	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	2	1	-

continued

Appendix. Table A.3. continued

									Shet	Ork	Dur	Wis	Sth	Ptl	Oban	Mpt	Pnt	IoM	York	Sow	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chip	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht324	14	13	11	14	24	10	I^(xIIb2)	8	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	4	-
ht325	14	13	11	15	22	9	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht326	14	13	11	15	22	10	I^(xIIb2)	27	-	1	1	-	-	-	-	1	-	1	2	-	-	3	-	-	-	-	-	-	1	2	-	-	-	-	-	4	10	1
ht327	14	13	11	15	22	11	I^(xIIb2)	5	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	1	-	-	-	-	1	-	1	-	-	-	-	-
ht328	14	13	11	15	23	10	I^(xIIb2)	7	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	3	2	-
ht329	14	13	11	15	23	11	I^(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht330	14	13	11	15	24	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht331	14	13	11	15	24	11	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht332	14	13	11	16	22	10	I^(xIIb2)	10	-	-	-	4	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	2	1
ht333	14	13	11	16	23	10	I^(xIIb2)	3	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-
ht334	14	13	11	16	24	11	I^(xIIb2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-
ht335	14	13	12	14	22	10	I^(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht336	14	13	12	14	22	11	I^(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht337	14	13	12	14	23	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht338	14	13	12	16	24	11	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht339	14	14	10	14	22	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht340	14	14	11	14	22	10	I^(xIIb2)	9	-	1	-	-	1	-	-	1	-	-	-	-	-	1	-	-	-	1	-	1	-	-	-	-	-	-	-	1	2	-
ht341	14	14	11	14	23	10	I^(xIIb2)	8	-	-	-	1	-	1	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	1	-	-	1	2	-
ht342	14	14	11	15	22	10	I^(xIIb2)	5	-	2	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht343	14	14	11	15	22	11	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht344	14	14	11	15	23	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht345	14	14	12	14	23	11	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht346	14	14	12	15	23	9	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht347	14	14	12	16	23	10	I^(xIIb2)	2	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht348	14	15	11	14	23	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht349	14	15	12	17	23	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht350	14	15	13	17	23	9	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht351	15	12	11	14	23	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht352	15	12	11	14	26	10	I^(xIIb2)	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht353	15	12	12	15	22	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht354	15	13	11	14	22	10	I^(xIIb2)	4	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	2	-	-
ht355	15	13	11	14	22	11	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht356	15	13	11	14	23	10	I^(xIIb2)	2	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht357	15	13	11	15	23	10	I^(xIIb2)	10	-	-	-	-	-	-	-	2	-	-	1	-	-	-	-	2	-	2	-	-	-	1	-	-	-	1	-	-	1	-
ht358	15	13	11	15	23	11	I^(xIIb2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
ht359	15	13	11	15	24	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht360	15	13	11	16	23	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht361	15	14	11	15	22	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht362	15	15	12	15	23	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht363	16	13	11	14	22	10	I^(xIIb2)	2	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht364	16	13	11	15	22	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht365	16	14	11	14	22	9	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht366	16	14	11	14	22	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht367	16	14	11	14	23	10	I^(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
ht368	12	13	11	16	23	10	IIb2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
ht373	13	13	11	17	23	9	IIb2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
ht374	13	13	11	17	23	10	IIb2	4	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	-	-	-
ht375	13	13	11	17	24	9	IIb2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht376	13	13	12	16	23	10	IIb2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht377	14	13	11	16	24	10	IIb2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht369	13	13	11	15	23	10	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht370	13	13	11	15	23	11	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht371	13	13	11	16	22	10	IIb2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.3. continued

									Shet	Ork	Dur	Wis	Sth	Ptl	Oban	Mpt	Pnt	loM	York	Sow	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chip	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht372	13	13	11	16	23	10	IIb2	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	1	-	-	-	-	-	-	-
ht378	12	12	11	13	24	10	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht379	12	13	11	12	22	9	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht380	12	13	11	13	22	9	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ht381	12	13	11	13	22	10	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht382	12	13	11	13	23	10	E3b	2	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht383	12	13	11	13	24	9	E3b	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	2	-	-
ht384	12	13	11	13	24	10	E3b	12	-	-	-	-	-	-	-	-	2	1	1	2	1	-	-	-	-	1	-	-	1	-	1	-	-	1	-	-	-	1
ht385	12	13	11	13	24	11	E3b	4	-	-	-	-	-	-	-	-	1	-	-	-	-	1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht386	12	13	11	13	24	12	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht387	12	13	11	13	25	9	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht388	12	13	11	13	25	10	E3b	5	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	1	-	-	-	-	-	-	-	1	-	1	-
ht389	12	13	11	14	24	9	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht390	12	13	11	14	24	10	E3b	4	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	2	-	-	1	-	-	-	-	-	-	-
ht391	12	13	11	14	25	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht392	12	13	12	13	23	10	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht393	12	13	12	13	24	10	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht394	12	13	15	13	24	10	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht395	12	14	11	13	24	10	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht396	12	14	11	13	24	11	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht397	14	14	11	16	24	10	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht398	12	12	11	15	24	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht399	13	12	11	14	23	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht400	13	12	11	15	23	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht401	13	12	12	16	24	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht402	14	12	11	13	23	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht403	14	12	11	14	23	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht404	14	12	11	14	24	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht405	14	12	11	15	23	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht406	15	11	11	14	23	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht407	15	12	11	14	22	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht408	15	12	11	14	23	11	J2	4	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	1	
ht409	15	12	11	14	24	10	J2	2	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	
ht410	15	12	11	14	26	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht411	15	12	11	15	23	10	J2	3	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-
ht412	15	12	11	15	23	11	J2	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht413	15	12	11	15	24	10	J2	5	-	-	-	-	-	2	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht414	15	12	11	15	25	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht415	15	12	11	15	24	11	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht416	15	12	11	16	24	10	J2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ht417	15	12	12	16	24	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht418	15	13	11	14	22	10	J2	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht419	16	12	11	14	22	10	J2	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht420	16	12	11	14	23	10	J2	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht421	16	12	11	15	23	9	J2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-
ht422	16	12	11	15	24	9	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht423	16	12	11	15	24	11	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht424	17	12	11	14	23	10	J2	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht425	17	12	11	15	24	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht426	13	12	11	14	23	10	J*(xJ2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht427	15	12	11	14	23	10	J*(xJ2)	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht428	15	12	11	15	24	10	J*(xJ2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-

continued

Appendix. Table A.3. continued

									Shet	Ork	Dur	Wis	Sth	Ptl	Oban	Mpt	Pnt	IoM	York	Sow	Utx	Ldl	Lgf	Rush	Cas	Nor	Hwf	Chip	Fav	Mdh	Dcr	Pnz	Jer	Gue	Bas	GD	Nrw	Lon
ht429	16	12	11	14	23	10	J*(xJ2)	4	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
ht430	16	12	11	14	24	10	J*(xJ2)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht431	17	12	10	14	23	10	J*(xJ2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht432	18	12	11	14	24	11	J*(xJ2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht433	12	12	14	16	23	10	P*(xR1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht434	12	12	15	14	23	10	P*(xR1)	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht435	12	13	12	13	23	10	P*(xR1)	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-
ht436	12	13	14	13	23	10	P*(xR1)	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-
ht437	12	13	10	15	22	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht438	12	13	11	14	22	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ht439	12	14	11	14	21	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht440	12	14	11	15	21	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht441	12	14	11	15	22	10	F*(xIJK)	2	-	-	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht442	12	14	11	15	22	11	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht443	12	14	11	15	23	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht444	12	14	14	14	23	11	F*(xIJK)	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht445	13	13	11	13	24	12	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht446	13	13	11	15	25	11	F*(xIJK)	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht447	13	14	10	15	22	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht448	13	14	11	15	21	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht449	13	14	11	15	22	10	F*(xIJK)	6	-	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-	-	1	-	-	-	-	1	-	-	-	1	1	-
ht450	13	14	11	15	23	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht451	13	14	13	15	21	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht452	15	13	11	14	23	10	F*(xIJK)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht453	12	13	11	14	25	10	K*(xPN3)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht454	12	13	11	15	23	11	K*(xPN3)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht455	12	13	12	13	23	10	K*(xPN3)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht456	12	13	13	13	23	10	K*(xPN3)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht457	12	15	11	15	21	11	K*(xPN3)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht458	13	12	11	14	23	10	K*(xPN3)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht459	12	14	14	14	23	11	N3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht460	12	14	14	14	24	11	N3	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-
ht461	12	14	14	15	23	11	N3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht462	12	14	15	14	23	10	N3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht463	12	13	11	15	23	10	M201	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
ht464	13	15	11	15	22	10	M201	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht465	13	15	11	15	22	11	M201	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ht466	12	13	11	13	24	10	c	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht467	13	15	11	16	21	10	d	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
							Total	2297	63	121	51	88	44	41	42	95	90	62	46	69	84	57	80	76	43	121	59	51	55	80	73	52	82	46	42	190	201	93

Legend on following page

Appendix. Table A.3. continued

Notes. Abbreviations as in Table 2.8, additionally, Lon=London

[a] The numbers 388-391 refer to the names of the 6 microsatellite loci used here (ie DYS388, 393 etc). Microsatellite haplotypes are given in terms of repeat size. The modal haplotypes and their one step neighbours defined by Wilson et al 2001 are highlighted. Blue (pale blue) is the AMH (one step neighbours), green (pale green) is 3.65 (one step neighbours) and tan (yellow) is 2.47 (and one step neighbours)

[b] Haplogroups are named using the YCC (2002) nomenclature, the mutations defining each hg are shown in Figure 2.3

[c] This sample is underived at the EURO1 UEP PCR kit. Time constraints did not allow this sample to be typed for further UEPs

[d] This sample is underived for all of the UEPs analysed in this thesis. Time contraints did not allow this sample to be typed for further UEPs

Appendix. Table A.4. Y-Chromosome Microsatellite Haplotype and UEP Information for the Surnames Studied

	Microsatellite Locus [a]								Count in Population																										
									Burnfather &				Causton & variants											Farrer & variants										Whittock & variants		
Haplotype Number	388	393	392	19	390	391	HG [b]	Total	Barn	Ban	Bairn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Coso	FLLD	Fai	Fara	Fare	Farr	Far	Pha	MCLD	SRB	SPCH	THWT	Wi	Whi	Wiu
ht1	10	12	14	15	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht2	10	13	11	15	24	10	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht3	10	13	13	14	23	11	R1*(xR1a[	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht4	11	13	13	14	23	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht5	11	13	13	15	24	10	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht6	12	11	13	14	23	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht7	12	11	13	14	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht8	12	12	11	14	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht9	12	12	13	13	24	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht10	12	12	13	13	25	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht11	12	12	13	14	23	10	R1*(xR1a[	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht12	12	12	13	14	23	11	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht13	12	12	13	14	24	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht14	12	12	13	14	24	11	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht15	12	12	13	14	25	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht16	12	12	13	14	25	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht17	12	12	13	15	23	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht18	12	12	13	15	23	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht19	12	12	13	15	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht20	12	12	13	16	23	12	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht21	12	12	14	14	22	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht22	12	12	14	14	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht23	12	12	14	14	25	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht24	12	13	11	14	23	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht25	12	13	11	14	24	10	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht26	12	13	11	14	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht27	12	13	11	14	25	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht28	12	13	11	15	23	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht29	12	13	12	13	23	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht30	12	13	12	14	22	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht31	12	13	12	14	22	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht32	12	13	12	14	23	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht33	12	13	12	14	23	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht34	12	13	12	14	24	9	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht35	12	13	12	14	24	10	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht36	12	13	12	14	24	11	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht37	12	13	12	14	25	11	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht38	12	13	12	15	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht39	12	13	12	15	25	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht40	12	13	12	15	26	12	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht41	12	13	12	17	24	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht42	12	13	13	11	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht43	12	13	13	12	23	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht44	12	13	13	12	24	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht45	12	13	13	12	24	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht46	12	13	13	13	23	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht47	12	13	13	13	23	11	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
ht48	12	13	13	13	24	10	R1*(xR1a[	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-
ht49	12	13	13	13	24	11	R1*(xR1a[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht50	12	13	13	13	24	12	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht51	12	13	13	13	25	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht52	12	13	13	13	26	10	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht53	12	13	13	14	21	11	R1*(xR1a[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.4. continued

								Barn	Ban	Bairn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Coso	FLLD	Fai	Fara	Fare	Farr	Fer	Pha	MCLD	SRB	SPCH	THWT	Wt	Wht	Wtu
ht54	12	13	13	14	22	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht55	12	13	13	14	22	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht56	12	13	13	14	23	10	R1*(xR1a1[	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	-	1	4	-	-	-	-	-	-
ht57	12	13	13	14	23	11	R1*(xR1a1[	31	-	-	-	-	-	1	-	-	2	-	1	1	1	-	-	-	2	2	-	-	20	1	-	-	-	-	-
ht58	12	13	13	14	23	12	R1*(xR1a1[	5	-	-	-	-	1	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
ht59	12	13	13	14	23	13	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht60	12	13	13	14	24	9	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht61	12	13	13	14	24	10	R1*(xR1a1[	47	-	-	-	1	-	1	-	1	1	-	-	-	-	2	-	-	2	-	-	-	34	-	3	2	-	-	-
ht62	12	13	13	14	24	11	R1*(xR1a1[	71	-	-	-	-	-	-	-	-	2	-	-	-	-	-	1	-	-	1	-	-	45	9	13	-	-	-	-
ht63	12	13	13	14	24	12	R1*(xR1a1[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht64	12	13	13	14	25	10	R1*(xR1a1[	11	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	9	-	-	-	-	-	-
ht65	12	13	13	14	25	11	R1*(xR1a1[	121	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	118	1	-	-	-	-	-
ht66	12	13	13	14	25	12	R1*(xR1a1[	3	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	2	-	-	-	-	-	-
ht67	12	13	13	14	26	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht68	12	13	13	14	26	11	R1*(xR1a1[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht69	12	13	13	15	23	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht70	12	13	13	15	23	11	R1*(xR1a1[	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht71	12	13	13	15	23	12	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht72	12	13	13	15	24	10	R1*(xR1a1[	7	-	-	-	-	-	-	-	-	-	-	-	-	2	3	-	-	1	-	-	1	-	-	-	-	-	-
ht73	12	13	13	15	24	11	t1*((xR1a1	5	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-
ht74	12	13	13	15	25	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht75	12	13	13	15	25	11	R1*(xR1a1[	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13	-	-	-	-	-	-
ht76	12	13	13	15	26	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht77	12	13	13	16	24	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht78	12	13	13	16	24	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht79	12	13	13	17	24	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht80	12	13	14	13	23	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht81	12	13	14	13	23	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht82	12	13	14	13	24	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht83	12	13	14	14	22	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht84	12	13	14	14	23	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht85	12	13	14	14	23	11	R1*(xR1a1[	6	-	-	-	-	-	-	-	-	-	2	1	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-
ht86	12	13	14	14	24	9	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht87	12	13	14	14	24	10	R1*(xR1a1[	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	2	-	-
ht88	12	13	14	14	24	11	R1*(xR1a1[	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-
ht89	12	13	14	14	24	12	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht90	12	13	14	14	25	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht91	12	13	14	14	25	11	R1*(xR1a1[	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	10	-	-	-	-	-	-
ht92	12	13	14	14	26	11	R1*(xR1a1[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht93	12	13	14	15	23	11	R1*(xR1a1[	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht94	12	13	14	15	24	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht95	12	13	14	15	24	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht96	12	13	14	15	25	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht97	12	13	14	15	25	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht98	12	13	14	16	24	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht99	12	13	14	16	24	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht100	12	13	15	14	24	10	R1*(xR1a1[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht101	12	13	15	14	24	11	R1*(xR1a1[	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht102	12	13	15	14	25	11	R1*(xR1a1[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht103	12	13	15	15	24	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht104	12	13	16	14	23	11	R1*(xR1a1[	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
ht105	12	14	11	16	25	11	R1*(xR1a1[	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht106	12	14	12	14	24	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht107	12	14	12	15	24	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht108	12	14	13	13	24	10	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht109	12	14	13	13	24	11	R1*(xR1a1[	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.4. continued

