A LONGITUDINAL MITOCHONDRIAL DNA ANALYSIS OF MODERN HUMAN HAIR SAMPLES FROM THE ANYANG REGION OF CHINA: IMPLICATIONS FOR ANCIENT DNA RESEARCH

BY

ARLENE LAHTI

A THESIS PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOLOGY
LAKEHEAD UNIVERSITY
THUNDER BAY, ONTARIO

OCTOBER, 2003

Your file Votre référence
ISBN: 0-612-96981-9
Our file Notre référence
ISBN: 0-612-96981-9

ABSTRACT

In 1998, Dr. E. Molto was selected to conduct pilot research on the Hei Heru, Shang Dynasty site in Anyang, China. A major component of this research is the ancient DNA analysis of human remains recovered at Anyang. Preliminary work on these ~3300 year old remains has led to an auxiliary study that focuses upon modern DNA extracted from hairs donated by individuals who currently live in the farming villages found within the greater Anyang Site area. This latter study uses a 'blind study' approach, to analyze human hairs in an effort to identify ancestor-descendent relationships in a controlled situation in order to validate the methods used to examine the ancient specimens. We secured 30 hair samples from 30 individuals who reside in three remote villages with a long history of occupation in the area. The hair samples include a series of identified lineages representing grandmothers, daughters and grandchildren. The analysis sought to identify individual matrilineages, validate interpretative methods relevant to the ancient Anyang remains, to critically address the implications of short-term mutations for longitudinal studies, and evaluate the Anyang data set with other mtDNA studies from Asia.

ACKNOWLEDGEMENTS

With special depths of gratitude, I would like to thank my supervisors, Dr. Scott Hamilton and Dr. El Molto. When other people may have believed I would never finish my masters, El and Scott never gave up. My experience in the Masters program and their experience with me as their student have made us friends as well, which is a priceless gift. As professors go, El and Scott are the best.

Thanks to Dr. Jigen Tang for organizing the collection of the hair samples from the small villages in China for this research project.

I would also like to thank the members of my graduate committee. The internal committee members, Dr. Murray Lankester, Dr. Kam Leung and Dr. Lada Malek, that provided a valuable critique of my research. My external advisor, Dr. Keith McKenny, taught me a lot about mitochondrial DNA.

To my fellow grad students, Amy Junnila and Curtis Hildebrandt, thanks for giving me strength, making me laugh when things were really hard, and for always being there for me and with me through the trials and tribulations of doing graduate work.

To other friends I've made through Lakehead University, Jenny Brown, Carolyn Murray and Laura Smith, thanks for all the little things, but nevertheless important contributions to my thesis.

Thanks to the special people who work behind-the-scenes in the university paperwork shuffle, Cindy Lamontagne and Eleanor Maunula. I know they made my life easier.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Over two decades have elapsed since the human mitochondrial DNA (mtDNA) genome was completely sequenced by a research team at Cambridge University, England (Anderson et al. 1981). Since that time, mtDNA has impacted scientific knowledge on many fronts including human aging (Wallace, 1995), human disease (Wallace, 1999) and human origins/population history (Merriwether et al. 1994, von Haesler et al. 1995). It is in the latter anthropological sphere that mtDNA gained its initial notoriety which continues to this day. The mtDNA based 'Out of Africa' hypothesis (Cann et al. 1987) rocked the anthropological world by proposing a fundamental paradigm shift whereby modern humans (*Homo sapiens*) were hypothesized to have evolved from one human female lineage in Africa approximately 150,000-200,000 years ago. Previously, paleoanthropologists had hypothesized that the human lineage represented by the genus 'Homo' could be traced back continuously for at least a million years (Wolpoff et al. 2001)! Though still acrimoniously debated, this molecular research illustrated the inherent value of mtDNA for deriving genealogical and historical origins. The haplotypic or clonal nature of mtDNA from exclusive maternal inheritance, concomitant with the absence of segregation and recombination, made mitochondrial DNA easier to interpret than nuclear DNA. Moreover, the non-coding D-Loop region of the mtDNA genome had a much higher mutation rate than nuclear DNA (Cann et al. 1984) which made it ideally suited for reconstructing population history. The 'Out-of-Africa' hypothesis also incorporated a long standing indirect approach to genealogical reconstruction first used by molecular researchers at the turn of the last century, namely the use of modern living