									Barn	Ban	Bairn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Cosa	FLLD	Fai	Fara	Fare	Farr	Fer	Pha	MCLD	SRB	SPCH	THWT	Wt	Wht	Wtu
ht110	12	14	13	14	22	10	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht111	12	14	13	14	23	10	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht112	12	14	13	14	23	11	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht113	12	14	13	14	23	12	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht114	12	14	13	14	24	10	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht115	12	14	13	14	24	11	R1*(xR1a1	6	-	-	-	2	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-	-	-	-	-	1	1	-	-
ht116	12	14	13	14	24	12	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht117	12	14	13	14	25	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht118	12	14	13	14	25	11	R1*(xR1a1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht119	12	14	13	15	23	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht120	12	14	13	15	24	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht121	12	14	13	15	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht122	12	14	13	16	25	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht123	12	14	14	14	23	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht124	12	14	14	14	24	11	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
ht125	12	14	14	14	25	11	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht126	12	14	14	14	26	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht127	12	14	15	14	25	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht128	12	15	13	14	23	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht129	12	15	13	14	24	10	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht130	12	15	13	14	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht131	12	15	13	14	25	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht132	12	15	13	14	25	11	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht133	12	15	14	15	24	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht134	12	17	13	15	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht135	13	13	9	16	26	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht136	13	13	11	15	23	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht137	13	13	12	14	23	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht138	13	13	12	14	24	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht139	13	13	12	15	23	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht140	13	13	13	14	23	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht141	13	13	13	14	23	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht142	13	13	13	14	23	12	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht143	13	13	13	14	24	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht144	13	13	13	14	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht145	13	13	13	14	25	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht146	13	13	13	15	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht147	13	13	13	16	23	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht148	13	13	13	16	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht149	13	13	14	14	25	12	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht150	13	14	13	14	22	12	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht151	13	14	13	14	24	11	R1*(xR1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht152	13	14	13	15	23	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht153	13	14	13	15	27	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht154	14	13	11	13	22	10	R1*(xR1a1	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht155	14	13	13	14	24	10	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht156	14	13	13	14	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht157	14	13	13	16	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht158	14	14	13	14	23	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht159	17	13	13	14	24	11	R1*(xR1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht160	10	13	11	15	23	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht161	10	13	11	15	24	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht162	10	13	11	15	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht163	10	13	11	16	24	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht164	10	13	11	16	25	10	R1a1	1	0	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.4. continued

									Barn	Ban	Bairn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Coso	FLLD	Fal	Fara	Fare	Farr	Fer	Pha	MCLD	SRB	SPCH	THWT	Wt	Wht	Wtu
ht165	10	13	11	16	26	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht166	10	13	11	17	24	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht167	10	13	11	17	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht168	10	13	12	16	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht169	10	14	11	15	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht170	11	13	11	16	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht171	12	12	11	15	22	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht172	12	12	11	15	23	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht173	12	12	11	15	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht174	12	12	11	16	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht175	12	12	13	14	24	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht176	12	13	11	13	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht177	12	13	11	14	24	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht178	12	13	11	14	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht179	12	13	11	14	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht180	12	13	11	14	26	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht181	12	13	11	15	23	10	R1a1	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-
ht182	12	13	11	15	24	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht183	12	13	11	15	24	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht184	12	13	11	15	25	10	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht185	12	13	11	15	25	11	R1a1	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	3	-	-	-	-	-	-
ht186	12	13	11	15	25	12	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht187	12	13	11	15	26	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht188	12	13	11	15	26	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht189	12	13	11	16	24	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht190	12	13	11	16	23	11	R1a1	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	-	-	-	-	-	-	-	-	-
ht191	12	13	11	16	24	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht192	12	13	11	16	25	10	R1a1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-
ht193	12	13	11	16	25	11	R1a1	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	2	-	-	10	-	-	-	-	-	-
ht194	12	13	11	16	26	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht195	12	13	11	16	26	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht196	12	13	11	16	26	12	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht197	12	13	11	17	24	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht198	12	13	11	17	24	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht199	12	13	11	17	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht200	12	13	11	17	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht201	12	13	11	17	26	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht202	12	13	12	13	23	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht203	12	13	12	15	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht204	12	13	13	14	25	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht205	12	13	13	15	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht206	12	13	14	15	25	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht207	12	14	11	15	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht208	12	14	11	16	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht209	12	14	11	16	25	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht210	12	14	13	13	24	11	R1a1	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht211	12	14	13	14	24	11	R1a1	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht212	13	13	11	15	25	11	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht213	13	13	11	16	24	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht214	13	14	11	14	22	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht215	13	14	11	15	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht216	14	13	11	15	25	11	R1a1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht217	14	14	11	16	25	10	R1a1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht218	10	14	12	16	23	11	P(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht219	11	15	12	14	23	10	P(xI1b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.4. continued

									Barn	Ban	Bairn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Coso	FLLD	Fai	Fara	Fare	Farr	Fer	Pha	MCLD	SRB	SPCH	THWT	Wt	Wht	Wtu
ht220	12	13	11	13	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht221	12	13	11	14	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht222	12	13	11	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht223	12	13	11	15	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht224	12	13	11	15	25	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht225	12	13	11	15	25	11	I*(xIIb2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
ht226	12	13	13	14	24	11	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
ht227	12	13	13	14	25	11	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht228	12	13	14	13	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht229	12	14	11	13	26	9	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht230	12	14	11	14	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht231	12	14	11	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht232	12	14	11	15	21	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht233	12	14	11	15	25	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht234	12	14	11	15	25	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht235	12	14	11	17	26	11	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht236	12	14	12	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht237	12	14	12	15	24	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht238	12	14	13	14	24	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht239	13	12	11	15	25	10	I*(xIIb2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
ht240	13	12	12	15	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht241	*13*	*13*	*11*	*14*	*22*	*10*	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht242	13	13	11	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht243	13	13	11	14	25	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht244	13	13	11	15	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht245	13	13	11	15	24	11	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht246	13	13	11	15	25	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht247	13	13	11	15	25	11	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht248	13	13	11	16	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht249	13	13	11	16	*24*	*10*	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht250	13	13	11	16	24	11	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
ht251	13	13	11	17	25	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht252	13	13	11	17	25	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht253	13	13	11	17	25	13	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht254	13	13	11	17	26	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht255	13	13	12	15	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht256	13	13	12	15	23	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht257	13	13	12	15	24	10	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht258	13	13	12	15	24	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht259	13	13	12	16	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht260	13	13	12	16	24	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht261	13	13	13	16	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht262	13	14	11	15	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht263	13	14	11	15	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht264	13	14	11	15	26	12	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht265	13	14	11	16	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht266	13	14	11	16	25	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht267	13	14	11	16	26	11	I*(xIIb2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-
ht268	13	14	11	17	24	11	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
ht269	13	14	11	17	25	11	I*(xIIb2)	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-
ht270	13	14	11	17	25	12	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht271	13	14	11	17	26	10	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht272	13	14	11	17	26	11	I*(xIIb2)	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	1	-	-	-	-	-	-	-	-
ht273	13	14	12	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht274	13	14	12	15	21	10	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht275	13	14	12	15	22	10	I*(xIIb2)	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-

continued

Appendix. Table A.4. continued

									Barn	Ban	Balrn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Cosa	FLLD	Fai	Fara	Fare	Farr	Fer	Pha	MCLD	SRB	SPCH	THWT	Wt	Wht	Wtu
ht276	13	14	12	15	23	9	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht277	13	14	12	15	23	10	I*(x11b2)	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	3	-	1
ht278	13	14	12	15	23	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht279	13	14	12	15	24	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht280	13	14	12	16	22	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht281	13	14	12	16	23	10	I*(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht282	13	14	12	16	23	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht283	13	14	12	16	24	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht284	13	14	12	17	23	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht285	13	14	13	15	22	10	I*(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht286	13	14	13	15	23	10	I*(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht287	13	14	13	16	23	10	I*(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht288	13	14	13	16	24	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht289	13	15	12	14	24	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht290	13	15	12	15	22	10	I*(x11b2)	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht291	13	15	12	15	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht292	13	15	12	15	23	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht293	13	15	12	15	23	13	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht294	13	15	12	15	24	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht295	13	15	12	15	25	10	I*(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
ht296	13	15	12	15	24	11	I*(x11b2)	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht297	13	15	12	16	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht298	13	15	13	15	22	10	I*(x11b2)	3	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht299	13	15	13	15	24	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht300	13	15	13	16	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht301	13	15	13	17	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht302	13	15	14	15	23	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht303	13	16	12	15	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht304	14	9	11	14	22	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht305	14	11	11	14	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht306	14	12	11	14	22	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht307	14	12	11	14	22	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht308	14	12	11	14	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht309	14	12	11	15	22	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht310	14	12	11	15	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht311	14	12	12	14	22	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht312	14	12	12	15	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht313	14	13	10	14	22	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht314	14	13	10	14	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht315	14	13	11	13	22	10	I*(x11b2)	2	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht316	14	13	11	13	25	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht317	14	13	11	14	21	9	I*(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht318	14	13	11	14	21	10	I*(x11b2)	2	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht319	14	13	11	14	22	10	I*(x11b2)	15	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	13	-	-	-	-	-	-
ht320	14	13	11	14	22	11	I*(x11b2)	4	7	1	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht321	14	13	11	14	22	12	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht322	14	13	11	14	23	10	I*(x11b2)	12	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	10	-	-	-	-	-	-
ht323	14	13	11	14	23	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht324	14	13	11	14	24	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht325	14	13	11	15	22	9	I*(x11b2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht326	14	13	11	15	22	10	I*(x11b2)	3	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-
ht327	14	13	11	15	22	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht328	14	13	11	15	23	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht329	14	13	11	15	23	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht330	14	13	11	15	24	10	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht331	14	13	11	15	24	11	I*(x11b2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.4. continued

									Barn	Ban	Bairn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Casn	FLLD	Fai	Fara	Fare	Farr	Fer	Pha	MCLD	SRB	SPCH	THWT	Wt	Wht	Wtu
ht332	14	13	11	16	22	10	I*(xIIb2)	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-
ht333	14	13	11	16	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht334	14	13	11	16	24	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht335	14	13	12	14	22	10	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht336	14	13	12	14	22	11	I*(xIIb2)	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht337	14	13	12	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht338	14	13	12	16	24	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht339	14	14	10	14	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht340	14	14	11	14	22	10	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
ht341	14	14	11	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht342	14	14	11	15	22	10	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht343	14	14	11	15	22	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht344	14	14	11	15	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht345	14	14	12	14	23	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht346	14	14	12	15	23	9	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht347	14	14	12	16	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht348	14	15	11	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht349	14	15	12	17	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht350	14	15	13	17	23	9	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht351	15	12	11	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht352	15	12	11	14	26	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht353	15	12	12	15	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht354	15	13	11	14	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht355	15	13	11	14	22	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht356	15	13	11	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht357	15	13	11	15	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht358	15	13	11	15	23	11	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht359	15	13	11	15	24	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht360	15	13	11	16	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht361	15	14	11	15	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht362	15	15	12	15	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht363	16	13	11	14	22	10	I*(xIIb2)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht364	16	13	11	15	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht365	16	14	11	14	22	9	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht366	16	14	11	14	22	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht367	16	14	11	14	23	10	I*(xIIb2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht368	12	13	11	16	23	10	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht369	13	13	11	15	23	10	IIb2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht370	13	13	11	15	23	11	IIb2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
ht371	13	13	11	16	22	10	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht372	13	13	11	16	23	10	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht373	13	13	11	17	23	9	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht374	13	13	11	17	23	10	IIb2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht375	13	13	11	17	24	9	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht376	13	13	12	16	23	10	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht377	14	13	11	16	24	10	IIb2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht378	12	12	11	13	24	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht379	12	13	11	12	22	9	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht380	12	13	11	13	22	9	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht381	12	13	11	13	22	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht382	12	13	11	13	23	10	E3b	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht383	12	13	11	13	24	9	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht384	12	13	11	13	24	10	E3b	7	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-
ht385	12	13	11	13	24	11	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.4. continued

									Barn	Ban	Barn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Coso	FLLD	Fai	Fara	Fare	Farr	Fer	Pha	MCLD	SRB	SPCH	THWT	Wt	Wht	Wtu
ht386	12	13	11	13	24	12	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht387	12	13	11	13	25	9	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht388	12	13	11	13	25	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht389	12	13	11	14	24	9	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht390	12	13	11	14	24	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht391	12	13	11	14	25	10	E3b	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht392	12	13	12	13	23	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht393	12	13	12	13	24	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht394	12	13	15	13	24	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht395	12	14	11	13	24	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht396	12	14	11	13	24	11	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht397	14	14	11	16	24	10	E3b	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht398	12	12	11	15	24	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht399	13	12	11	14	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht400	13	12	11	15	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht401	13	12	12	16	24	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht402	14	12	11	13	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht403	14	12	11	14	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht404	14	12	11	14	24	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht405	14	12	11	15	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht406	15	11	11	14	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht407	15	12	11	14	22	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht408	15	12	11	14	23	11	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht409	15	12	11	14	24	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht410	15	12	11	14	26	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht411	15	12	11	15	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht412	15	12	11	15	23	11	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht413	15	12	11	15	24	10	J2	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht414	15	12	11	15	25	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht415	15	12	11	15	24	11	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht416	15	12	11	16	24	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht417	15	12	12	16	24	10	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
ht418	15	13	11	14	22	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht419	16	12	11	14	22	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht420	16	12	11	14	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht421	16	12	11	15	23	9	J2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht422	16	12	11	15	24	9	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht423	16	12	11	15	24	11	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht424	17	12	11	14	23	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht425	17	12	11	15	24	10	J2	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht426	13	12	11	14	23	10	J*(xJ2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht427	15	12	11	14	23	10	J*(xJ2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht428	15	12	11	15	24	10	J*(xJ2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht429	16	12	11	14	23	10	J*(xJ2)	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht430	16	12	11	14	24	10	J*(xJ2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht431	17	12	10	14	23	10	J*(xJ2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht432	18	12	11	14	24	11	J*(xJ2)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht433	12	13	14	13	23	10	P*(xR1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht434	12	12	14	16	23	10	P*(xR1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht435	12	12	15	14	23	10	P*(xR1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht436	12	13	12	13	23	10	P*(xR1)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht437	12	13	10	15	22	10	F*(xUK)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

continued

Appendix. Table A.4. continued

									Barn	Ban	Bairn	Cast	Cas	Caw	Cos	Cau	Cor	Cost	Caus	Core	Coso	FLLD	Fai	Fara	Fare	Farr	Fer	Pha	MCLD	SRB	SPCH	THWT	Wt	Wht	Wtu
ht438	12	13	11	14	22	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht439	12	14	11	14	21	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht440	12	14	11	15	21	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht441	12	14	11	15	22	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht442	12	14	11	15	22	11	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht443	12	14	11	15	23	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht444	12	14	14	14	23	11	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht445	13	13	11	13	24	12	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht446	13	13	11	15	25	11	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht447	13	14	10	15	22	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht448	13	14	11	15	21	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht449	13	14	11	15	22	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht450	13	14	11	15	23	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht451	13	14	13	15	21	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht452	15	13	11	14	23	10	F*(xDYS)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
								0																											
ht453	12	13	11	14	25	10	K*(xPN3)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht454	12	13	11	15	23	11	K*(xPN3)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht455	12	13	12	13	23	10	K*(xPN3)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht456	12	13	13	13	23	10	K*(xPN3)	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht457	12	15	11	15	21	11	K*(xPN3)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
ht458	13	12	11	14	23	10	K*(xPN3)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
								0																											
ht459	12	14	14	14	23	11	N3	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht460	12	14	14	14	24	11	N3	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht461	12	14	14	15	23	11	N3	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht462	12	14	15	14	23	10	N3	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
								0																											
ht463	12	13	11	15	23	10	M201	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht464	13	15	11	15	22	10	M201	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ht465	13	15	11	15	22	11	M201	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
								0																											
ht467	12	13	11	13	24	10	c	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
ht468	13	15	11	16	21	10	d	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
							Total	549	3	2	6	13	7	5	1	4	7	5	5	1	6	8	8	7	25	8	1	1	367	12	18	8	15	5	1

Notes. Abbreviations as Table 3.3, additionally: Barn=Barnfather, Ban-Banfather, Bairn=Bairnsfather; Cast=Caston, Cas=Cason, Caw=Cawston, Cos=Costen, Cau=Causon, Cor=Corston, Cost=Costin, Caus=Causton, Core=Corsten, Coso=Coston; Fai=Fairer, Fara=Farrar, Fare=Farrer; Farr=Farrow, Fer=Ferrer, Pha=Pharoah; Wt=Whittock, Wht=Whytock, Wtu=Whittuck. All haplotypes listed in Appendix Table A.3 are included in this Table for comparative purposes; haplotype numbers are also the same as those in Appendix Table A.3

a The numbers 388-391 refer to the names of the 6 microsatellite loci used here (ie DYS388, 393 etc). Microsatellite haplotypes are given in terms of repeat size. The modal haplotypes and their one step neighbours defined by Wilson et al 2001 are highlighted. Blue (pale blue) is the AMH (one step neighbours), green (pale green) is 3.65 (one step neighbours) and tan (yellow) is 2.47 (and one step neighbours)

b Haplogroups are named using the YCC (2002) nomenclature, the mutations defining each hg are shown in Figure 2.3

c This sample is underived at the EURO1 UEP PCR kit. Time constraints did not allow this sample to be typed for further UEPs

d This sample is underived for all of the UEPs analysed in this thesis. Time contraints did not allow this sample to be typed for further UEPs

Appendix. Table A.5. mtDNA HVSI Sequence Data for the London Population

Sequence Number	*Hg*	*HVSI Sequence from 16,040-16,399 (less 16,000)*										*Count (and frequency)*
1	CRS	0	-	-	-	-	-	-	-	-	-	19 (0.153)
2	T	69	126	145	172	294	296	324	-	-	-	1 (0.008)
3	J	69	126	213	-	-	-	-	-	-	-	1 (0.008)
4	H	80	189	356	-	-	-	-	-	-	-	1 (0.008)
5	H	92	140	311	-	-	-	-	-	-	-	1 (0.008)
6	U5*	92	265	270	292	362	-	-	-	-	-	1 (0.008)
7	L1b	93	126	187	189	223	264	270	278	311	318T	1 (0.008)
8	K	93	192	224	311	318T	-	-	-	-	-	1 (0.008)
9	H	93	221	-	-	-	-	-	-	-	-	1 (0.008)
10	W	93	223	292	-	-	-	-	-	-	-	1 (0.008)
11	K	93	224	311	-	-	-	-	-	-	-	3 (0.024)
12	T	111	126	294	304	-	-	-	-	-	-	1 (0.008)
13	U4	111	140	356	-	-	-	-	-	-	-	1 (0.008)
14	+	114	224	270	-	-	-	-	-	-	-	1 (0.008)
15	+	114	263	-	-	-	-	-	-	-	-	1 (0.008)
16	H	114	-	-	-	-	-	-	-	-	-	1 (0.008)
17	U5a1*	114A	192	256	270	294	-	-	-	-	-	1 (0.008)
18	+	114A	263	-	-	-	-	-	-	-	-	1 (0.008)
19	H	118Ains	239G	-	-	-	-	-	-	-	-	1 (0.008)
20	H	124	-	-	-	-	-	-	-	-	-	1 (0.008)
21	JT	126	-	-	-	-	-	-	-	-	-	4 (0.032)
22	+	126	145	162C	192	222	261	-	-	-	-	1 (0.008)
23	T	126	163	186	189	294	-	-	-	-	-	1 (0.008)
24	J*	126	180DEL	183C	189	207	-	-	-	-	-	1 (0.008)
25	+	126	182C	183C	-	-	-	-	-	-	-	1 (0.008)
26	J2	126	193	278	372G	-	-	-	-	-	-	1 (0.008)
27	J2	126	193	-	-	-	-	-	-	-	-	1 (0.008)
28	T	126	278	294	296	304	360	-	-	-	-	1 (0.008)
29	T	126	294	296	304	-	-	-	-	-	-	1 (0.008)
30	T	126	294	296	324	-	-	-	-	-	-	1 (0.008)
31	T	126	294	-	-	-	-	-	-	-	-	1 (0.008)
32	JT	126	390a	-	-	-	-	-	-	-	-	1 (0.008)

continued

Appendix. Table A.5. continued

Sequence Number	Hg	HVSI Sequence from 16,040-16,399 (less 16,000)										Count (and frequency)
33	I	129	172	223	311	335N	391	-	-	-	-	1 (0.008)
34	I	129	218	223	263	-	-	-	-	-	-	1 (0.008)
35	+	129	223	360	-	-	-	-	-	-	-	1 (0.008)
36	U5*	147	183C	189	270	-	-	-	-	-	-	1 (0.008)
37	+	153	288	360	-	-	-	-	-	-	-	1 (0.008)
38	H	162	172	209	-	-	-	-	-	-	-	1 (0.008)
39	U5a1*	162	189	234	256	270	362	-	-	-	-	1 (0.008)
40	H	162	-	-	-	-	-	-	-	-	-	2 (0.016)
41	+	172	189	194C	-	-	-	-	-	-	-	1 (0.008)
42	H	172	-	-	-	-	-	-	-	-	-	2 (0.016)
43	U5b	174	189	192	270	311	-	-	-	-	-	1 (0.008)
44	U4	179	284C	356	-	-	-	-	-	-	-	1 (0.008)
45	+	182C	183C	189	234	319	324	-	-	-	-	1 (0.008)
46	K	182C	183C	224	311	-	-	-	-	-	-	1 (0.008)
47	+	183C	189	172	223	278	-	-	-	-	-	1 (0.008)
48	H	183C	189	356	362	-	-	-	-	-	-	1 (0.008)
49	+	183C	189	-	-	-	-	-	-	-	-	2 (0.016)
50	U5b*	189	192	270	398	-	-	-	-	-	-	1 (0.008)
51	U4	189	356	362	-	-	-	-	-	-	-	1 (0.008)
52	+	189	209	239	352	-	-	-	-	-	-	1 (0.008)
53	X	189	223	278	-	-	-	-	-	-	-	1 (0.008)
54	U5b	189	270	300	-	-	-	-	-	-	-	1 (0.008)
55	+	189	-	-	-	-	-	-	-	-	-	2 (0.016)
56	L3b	189G	223	274	278	294	362	-	-	-	-	1 (0.008)
57	U5a1*	192	256	270	320	399	-	-	-	-	-	1 (0.008)
58	+	193	219	360	-	-	-	-	-	-	-	1 (0.008)
59	+	193	219	362	-	-	-	-	-	-	-	1 (0.008)
60	H	212	-	-	-	-	-	-	-	-	-	1 (0.008)
61	H	218	-	-	-	-	-	-	-	-	-	1 (0.008)
62	H	221	291	-	-	-	-	-	-	-	-	1 (0.008)
63	C	223	249	295	298	311	325	327	-	-	-	1 (0.008)
64	W	223	292	-	-	-	-	-	-	-	-	1 (0.008)

continued

Appendix. Table A.5. continued

Sequence Number	*Hg*	*HVSI Sequence from 16,040-16,399 (less 16,000)*										*Count (and frequency)*
65	K	224	-	-	-	-	-	-	-	-	-	1 (0.008)
66	K	224	245	311	-	-	-	-	-	-	-	1 (0.008)
67	K	224	311	-	-	-	-	-	-	-	-	4 (0.032)
68	H	234	-	-	-	-	-	-	-	-	-	1 (0.008)
69	U5*	239	270	-	-	-	-	-	-	-	-	1 (0.008)
70	H	239G	-	-	-	-	-	-	-	-	-	1 (0.008)
71	U5a1	256	270	399	-	-	-	-	-	-	-	1 (0.008)
72	+	261	304	-	-	-	-	-	-	-	-	1 (0.008)
73	H	263	-	-	-	-	-	-	-	-	-	1 (0.008)
74	U5*	270	-	-	-	-	-	-	-	-	-	1 (0.008)
75	+	278	360	-	-	-	-	-	-	-	-	1 (0.008)
76	H	278	-	-	-	-	-	-	-	-	-	1 (0.008)
77	H	286	-	-	-	-	-	-	-	-	-	1 (0.008)
78	H	287	311	-	-	-	-	-	-	-	-	1 (0.008)
79	H	291	-	-	-	-	-	-	-	-	-	1 (0.008)
80	V	298	311	-	-	-	-	-	-	-	-	1 (0.008)
81	V	298	-	-	-	-	-	-	-	-	-	3 (0.024)
82	H	304	-	-	-	-	-	-	-	-	-	4 (0.032)
83	H	311	-	-	-	-	-	-	-	-	-	1 (0.008)
84	+	342	-	-	-	-	-	-	-	-	-	1 (0.008)
85	H	354	-	-	-	-	-	-	-	-	-	1 (0.008)
86	U4	356	362	-	-	-	-	-	-	-	-	1 (0.008)
87	+	357	360	-	-	-	-	-	-	-	-	1 (0.008)
88	+	360	-	-	-	-	-	-	-	-	-	1 (0.008)
89	H	362	-	-	-	-	-	-	-	-	-	1 (0.008)
											Total	124

Notes: Hgs have been assigned on the basis of HVSI sequence information only. Where the sequence information could not unambiguously assign a sequence to a hg, it was left undesignated (indicated by "+"). Time constraints did not allow RFLPs to be assayed. Sequences that were not assigned to hgs were not used in analyses that required hg information, and were only used where HVSI sequence information was required.

Appendix. Table A.6. RFLP Screening Reuslts and Hg Designations for Lemba, Bantu and Yemen-Sena Samples Not Assigned to a Hg Using HVSI Sequence Data

Sequence Number	*HVSI Sequence (16,040-16,399)*	*RFLP Results* [a]	*RFLP-based Haplogroup*
48[b]	93 223 278 362	+10084 TaqI.	L3b
61	114 189 192 223 293T 311 316	-10871MnlI, -2349 MboI, -10084 TaqI, +8616 MboI.	L3* (non-L3e, non-L3b, non-L3d).
78[b]	126 153 233C 257 294 325	-13366 BamHI, -111718 HaeIII, +10871 MnlI, -12308 HinfI.	N* (non-T, non-HV, non-U, non-K).
117[b]	129 172 173 188a 223 256 278 293 294 311 360 368	+3592 HpaI, +12810 RsaI.	L1c
211[b]	183c 189 223 278	-3592 HpaI, +14465 AccI, +10871 MnlI	X
216[b,c]	185 223 327	+2349 MboI.	L3e1a
224	189 223 270Del 278	-2349 MboI, -10084 TaqI, +8616 MboI.	L3* (non-U, non-X, non-L1, non-L2, non-L3e, non-L3b, non-L3d).
225	189 223 278	-3592 HpaI, -14465 AccI, -10871 MnlI,	L3* (non-X, non-L1, non-L2, non-L3e, non-L3b, non-L3d).
240[b]	223 239 323Del	+2349 MboI.	L3e1
255	223 327	+2349 MboI.	L3e1

[a] RFLP analysis was kindly performed by the Torroni lab on samples thst could not be unambiguosly assigned to a hg based on HVSI sequence data alone.