populations to reconstruct phylogenetic patterns based on accumulated differences. Later Sarich and Wilson (1967) suggested that the timing of the separation of past species could be inferred by assuming constant rates of mutational change, through the concept of 'molecular clocks'. This method first computes genetic relationships based on 'mutational' differences between modern samples and then, assuming constant 'mutation rates', projects the length of time for the separation of extant populations. For the 'Out of Africa'/ Eve Hypothesis it was suggested that since modern African populations invariably have the largest number of mutations in the mtDNA D-Loop region, they must represent the founding lineage of human evolution.

With the development of the polymerase chain reaction (PCR) in the 1980s the ability to recover ancient DNA became a reality. This provided researchers with the ability for 'direct' sampling of ancient human genomes. Again mtDNA provided yet another advantage over nuclear DNA, namely, a high copy number per cell which increased the probability of recovering ancient DNA. For example, in bone, which is the tissue most often used in ancient DNA research, each bone cell has several hundred mitochondria but only one nucleus, which increases the probability of recovering a desired locus of DNA. Numerous early success stories for recovering and interpreting ancient mtDNA in archaeological populations were reported in the literature including the recovery and sequencing of Neandertal mtDNA (Krings et al. 1997). These early successes also revealed major problems: contamination by modern DNA and inconsistent results. Due to the latter problems, most researchers reverted to the 'indirect' approach to reconstructing population history.

In 1998, pilot mtDNA research on ancient Chinese samples from the Shang Dynasty capitol of Anyang was initiated by Dr. E. Molto under the auspices of the Paleo-DNA laboratory and the Department of Anthropology at Lakehead University. This research was part of a larger archaeological program conducted jointly between the University of Minnesota and the Institute of Archaeology (IA) of the Chinese Academy of Social Sciences (CASS), People's Republic of China, with the permission of the China State Bureau of Cultural Relics. Bone and tooth samples from the Hei Heru cemetery of the Anyang Site, Henan Province, China, dating to circa, 1300 B.C., were the subject of a Master of Science Thesis by Ms. Alison Graver (Graver, 1999). Though ancient mtDNA was recovered, the preservation was at best poor and inconsistent from sample to sample (Graver et al. 2000). This reduced the interpretive aspects of the ancient DNA (aDNA) research. To further the interpretive framework for this ancient DNA initiative, hair samples from modern individuals who claimed a long tenure in the Anyang region were collected by Dr. Jigen Tang, Director of the Anyang Research Station. These samples represented succinct lineages of grandmothers-mothers-children to help understand the haplotypic variation in this region over 'limited' time. It was reasoned that since China covers such a large geographical region and the large Chinese population may be relatively heterogeneous, a better understanding of the mtDNA variation in this well circumscribed area of historical significance (origin of the Shang Dynasty) was required. Subsequently, Yao et al. (2000) suggests that the southern Han Chinese populations are older and that with mtDNA expansion to the north and east, HV1 diversity generally decreased. This retrospective study incorporated a large cross-sectional database of 1218

individuals from across China. The implications of this study for Chinese origins are significant and contrast the initial hypothesis for the aDNA program from Anyang.

The modern hair samples collected from 30 Han Chinese in the Anyang District of Central China form the basis for this mtDNA thesis. The research design is both blind and longitudinal, which is essential for examining the following objectives:

(1) To test the efficacy of mtDNA for lineage assignment of these individuals;

(2) To examine the mtDNA for sources of variation particularly heteroplasmy;

(3) To compare these data to the aDNA results previously generated for ancient Anyang samples; and

(4) To test the fit of these data relative to the homogeneity hypothesis for Han Chinese recently proposed by Oota et al. (2002).