[b] >1 individuals with this sequence appear in the dataset, due to DNA limitations only one DNA sample was sent to the Torroni lab for RFLP analysis

[c] This sequence was assigned to hg L3e1 by the Torroni lab, but has been re-classified as L3e1a here based on the recent study of Salas *et al.* (2002)

Appendix. Table A.7. mtDNA HVSI Sequence Data for the Populations Studied and Comparative Populations

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)												Count (and Frequency) in the Studied Populations and Comparison Populations							
														Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
1	H	0												-	-	-	-	-	-	-	8 (0.103)
2	R*	0												-	-	-	-	-	-	2 (0.043)	-
3	U*	0												-	-	-	4 (0.071)	-	-	-	-
4	HV1	67	183	260	327A									-	-	-	-	-	1 (0.014)	-	-
5	HV1	67	183	327A										-	-	-	2 (0.036)	-	-	-	-
6	HV1	67	183C	189	197	360								-	-	-	-	-	-	-	1 (0.013)
7	HV1	67	183C	189										-	-	-	-	-	-	-	2 (0.026)
8	HV1	67	274											-	-	-	-	13 (0.2)	-	-	1 (0.013)
9	HV1	67	278	362										-	-	-	-	-	3 (0.041)	-	-
10	J*	69	93	126	261	274	319	355						-	-	-	-	-	-	-	1 (0.013)
11	J*	69	93	126	261	274	355							-	-	-	-	-	-	-	1 (0.013)
12	J1b	69	126	136	145	221	261							-	-	-	-	4 (0.062)	-	-	-
13	J1b	69	126	136	145	261								-	-	-	-	1 (0.015)	-	-	-
14	J1b1	69	126	145	185	222	261							-	-	-	1 (0.018)	-	-	-	-
15	J1b	69	126	145	222	261								-	-	-	-	6 (0.092)	-	-	-
16	J1	69	126	145	261	399								-	-	-	-	-	1 (0.014)	-	-
17	J*	69	126	192										-	-	-	-	1 (0.015)	-	-	-
18	J2	69	126	193	212	300	309							-	-	-	1 (0.018)	-	-	-	-
19	J2	69	126	193	300	309								-	-	-	2 (0.036)	-	1 (0.014)	-	-
20	J*	69	126	214	231									-	-	-	-	4 (0.062)	-	-	-
21	J*	69	126	261	274	355								-	-	-	-	-	-	-	1 (0.013)
22	J*	69	126	261	297									-	-	-	2 (0.036)	-	-	-	-
23	J*	69	126	390										-	-	-	-	1 (0.015)	-	-	-
24	R2	71	93	265	274									-	-	-	1 (0.018)	-	-	-	-
25	R2/N*	71	188	223	362									-	-	-	-	1 (0.015)	-	-	-
26	R2/N*	71	188	223										-	-	-	-	5 (0.077)	-	-	-
27	L3*	75	153	223	319									-	-	-	-	-	-	1 (0.022)	-
28	L1c	78	129	183C	184G	189	223	265C	286G	294	311	320	360	-	-	-	1 (0.018)	-	-	-	-
29	H	80	183C	189	356	360								-	-	-	-	-	-	-	2 (0.026)
30	U*	86	119											-	-	-	-	3 (0.046)	-	-	-
31	M*	86	148	223	259	278	319	399						-	-	-	1 (0.018)	-	-	-	-

continued

Appendix. Table A.7. continued

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)														Count (and Frequency) in the Studied Populations and Comparison Populations							
																Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
32	L2	86	182C	183C	189	223	278	294	309	390						-	-	-	-	-	-	1 (0.022)	-
33	M1	92	129	183C	189	223	249	287	311	359						-	-	-	-	-	1 (0.014)	-	-
34	L1a1a	93	129	148	168	172	187	188G	189	223	230	278	293	311	320	1 (0.008)	-	-	-	-	-	-	-
35	L1d	93	129	187	189	223	230	239	243	311	325					-	1 (0.015)	-	-	-	-	-	-
36	M1a	93	129	189	213	223	249	311	350							-	-	-	-	-	-	1 (0.022)	-
37	M1a	93	129	189	213	223	249	311	359							-	-	-	-	-	-	1 (0.022)	-
38	M1a	93	129	189	223	249	311	359								-	-	-	-	-	-	1 (0.022)	-
39	L1c	93	129	183C	189	223	278	294	311	360						-	-	-	1 (0.018)	-	-	-	-
40	L1a	93	148	172	187	188G	189	223	230	311	320					2 (0.017)	1 (0.015)	-	-	-	-	-	-
41	L2a1b	93	169	182C	183C	189	223	278A	290	309	390					-	1 (0.015)	-	-	-	-	-	-
42	L3a1a	93	192	209	223	292	311									-	-	-	-	-	1 (0.014)	-	-
43	L3a1	93	209	223	266	290	311									-	-	-	-	-	1 (0.014)	-	-
44	L3a1	93	209	223	292	311										-	-	-	-	3 (0.046)	-	-	-
45	L2c2	93	223	264	265	278	311	390								3 (0.025)	-	-	-	-	-	-	-
46	L2	93	223	278	294	309	368	390								-	-	-	1 (0.018)	-	-	-	-
47	L2a	93	223	278	294	311	390									1 (0.008)	-	-	-	-	-	-	-
48[d]	L3b	93	223	278	362											2 (0.017)	-	-	-	-	-	-	-
49	K	93	224	311												-	-	-	-	-	1 (0.014)	-	5 (0.064)
50	L3a2	93G	223	287A	293T	311	355	362	399							-	-	-	-	-	1 (0.014)	-	-
51	L2a1b	94	182C	183C	223	278	290	294	309	390						1 (0.008)	-	-	-	-	-	-	-
52	L2*	95	148	183C	189	223	224	278	390							1 (0.008)	-	-	-	-	-	-	-
53	L2a1b	95	182C	183C	189	223	278	290	294	309	390					1 (0.008)	-	-	-	-	-	-	-
54	L2a1b	95G	182C	183C	189	192	223	278	290	294	309	390				1 (0.008)	-	-	-	-	-	-	-
55	U5a1a	107G	256	270	293											-	-	1 (0.034)	-	-	-	-	-
56	F	108	129	162	172	304										-	-	-	1 (0.018)	-	-	-	-
57	L3*	111	184	223	304											-	-	-	-	-	2 (0.027)	-	-
58	M*	111	223	235	362											-	-	-	-	-	-	-	1 (0.013)
59	L2c2	111	223	264	278	311	390									1 (0.008)	-	-	-	-	-	-	-
60	L3b1	114	124	223	278	362										-	1 (0.015)	-	-	-	-	-	-
61[d]	L3*	114	189	192	223	293T	311	316								1 (0.008)	-	-	-	-	-	-	-
62	L2b	114A	129	145	213	223	278	311	390							-	-	-	-	-	1 (0.014)	-	-

continued

Appendix. Table A.7. continued

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)											Count (and Frequency) in the Studied Populations and Comparison Populations							
													Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
63	L2b	114A	129	145	213	223	278	390					-	-	-	-	-	1 (0.014)	-	-
64	L2b	114A	129	145	213	223	278						-	-	-	-	-	-	1 (0.022)	-
65	L2*	114A	129	213	223	278	354	390					1 (0.008)	1 (0.015)	-	-	-	-	-	-
66	L2b	114A	213	223	278								-	-	-	-	-	-	1 (0.022)	-
67	L3b1	124	172	223	278	362							1 (0.008)	-	-	-	-	-	-	-
68	L3b2	124	183C	189	223	278	304	311					2 (0.017)	-	-	-	-	-	-	-
69	L3b1	124	223	234	278	362							-	1 (0.015)	-	-	-	-	-	-
70	L3b2	124	223	278	311	362							-	1 (0.015)	-	-	-	-	4 (0.087)	-
71	L3b	124	223	278	362								-	-	-	1 (0.018)	-	-	-	-
72	L3d	124	223	311									1 (0.008)	-	-	-	-	-	-	-
73	L3d	124	223	319									-	-	-	1 (0.018)	-	-	-	-
74	L3d1	124	223	319									8 (0.067)	-	2 (0.069)	-	-	1 (0.014)	-	-
74	T2	126	129	294	296	304							-	-	-	-	-	-	-	2 (0.026)
75	T1	126	136	163	186	189	294						-	-	-	-	-	1 (0.014)	-	-
76	T*	126	146	189	292	294	296						-	-	-	1 (0.018)	-	-	-	-
77	N*	126	153	233C	257	294	325						-	-	2 (0.069)	-	-	-	-	-
78	T1	126	163	186	189	294							-	-	-	2 (0.036)	-	2 (0.027)	-	2 (0.026)
79	pre-HV	126	172	184A	362								-	-	-	-	-	1 (0.014)	-	-
80	L1b	126	186	189	288	292	294	296	311				-	-	-	-	-	1 (0.014)	-	-
81	L1b	126	187	189	223	264	270	278	289	293	311	362	-	-	-	-	-	1 (0.014)	-	-
82	L1b	126	187	189	223	264	270	278	289	293	311		-	-	-	-	-	3 (0.041)	-	-
83	L1b	126	187	189	223	264	270	278	311				1 (0.008)	-	-	-	-	-	-	-
84	U*	126	231	318C									-	-	-	1 (0.018)	-	-	-	-
85	pre-HV	126	234	355	362								-	-	-	1 (0.018)	-	-	-	-
86	T*	126	294	295									-	-	-	-	-	-	-	1 (0.013)
87	T2	126	294	296	304	362							-	-	-	-	-	-	-	1 (0.013)
88	T*	126	294	296	320								-	-	-	-	2 (0.031)	-	-	-
89	T*	126	294	296									-	-	-	-	-	1 (0.014)	-	-
90	pre-HV	126	304	311	362								-	-	-	-	1 (0.015)	-	-	-
91	pre-HV	126	304	362									-	-	-	1 (0.018)	8 (0.123)	-	-	-
92	pre-HV	126	305T	362									-	-	-	-	-	1 (0.014)	6 (0.130)	-

continued

Appendix. Table A.7. continued

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)													Count (and Frequency) in the Studied Populations and Comparison Populations							
															Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
93	pre-HV	126	311	362											-	-	-	-	-	-	1 (0.022)	-
94	pre-HV	126	355	362											-	-	-	1 (0.018)	1 (0.015)	3 (0.041)	-	-
95	pre-HV	126	362												-	-	-	1 (0.018)	-	1 (0.014)	-	2 (0.026)
96	pre-JT	126													-	-	-	-	-	1 (0.014)	-	-
97	L2*	129	145	187	189	212	223	230	243	311	390				-	1 (0.015)	-	-	-	-	-	-
98	L1c2	129	145	187	189	213	223	234	265C	278	286G	294	311	360	1 (0.008)	-	-	-	-	-	-	-
99	L1c2	129	145	187	189	213	223	234	265C	278	286G	294	311		1 (0.008)	-	-	-	-	-	-	-
100	L1c2	129	145	187	188G	189	213	223	234	265C	278	286G	311	360	1 (0.008)	-	-	-	-	-	-	-
101	L1a	129	148	165	168	172	187	188G	189	223	230	311	320		-	-	-	2 (0.036)	-	-	-	-
102	L1e	129	148	166	183del	187	189	223	278	311	355	362			-	-	-	-	-	-	3 (0.065)	-
103	L1e	129	148	166	183del	188A	189	223	278	311	355	362			-	-	-	-	-	-	1 (0.022)	-
104	L1a	129	148	168	172	187	188G	189	223	230	278	293	311	320	4 (0.034)	10 (0.149)	5 (0.172)	-	-	-	-	-
105	L1a	129	148	168	172	187	188G	189	223	230	293	311			-	-	-	-	-	2 (0.027)	1 (0.022)	-
106	L1a2	129	148	169	172	187	188A	189	223	230	239G	261	278	311	320	-	1 (0.015)	-	-	-	-	-
107	L1a	129	148	172	187	188G	189	223	230	311	320				-	-	-	-	-	1 (0.014)	-	-
108	I	129	148	223	391										-	-	-	-	-	-	-	1 (0.013)
109	M1*	129	154	189	223	249	311								-	-	-	-	-	-	1 (0.022)	-
110	L1d	129	162	187	189	212	223	230	243	311	390				2 (0.017)	-	-	-	-	-	-	-
111	L1c	129	163	187	189	209	223	278	293	294	311	360			1 (0.008)	1 (0.015)	-	-	-	-	-	-
112	L1d	129	166C	186	187	189G	212	223	230	243	311	390T			-	1 (0.015)	-	-	-	-	-	-
113	L1	129	166	187	189	209	213	215	223	256	278	298	311		-	-	-	-	-	1 (0.014)	-	-
114	L1	129	166	187	189	209	213	223	256	266	278	298	311		-	-	-	-	-	1 (0.014)	-	-
115	V	129	166	192	255	298	311								-	-	-	-	-	1 (0.014)	-	-
116	L1c1	129	172	173	188A	189	223	256	278	293	294	311	360	368	2 (0.017)	-	-	-	-	-	-	-
117	L3*	129	172	174	192	218	223	256A	311A						1 (0.008)	-	-	-	-	-	-	-
118	L1d	129	179	187	189	223	230	243	290	311					1 (0.008)	-	-	-	-	-	-	-
119	M1	129	182C	183C	189	223	240	249	311						-	-	-	-	-	1 (0.014)	-	-
120	M1	129	182C	183C	189	223	249	294	311	359					-	-	1 (0.034)	-	-	-	-	-
121	M1a	129	182C	183C	189	223	249	311	359						-	-	-	1 (0.018)	-	-	-	-
122	L2/L3*	129	182DEL	189A	215	223	278	294	311						1 (0.008)	-	-	-	-	-	-	-
123	U2	129C	182C	183C	189	260	356	362							-	-	-	1 (0.018)	-	-	-	-

continued

Appendix. Table A.7. continued

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)												Count (and Frequency) in the Studied Populations and Comparison Populations							
														Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
124	U2	129C	183C	189	362									-	-	-	-	-	-	-	2 (0.026)
125	L1d	129	183C	189	212	223	230	243	311	390				-	1 (0.015)	-	-	-	-	-	-
126	L1c3	129	183C	189	215	223	278	294	311	356	360	390C		1 (0.008)	-	-	-	-	-	-	-
127	L1c	129	183C	189	215	223	278	294	311	360				-	1 (0.015)	-	-	-	-	-	-
128	L1c	129	183C	189	215	223	278A	294	311	360				-	2 (0.030)	-	-	-	-	-	-
129	M1a	129	183C	189	223	249	271	311	359					-	-	-	-	-	1 (0.014)	-	-
130	M1a	129	183C	189	223	249	278	311	359					-	-	-	-	-	1 (0.014)	-	-
131	M1a	129	183C	189	223	249	311	359						-	-	-	1 (0.018)	-	1 (0.014)	-	-
132	U1a	129	183C	189	249	288								-	-	-	-	-	-	-	1 (0.013)
133	L2d1	129	183C	189	278	300	311	354	390C	399				1 (0.008)	-	-	-	-	-	-	-
134	L1d	129	186	187	189	212	223	230	243	311	390T			-	1 (0.015)	-	-	-	-	-	-
135	L1d	129	187	189	212	223	230	243	291	311				1 (0.008)	-	-	-	-	-	-	-
136	L1d	129	187	189	212	223	230	243	311	390				3 (0.025)	3 (0.045)	-	-	-	-	-	-
137	L1c	129	187	189	214	223	265C	278	286A	291	294	311	360	1 (0.008)	-	-	-	-	-	-	-
138	L1*	129	187	189	218	223	227	239	243	294	311			-	1 (0.015)	-	-	-	-	-	-
139	L1d1	129	187	189	223	230	239	243	294	311	320			2 (0.017)	-	-	-	-	-	-	-
140	L1d1	129	187	189	223	230	239	243	294	311	325	362		1 (0.008)	-	-	-	-	-	-	-
141	L1d1	129	187	189	223	230	239	243	294	311	325			-	1 (0.015)	-	-	-	-	-	-
142	L1d1	129	187	189	223	230	239	243	294	311				-	1 (0.015)	-	-	-	-	-	-
143	L1d	129	187	189	223	230	243	311	390					-	1 (0.015)	-	-	-	-	-	-
144	L1d	129	187	189	223	230	243	311						1 (0.008)	-	-	-	-	-	-	-
145	L1d1	129	187	189	223	239	243	261	294	311				-	1 (0.015)	-	-	-	-	-	-
146	L1c	129	187	189	223	239	243	294	311					1 (0.008)	1 (0.015)	-	-	-	-	-	-
147	L1c1	129	187	189	223	278	293	294	311	360				-	2 (0.030)	-	-	-	-	-	-
148	L1d	129	187	189	230	234	243	266A	311					3 (0.025)	3 (0.045)	-	-	-	-	-	-
149	L1d	129	187	189	230	234	243	266G	311					-	1 (0.015)	-	-	-	-	-	-
150	L2*	129	189	212	223	230	243G	311	390					1 (0.008)	-	-	-	-	-	-	-
151	M1a	129	189	223	249	311	359							-	-	-	-	-	-	2 (0.043)	-
152	L2d	129	189	278	300	352	354	390	399					-	-	-	1 (0.018)	-	-	-	-
153	L2	129	192	223	278	294	309	390						-	-	-	-	-	1 (0.014)	-	-
154	L2	129	223	242A	278	294	309							-	-	-	-	-	-	1 (0.022)	-

continued

Appendix. Table A.7. continued

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)										Count (and Frequency) in the Studied Populations and Comparison Populations							
												Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
155	I	129	223	256G	391							-	-	-	-	-	-	-	1 (0.013)
156	I	129	223	264	270	311	319	362	391			-	-	-	-	-	-	-	1 (0.013)
157	L2	129	223	278	294	309	390					-	-	-	2 (0.036)	-	-	-	-
158	I	129	223	291								-	-	1 (0.034)	-	-	-	-	-
159	I	129	223	391								-	-	-	-	-	-	-	2 (0.026)
160	L1*	140	187	189	223	239	243	294	311			-	1 (0.015)	-	-	-	-	-	-
161	L2	145	150	189	223	278	294	309				-	-	-	-	-	-	1 (0.022)	-
162	N1b	145	176	223	261	311						-	-	-	1 (0.018)	-	-	-	-
163	N1b	145	176A	223	390							-	-	-	-	-	-	-	5 (0.064)
164	N1a	147G	170	172	223	248	355					-	-	-	-	-	1 (0.014)	-	-
165	N1a	147G	172	223	248	355						-	-	-	-	-	1 (0.014)	-	-
166	L1a	148	172	187	188A	189	214	223	230	234	311	-	-	-	-	-	1 (0.014)	1 (0.022)	-
167	L1a	148	172	187	188G	189	223	230	311	320		9 (0.076)	6 (0.090)	4 (0.138)	3 (0.054)	-	-	-	-
168	L1a	148	172	187	188G	189	223	230	311N	320		-	1 (0.015)	-	-	-	-	-	-
169	L1a	148	172	187	188G	189	223	224del	230	311	320	-	-	-	1 (0.018)	-	-	-	-
170	L2*	148	183C	189	223	224	278	390				1 (0.008)	-	-	-	-	-	-	-
171	L3*	148	192	223	234	311						-	-	-	-	-	-	2 (0.043)	-
172	L2a	148	223	234	249	278	294	295	390			-	1 (0.015)	-	-	-	-	-	-
173	H	153	218									-	-	-	1 (0.018)	-	-	-	-
174	V	153	298									-	-	-	-	-	-	-	3 (0.038)
175	W	166	192	223	292	343						-	-	-	-	-	-	3 (0.065)	-
176	W	166	192	223	292							-	-	-	-	-	-	1 (0.022)	-
177	L2	166del	183C	189	223	278	292	294	309	390		-	-	-	-	-	1 (0.014)	-	-
178	H01	167	274	304	482							-	-	-	-	-	-	-	1 (0.013)
179	U3	168	189	235	311	343						-	-	-	-	-	-	1 (0.022)	-
180	L3*	169	213	223	256	278	311	344				-	-	-	-	-	-	1 (0.022)	-
181	L3*	169	223	256	278	305	311	320				-	-	-	-	-	1 (0.014)	-	-
182	L3*	169	223	278	298	311						-	-	-	-	-	1 (0.014)	-	-
183	L3*	169	223	278	311							-	-	-	-	-	1 (0.014)	-	-
184	L3*	169	223	278								-	-	-	-	3 (0.046)	1 (0.014)	-	-
185	L	169	231	278	311							-	-	-	-	-	-	1 (0.022)	-

continued

Appendix Table A.7. continued

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)													Count (and Frequency) in the Studied Populations and Comparison Populations							
															Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
186	M*	169+C	183C	189	223	274	311	319	320						-	-	-	1 (0.018)	-	-	-	-
187	U6a1	172	183C	184	189	219	278	354							-	-	-	-	-	1 (0.014)	-	-
188	L3e2b	172	183C	186	189	223	292	320							1 (0.008)	-	-	-	-	-	-	-
189	U6	172	183C	189	219	278	311	360							-	-	-	-	-	-	-	1 (0.013)
190	U6a1	172	183C	189	219	278	311								-	-	-	-	-	1 (0.014)	-	-
191	U6a1	172	183C	189	219	278									-	-	-	-	-	1 (0.014)	-	-
192	L3e2b	172	183C	189	223	320									2 (0.017)	-	-	-	-	-	-	-
193	L1c2	172	187	189	223	265C	278	286G	294	311	360				1 (0.008)	-	-	-	-	-	-	-
194	?	172	189	218	230	234	243	311							-	1 (0.015)	-	-	-	-	-	-
195	H	172	192	456											-	-	-	-	-	-	-	1 (0.013)
196	U6a*	172	219	278											-	-	-	-	-	-	-	2 (0.026)
197	L3a2	172	223	287	293T	311	354	355	362	399					-	-	-	-	-	1 (0.014)	-	-
198	L3a2	172	287	293T	311	355	362	399							-	-	-	-	-	1 (0.014)	-	-
199	K	176	223	224	278	311									-	-	-	1 (0.018)	-	-	-	-
200	L3e1	176	223	327											-	1 (0.015)	-	-	-	-	-	-
201	L1d	182C	183C	187	189	223	230	243	274	278	290	300	311		1 (0.008)	-	-	-	-	-	-	-
202	L2a1b	182C	183C	189	192	223	278	290	294	309	390				5 (0.042)	1 (0.015)		1 (0.018)				
203	M1	182C	183C	189	223	249	311								-	-	-	-	-	2 (0.027)	-	-
204	L3*	182C	183C	189	223	260	264	311	362						-	-	-	-	-	1 (0.014)	-	-
205	L2a1b	182C	183C	189	223	278	290	294	292	390						-	-	-	-	-	-	-
206	L2a1b	182C	183C	189	223	278	290	294	309	390					15 (0.126)	2 (0.030)	-	2 (0.036)	-	-	-	-
207	L2a1b	182C	183C	189	223	278A	290	294	309	390					-	1 (0.015)	-	-	-	-	-	-
208	L2a1b	182C	183C	189	223	278A	290	294	390						1 (0.008)	-	-	-	-	-	-	-
209	L2a1b	182C	183C	192	223	278	290	294	309	390					1 (0.008)	-	-	-	-	-	-	-
210	X	183C	189	223	278										-	-	1 (0.034)	2 (0.036)	-	-	-	2 (0.026)
211	U2	183C	189	234	266	294									-	-	-	-	-	-	-	1 (0.013)
212	U1	183C	189	249											-	-	-	-	1 (0.015)	-	-	-
213	H	184	265T	399											-	-	-	-	-	-	-	2 (0.026)
214	L3*	185	223	260	311										-	-	-	-	-	1 (0.014)	-	-
215	L3e1a	185	223	327											2 (0.017)	2 (0.030)	-	-	-	-	-	-
216	L3e1a	185	223												1 (0.008)	-	-	-	-	-	-	-

continued

Appendix. Table A.7. continued

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)										Count (and Frequency) in the Studied Populations and Comparison Populations							
												Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
217	L1d2	187	189	223	230	234	243	249	311			-	1 (0.015)	-	-	-	-	-	-
218	L1d	187	189	223	230	243	274	278	290	300	311	-	2 (0.030)	-	-	-	-	-	-
219	L3*	188	189	207T	220	223	260	261	311	362		-	-	-	-	-	1 (0.014)	-	-
220	L2	189	192	223	230	278	294	309				-	-	-	-	-	-	1 (0.022)	-
221	L2a	189	192	223	278	294	309	390				-	-	-	-	-	2 (0.027)	-	-
222	L2a1	189	192	223	278	294	309					1 (0.008)	-	-	-	-	-	4 (0.087)	-
223	L3*	189	223	270DEL	278							-	-	1 (0.034)	-	-	-	-	-
224	L3*	189	223	278								-	-	1 (0.034)	-	-	-	-	-
225	L2a1	189	223	278A	294	309	390					-	1 (0.015)	-	-	-	-	-	-
226	K	192	210	224	311							-	-	-	-	4 (0.062)	-	-	-
227	U5a1*	192	243	256	270	390	399					-	-	-	-	-	-	-	1 (0.013)
228	U5a1*	192	256	265C	270							-	-	-	-	-	-	-	1 (0.013)
229	D	207C	223	260	264	311	320	356	362			-	-	-	-	-	1 (0.014)	-	-
230	D	207T	217	220	223	260	261	311	362			-	-	1 (0.034)	-	-	-	-	-
231	L2a	209	223	278	294	301	354	390				-	-	-	-	-	1 (0.014)	-	-
232	L3a1a	209	223	292	311							-	-	-	1 (0.018)	-	2 (0.027)	-	-
233	L3f	209	223	311								1 (0.008)	2 (0.030)	-	-	-	-	-	-
234	L3a1	209	223	355								1 (0.008)	-	-	-	-	-	-	-
235	?	214	217	335								-	-	-	-	1 (0.015)	-	-	-
236	K	223	224	234	266	311						-	-	-	-	-	-	-	1 (0.013)
237	K	223	224	234	311							-	-	-	-	-	-	-	4 (0.051)
238	L2/L3*	223	224	278	311							-	-	-	-	-	1 (0.014)	-	-
239	L3e1	223	239	323DEL								2 (0.017)	-	-	-	-	-	-	-
240	L3*	223	260	265	311							-	-	-	-	-	1 (0.014)	-	-
241	L3*	223	260	311								-	-	-	-	-	1 (0.014)	-	-
242	L2c2	223	264	278	311	390						1 (0.008)	1 (0.015)	-	-	-	-	-	-
243	L3*	223	270	311								-	-	-	-	-	1 (0.014)	-	-
244	L3a2	223	274	293T	311	355	362	399				-	-	-	-	-	1 (0.014)	-	-
245	L2a1	223	278	286	291	294	309	390				2 (0.017)	-	-	-	-	-	-	-
246	L2a1	223	278	286	294	309	390					3 (0.025)	-	5 (0.172)	-	-	-	-	-
247	L2a1	223	278	286	294	309						2 (0.017)	-	-	-	-	-	-	-

continued

Appendix. Table A.7. continued

Sequence Number	Hg[a]	HVSI Sequence (16,040-16,399)					Count (and Frequency) in the Studied Populations and Comparison Populations							
							Lem[b]	Ban[b]	YS[b]	YH[c]	YJ[c]	Eth[c]	EJ[c]	AshJ[c]
248	L2a1	223	278	294	309	390	1 (0.008)	2 (0.030)	-	-	-	-	-	-
249	L3e1	223	311	327			1 (0.008)	-	-	-	-	-	-	-
250	M*	223	311	362	400		-	-	-	1 (0.018)	-	-	-	-
251	L3*	223	311	362			-	-	-	-	2 (0.031)	-	-	-
252	L3e1	223	311	323DEL	327		1 (0.008)	-	-	-	-	-	-	-
253	N*	223	319				-	-	-	-	-	-	1 (0.022)	-
254	L3e1	223	327				1 (0.008)	-	-	-	-	-	-	-
255	L3e3	223	265T				-	1 (0.015)	1 (0.034)	1 (0.018)	-	-	-	-
256	L3e1	223	323DEL	327			1 (0.008)	-	-	-	-	-	-	-
257	M*	223					-	-	-	1 (0.018)	-	-	-	-
258	K	224	234	311			-	-	-	-	-	-	-	6 (0.077)
259	K	224	311				-	-	-	-	-	-	-	3 (0.038)
260	U5a1a	256	270	293	399		-	-	-	2 (0.036)	-	-	-	-
261	U*	260	278				-	-	-	1 (0.018)	-	-	-	-
262	U2i	261	278	311			-	-	-	-	-	1 (0.014)	-	-
263	R2	286	311	320			-	-	1 (0.034)	-	-	-	-	-
264	R2	286	320				-	-	2 (0.069)	-	-	-	-	-
265	U7	291	304	318T			-	-	-	-	-	-	-	1 (0.013)
266	H	311	362	482			-	-	-	-	-	-	-	2 (0.026)
267	HV*	311					-	-	-	-	-	1 (0.014)	-	-
268	U3	343	390					-	-	-	-	-	-	2 (0.026)
						Total	119	67	29	56	65	74	46	78

[a] Lemba, Bantu, and Yemen-Sena sequences were assigned to hgs based on sequence data (but see (d) below). Data for the remaining populations was provided by M B Richards

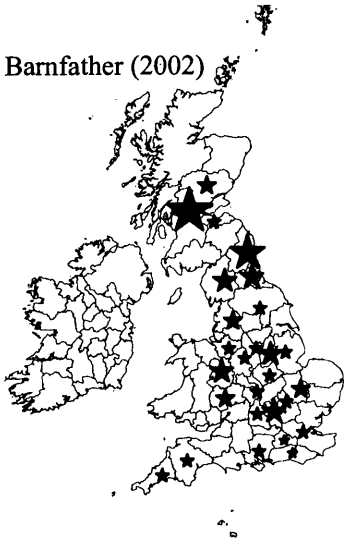
[b] This study

[c] Data kindly provided by M B Richards

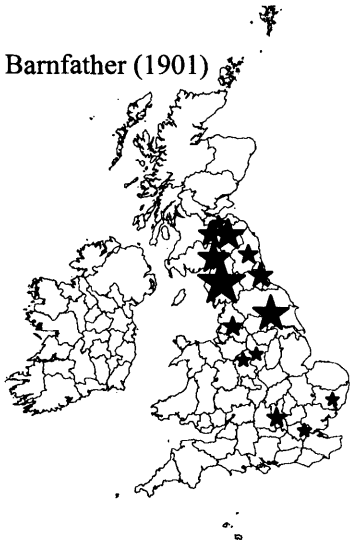
[d] Hg assignment was ambiguous based on sequence data alone, therefore they were assigned to hgs using RFLPs (kindly performed by the Torroni lab, see Appendix, Table A.6)

Appendix. Figure A.1. Maps Showing the Distribution of the Studied Surnames in England, Wales and Scotland in 2002 and 1901 by County.