The thesis is organized as follows. Chapter 2 provides an overview of ancient DNA research with special emphasis on the Shang Dynasty archaeology; Chapter 3 details the research design, including materials and methods; Chapter 4 presents the results, while the interpretation and significance of the results form the body of Chapter 5.

CHAPTER 2: DEVELOPMENT OF ANCIENT DNA STUDIES

Ancient DNA research has had a short but controversial history that has rapidly changed with the development of technology suitable for the extraction, amplification and sequencing of DNA molecules. Like many new scientific endeavors, there was early optimism for ancient DNA research based on protocols originally developed for modern DNA studies. Specifically, some modern protocols using molecular cloning were successful in recovering nucleic acids from ancient tissue. An important study done by Higuchi et al. (1984), reported that mtDNA was extracted from a quagga, a now extinct relative of the zebra. Svante Pääbo (1984, 1985a, 1985b), another pioneer in ancient DNA studies, was able to extract DNA from the ancient tissue of Egyptian mummies. In addition, the development of PCR technology (Mullis and Faloona, 1987) offered exciting potential for allowing trace amounts of endogenous DNA to be targeted and amplified. For instance, Pääbo et al. (1988) used PCR to analyze mtDNA from a 7000-year-old human brain. In fact, PCR based technology revolutionized the ancient DNA field by allowing researchers to genetically characterize almost any well preserved ancient biological material.

Unfortunately, the early optimism about the broad application of ancient DNA research was overshadowed by the identification of new obstacles. In some circumstances these limitations relate to the degree of preservation of ancient DNA, and in other situations the problems relate to the research and development of more effective processing and analysis protocols of this new and developing discipline. A particularly serious problem relates to the fact that DNA can degrade rapidly and only small

fragments may remain (Pääbo, 1986). This is particularly serious in the context of rapid improvement in methodologies whereby even small remnant portions of the DNA strand may be recovered, amplified and subjected to analysis. In this situation, amplification of minute amounts of modern DNA contamination may be sufficient to obscure and mask the fragments of ancient DNA. Continued research has demonstrated that modern DNA contamination of ancient template can generate false-positive results. PCR, an extremely sensitive technique, will preferentially amplify modern intact DNA over degraded ancient DNA (Brown and Brown, 1992, Handt et al. 1994). This may lead to inadvertent treatment of modern contamination as the DNA of interest. Failure to adequately address and control for contamination issues has led some to challenge the authenticity of some studies that report the successful extraction of very ancient DNA. For example, the recovery of DNA from a 12-20 million-year-old Magnolia leaf (Golenberg et al. 1990), and 120-135 million-year-old weevil (Cano et al. 1993) have not been reproducible. This problem, coupled with time consuming protocols and low success rates because of the often poor quality of ancient DNA, led to the general trend of using modern DNA samples from different populations to reconstruct past population genetic histories (Higuchi et al. 1987, Merriwether et al. 1995, Starikovskaya et al. 1998).

In general, as DNA researchers have addressed these problems, it has become clear that extremely rigorous sample preparation, DNA extraction and amplification standards are required to minimize the risk of modern contamination, and to correctly identify compromised samples as part of the preliminary analytic process.

This research was conducted at the LU Paleo-DNA Laboratory that specializes in

the extraction and analysis of ancient samples. Therefore, the standard analytical procedures are designed to address and control for the risk of modern contamination and other complications affecting the credible recovery of ancient and degraded DNA. These controls are reviewed in some detail in Chapter 3.