Barnfather (2002)

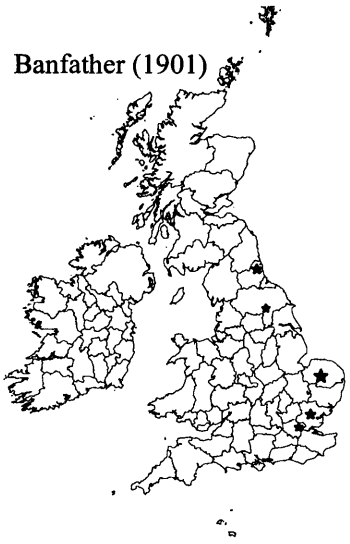


Barnfather (1901)



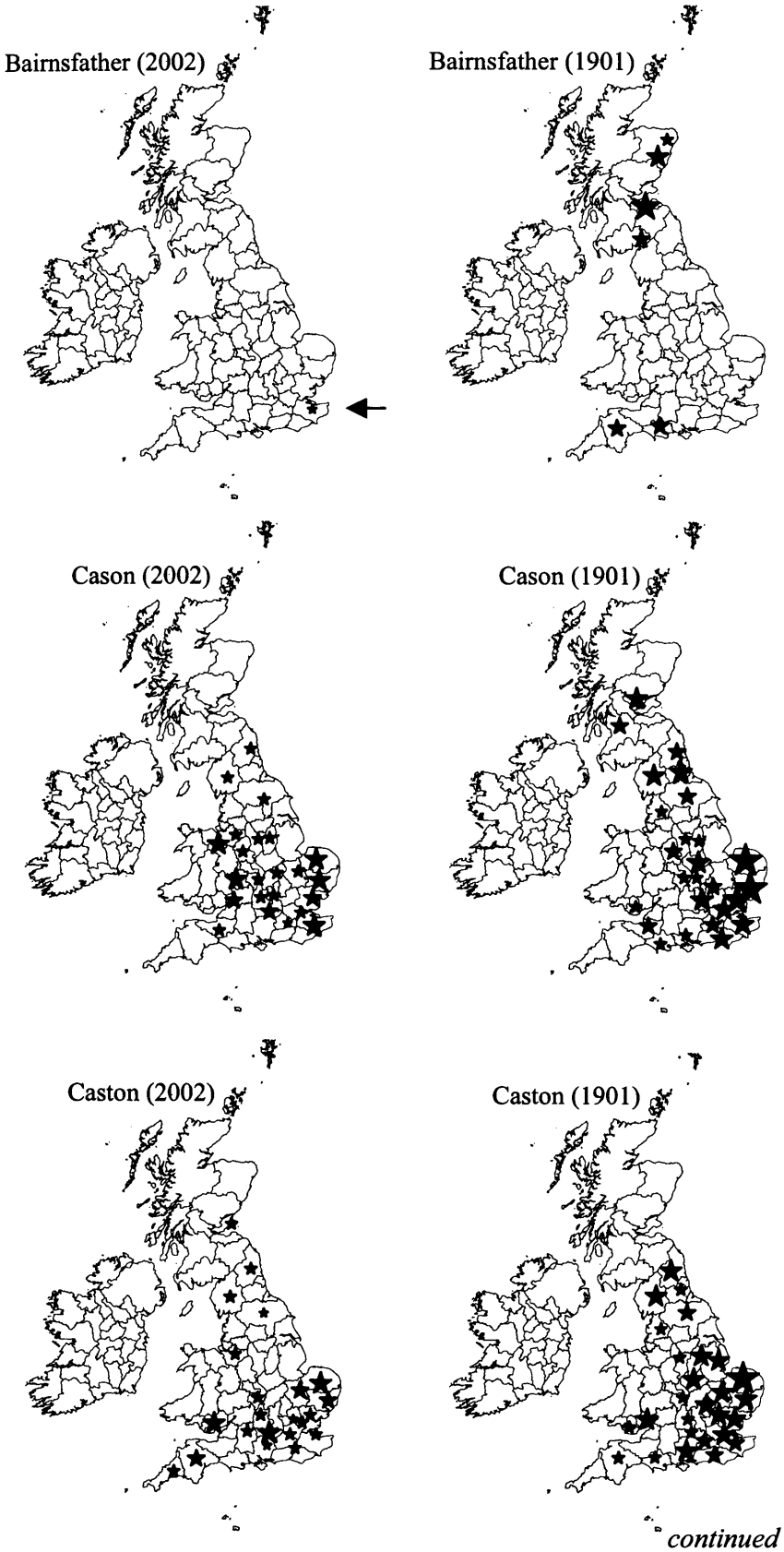
Banfather (2002)

Banfather (1901)



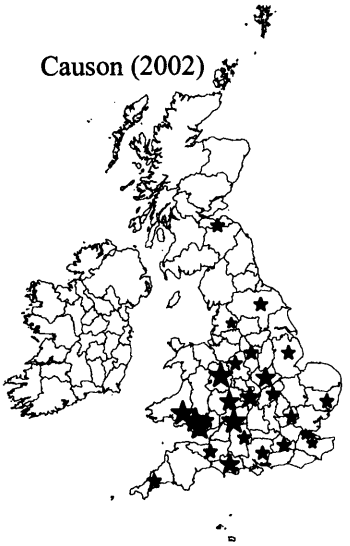
continued

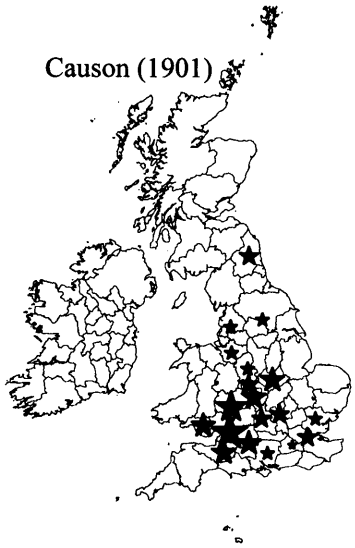
Appendix. Figure A.1. continued

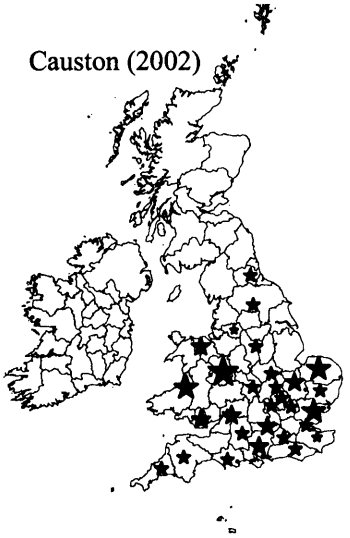


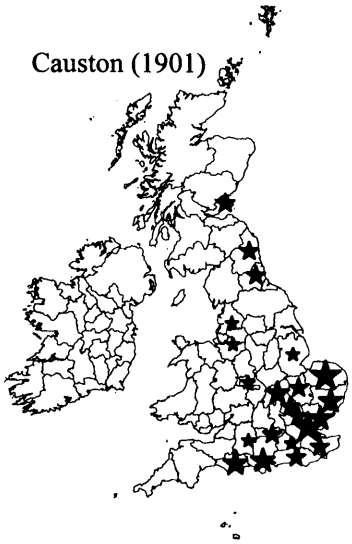
continued

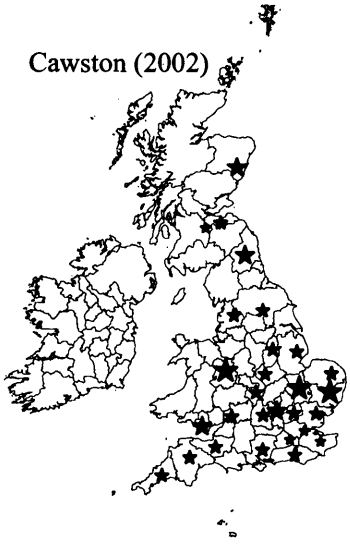
Appendix. Figure A.1. continued

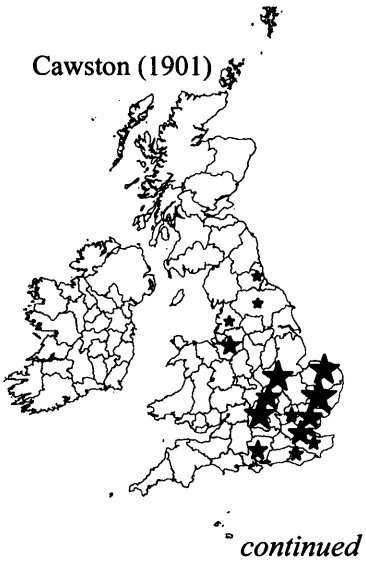






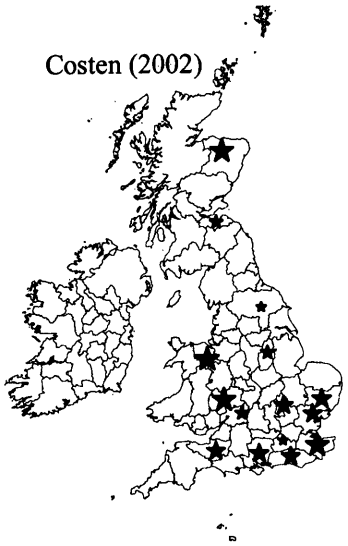


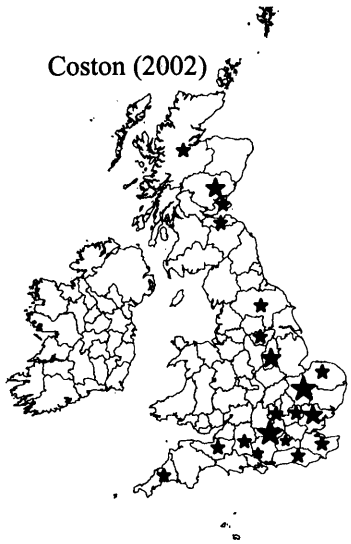


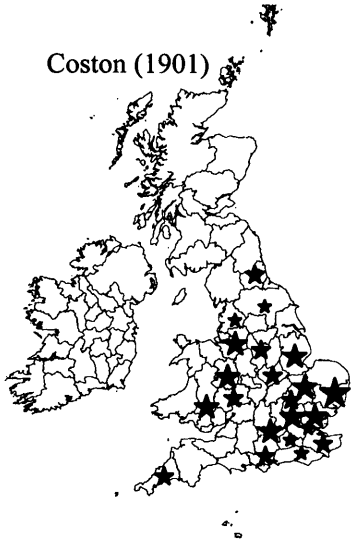


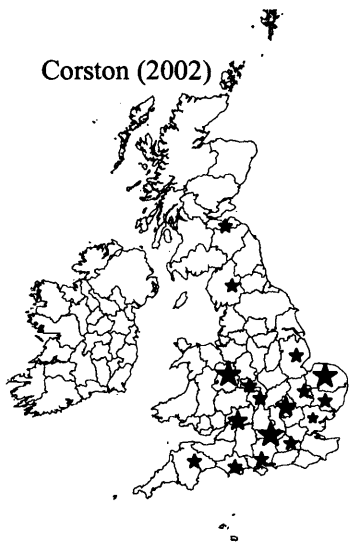
continued

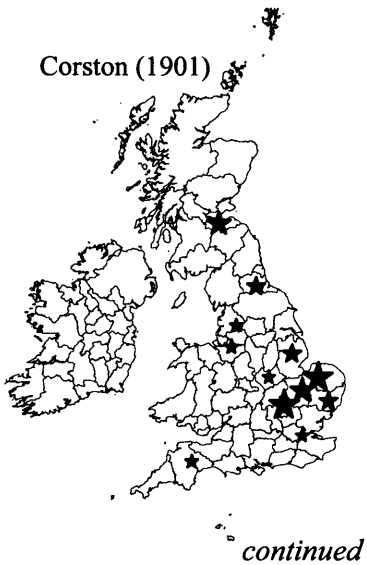
Appendix. Figure A.1. continued





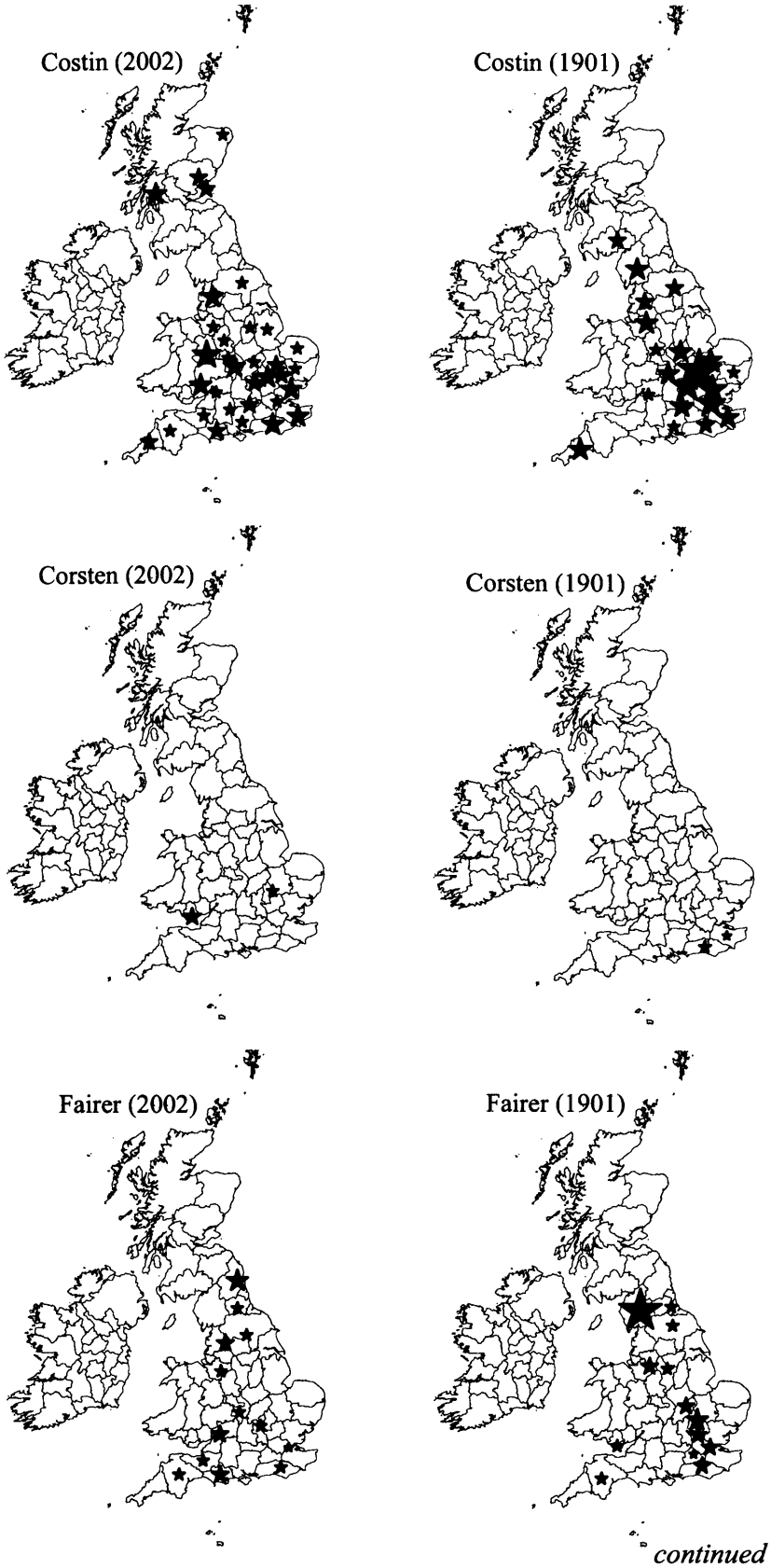






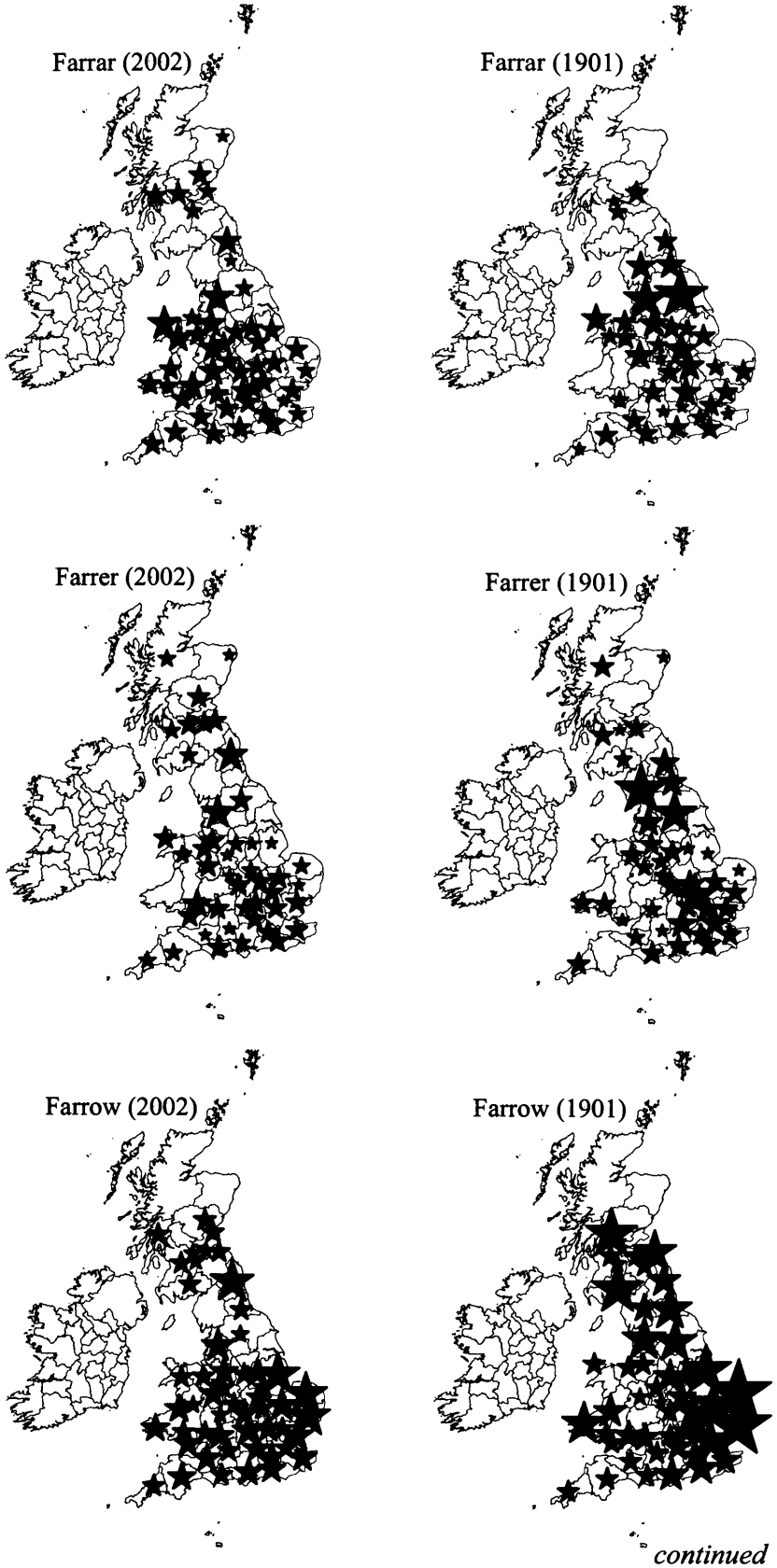
continued

Appendix. Figure A.1. continued



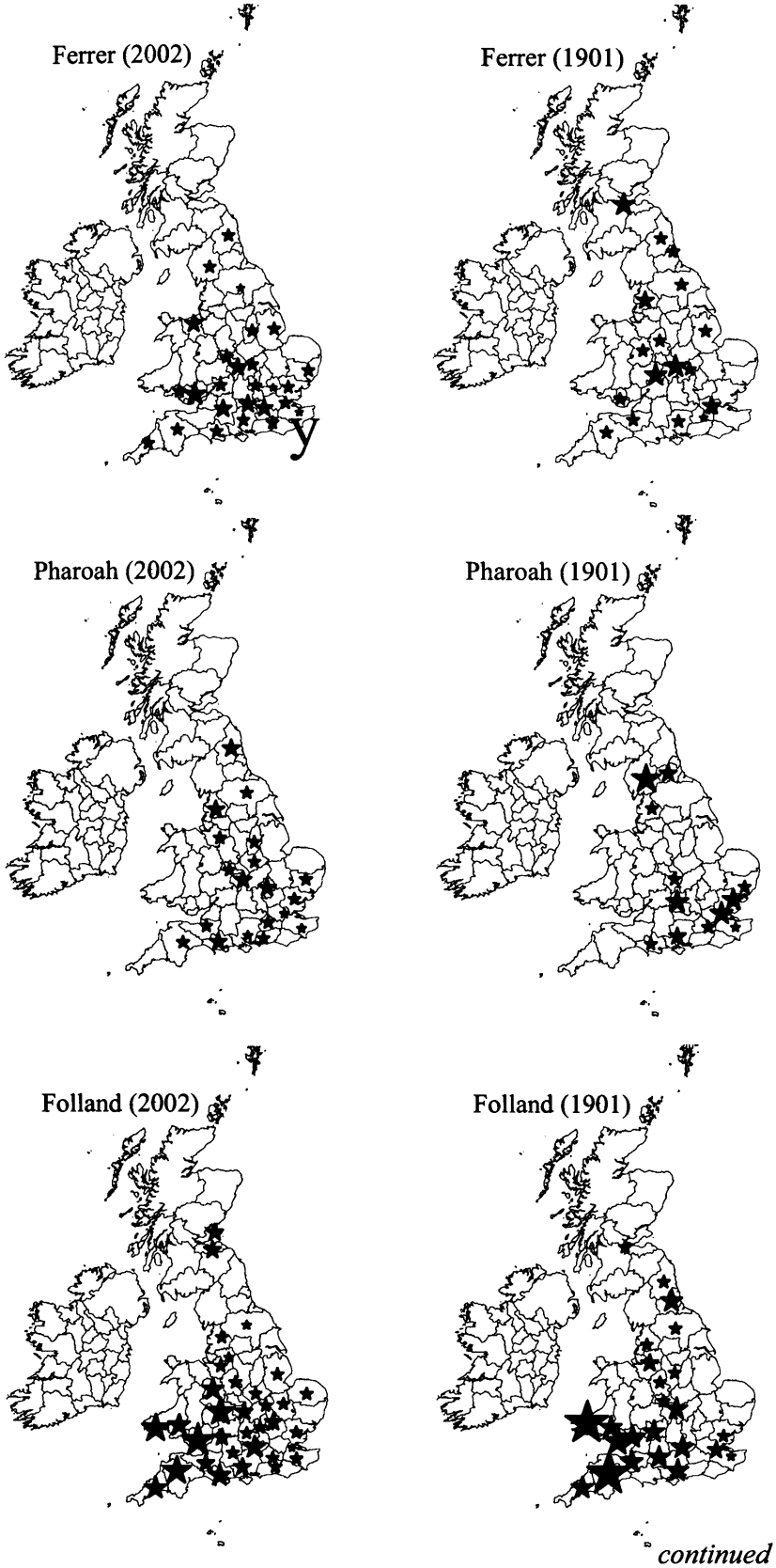
continued

Appendix. Figure A.1. continued



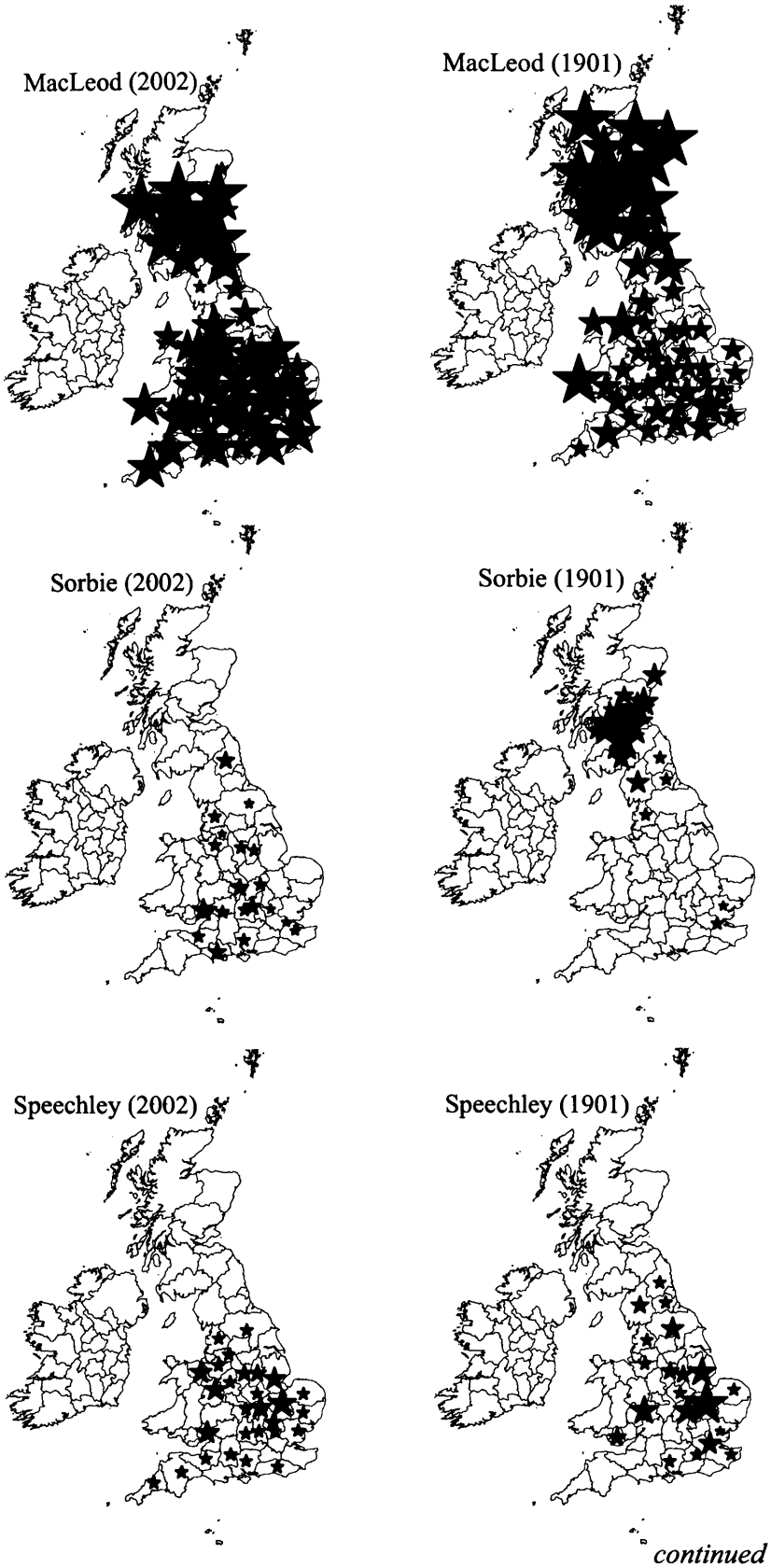
continued

Appendix. Figure A.1. continued



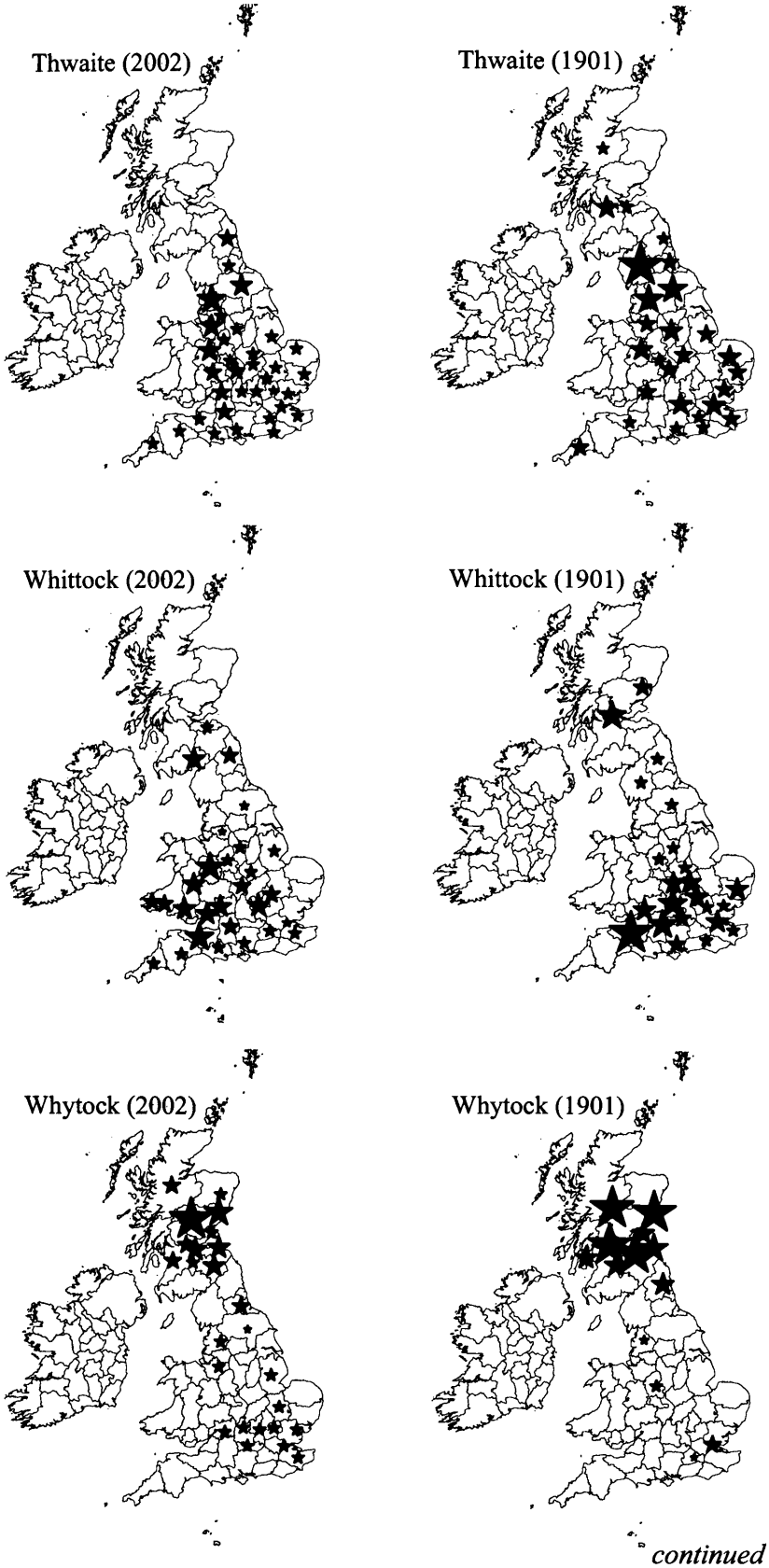
continued

Appendix. Figure A.1. continued



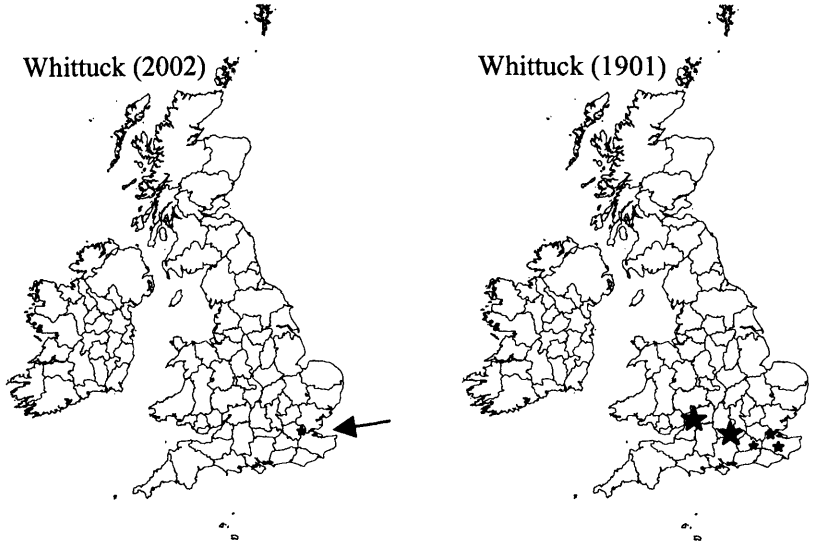
continued

Appendix. Figure A.1. continued



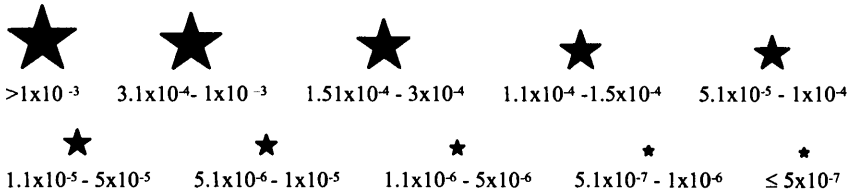
continued

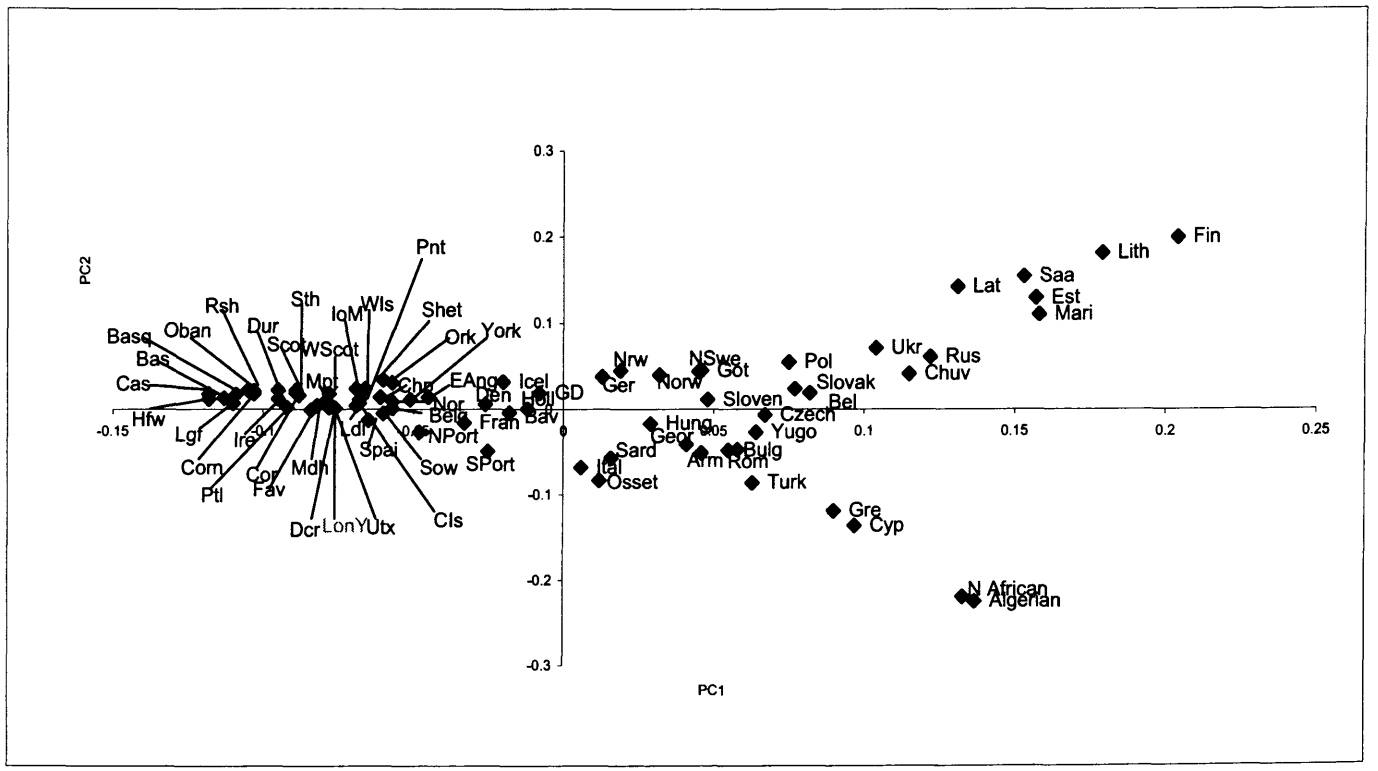
Appendix. Figure A.1. continued



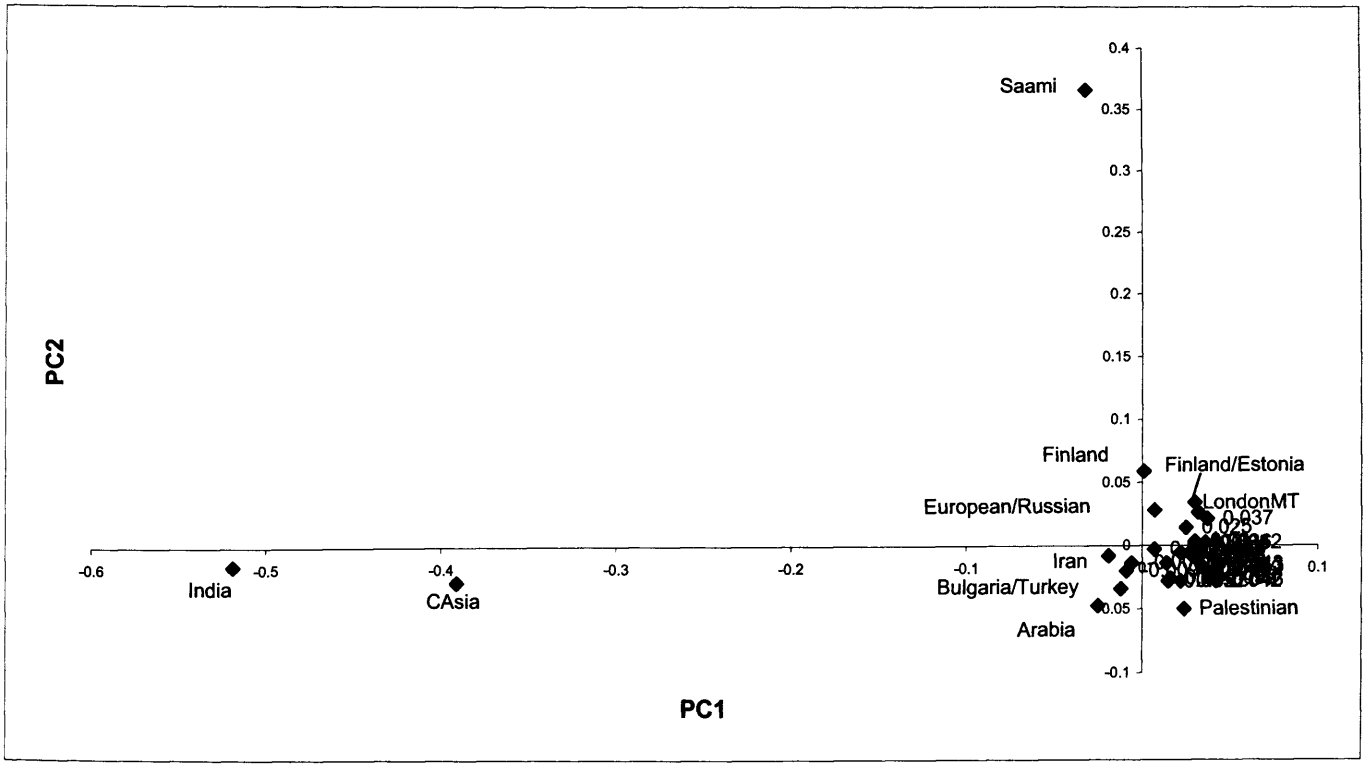
Notes: 2002 information has been taken from the BT Telephone Directory, and 1901 information is from the 1901 Census. Shown are the distributions of the surnames expressed as a percentage of the total population of each county (where the size of each county has been obtained from the 2001 and 1901 Censuses). The size of the star is proportional to the frequency of the name. The surnames were expressed as percentages to control for the large population sizes seen in metropolitan regions.

Key: The stars represent the frequency of the surnames in each county. Note that to best display the data not all of the intervals are equal



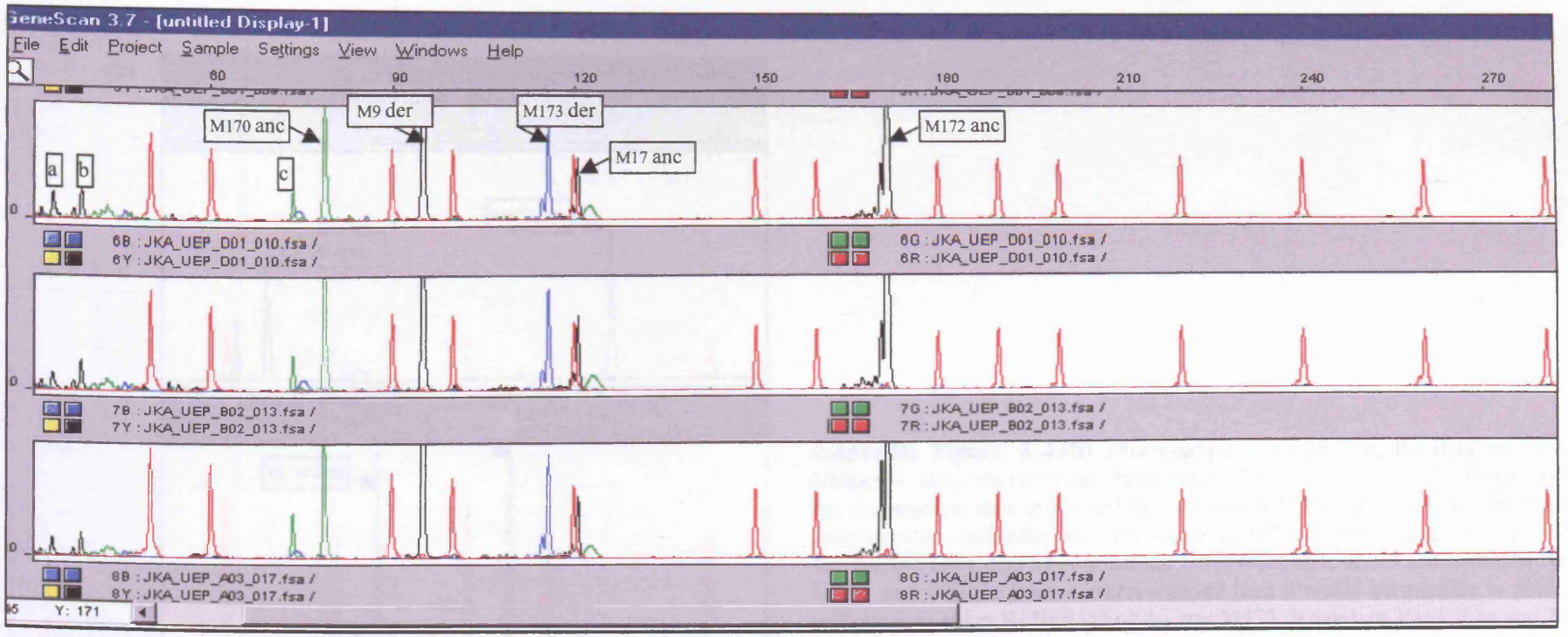


Appendix. Figure A.2. Y-Chromosome PC Plot of LondonY, the BCD and the RD Using Hg Frequencies. As each of the populations from the BCD and the European populations from Chapter 2 have been used in this PC plot it is difficult to differentiate many of the British populations, which gorup together in the negtive half of the x-axis and many of the remaining European populations to the positive half. Therefore it was difficult to assess the relationship of LondonY to the European populations. For this reason the Y chromosome PC plots in Chapter 4 were drawn by clustering the BCD into "AllBrit". The 3 European populations analysed in Chapter 2 (Norway, North Germany/Denmark, and Basques) were also exlcuded from the plots in Chapter 4 to increase their clarity, as these 3 populations cluster very close to equivalent populations from the RD. Abbreviations as Table 2.8 and Figure 4.5.



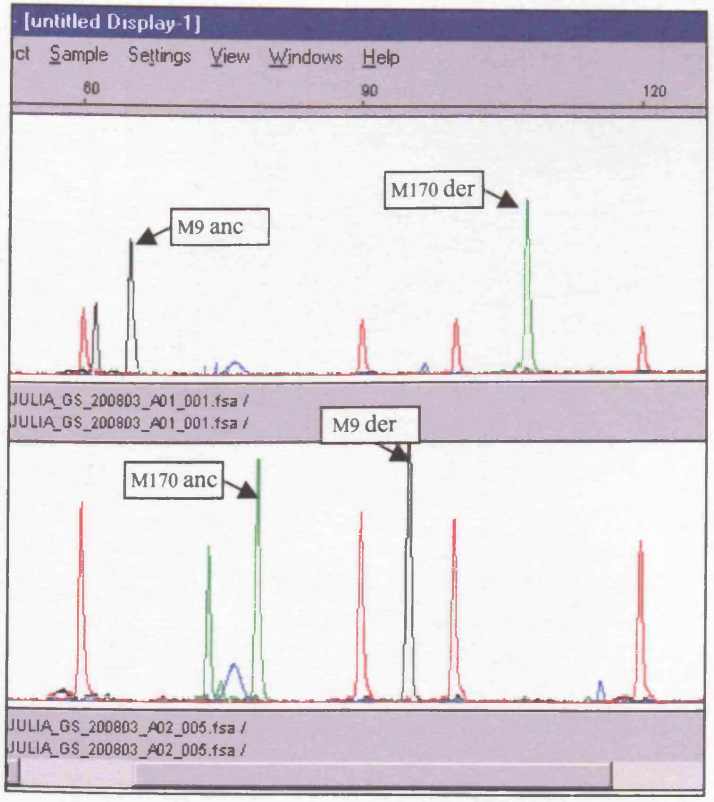
Appendix. Figure A.3. mtDNA PC of LondonMT and European Comparison Populations Using HG Frequencies. Unlike the plot shown in Figure 4.7, all of the European comparison populations are used. This highlights the fact that India, CAsia and Saami fall as such extreme outliers that the remaining populations are predominantly forced into an undifferentiated cluster. Therefore these outlying populations were excluded from the plots shown in Chapter 4.

(a)



Appendix Figure A.4. Screen Capture of the GeneScan Output for the Euro1 PCR Multiplex Kit Electrophoresed on an ABI 3700 Sequencer. (a)This example is taken from the data presented in Chapter 3. There are three horizontal "lanes", each of which represents a different individual. The red peaks show the ROX size standard, and the blue, black and green peaks relate to the assayed alleles, which have been fluorescently tagged with FAM, NED, and HEX dies, respectively. Each assayed allele has been labelled, stating whether the observed state is ancestral or derived (e.g. "M170 anc", "M9 der"). As of the size of each of the labelled alleles in each individual is the same all 3 individuals have the same genotype (classified as M173 derived or R*(xR1a1)). See (b) for an example of individuals having a different genotype for two of the assayed alleles. Note that in this example the marker 92R7 has failed to amplify, but as this marker is not necessary for haplogroup designation (see p 115 and Figure 2.6) these samples were not excluded from the analysis. As can be seen there are three peaks (labelled a, b, c) which do not correlate to the assayed alleles, however, these peaks do not cause confusion in interpreting the genotypes here as their respective sizes do not correlate to any of the expected allele sizes (see for example Table 2.4).

(b)



Appendix Figure A.4.(b) This example is taken from the data presented in Chapter 3. Here the two individuals have different genotypes. The first individual has the ancestral state at M9 and derived state at M170 and given the other markers assayed in this multiplex this correlates to an individual belonging to hg I*(xI1b2). The second individual by contrast has the derived state at M9 and ancestral state at M170, and with the other markers assayed here it would be possible to place this individual in either R1*(xR1a1) it he was M173 derived, or R1a1 if he was M173 and M17 derived. As with the examples in (a) there are peaks present which do not correlate to assayed alleles, however as they do not have sizes like those expected there is not any confusion with the alleles that are expected.