The Role of Mitochondrial DNA in Ancient DNA Studies

Most of the early studies into modern and ancient DNA utilized mtDNA for population genealogical analysis and aDNA analysis. The entire mitochondrial genome was fully determined by the early 1980s (Anderson et al. 1981), and subsequently verified by Andrews et al. (1999). The genetic determination of the mitochondrial genome showed that mtDNA offered some analytic advantages. For example, mtDNA is maternally inherited (Giles et al. 1980), unlike nuclear DNA, which is subject to recombination and independent assortment. Maternal inheritance of mtDNA allows family lineages to be tracked through time (Allen et al. 1998). Particularly important for ancient DNA studies is the high copy number of mitochondria in a single cell. While there is only one nucleus, there are between hundreds and thousands of mitochondria per cell (Tamarin, 1996). The sheer number of mitochondria significantly improves the probability of extracting a portion of intact DNA for analysis. Thus, maternal inheritance and high copy number allow for population genealogical studies by mtDNA analysis of modern and ancient materials.

However, mtDNA has general and specific problems that have been identified during early trials. For example, the mitochondrial genome has a higher mutation rate

that over time enables the genetic differentiation of maternally related individuals (Parsons et al. 1997). This higher mutation rate has been used to assess mtDNA biodiversity in differentiating between racial groups (Horai and Hayasaka, 1990) and broadly defined ethnic populations (Melton et al. 2001). But an accelerated mutation rate has not always been useful for reconstructing certain population histories. Studies have shown that regions such as Europe don't reveal sufficient variability to differentiate individuals and assign them to a particular geographic area because the population has been so mobile and interbred (Melton et al. 1997). In contrast, it is possible to identify people of Aboriginal ancestry in North and South America because these populations have a long history of isolation from other human populations. Individuals of Amerind background can generally be genetically typed to one of four founding haplotypes (Bailliet et al. 1994, Horai et al. 1993, Merriwether et al. 1995). Thus mtDNA data can be useful in identifying and tracking some populations from a genetic perspective, except in places like Europe where genetic admixture has been so historically prevalent.

Research design and sampling strategies for genetic research have played a significant role in our understanding of mitochondrial variation at the macro scale level. Many mtDNA studies to date have followed a cross-sectional research design, whereby samples are chosen from a diverse group of unrelated individuals within a large population area. Studies from Europe (Baasner et al. 1998, Cali et al. 2000), North America (Melton et al. 2001), and Asia (Lee et al. 1997, Seo et al. 1998, Tsai et al. 2001, Tzen et al. 2001) have assessed the genetic diversity of the mtDNA hypervariable regions. From a large sample of 2282 individuals, Melton et al. (2001) reports that high diversity

exists within populations, but that little of this variability was specific to sub-populations. That is, variability within geographically specific sub-population was as great as that noted between sub-populations. As well, many of these cross-sectional studies have also sampled from a variety of biological materials. Tully et al. (2000) notes the challenge now is to understand how the complex 'micro-evolutionary' processes of mtDNA mutation, segregation, and transmission affect the more traditional population genetic patterns of mtDNA variation that have long been studied (i.e. cross-sectional studies).

Another potentially confounding issue relates to the phenomenon of heteroplasmy that can occur due to the clonal nature of mtDNA replication and high copy number. Heteroplasmy is the mixture of two bases at a single site in a mtDNA sequence (Stewart et al. 2001, Tzen et al. 2001). Due to the rare occurrence of heteroplasmy, it has proved useful in reconstructing genetic relationships (Wilson et al. 1997). One well-known DNA case involving heteroplasmy was the identification of remains suspected to be the Tsar Nicolas II of Russia. The identification of this individual was confirmed due to a unique point mutation heteroplasmy (Debenham, 1996, Ivanov et al. 1996).

However, heteroplasmy also has the potential to limit the successful reconstruction of genetic relationships. The existence of two distinct peaks at one site in an electropherogram can cause interpretative uncertainty. The difference of one or more sites in the comparison of two genetic profiles can be sufficient to differentiate them as not maternally related. Consequently, a heteroplasmic event could confuse the identification of an ancestor/descendent relationship. Further compounding the potential uncertainty, the level of heteroplasmy can vary among family members, tissues and age

groups (Baasner et al. 1998, Calloway et al. 2000, Gocke et al. 1997, Stewart et al. 2001). Specifically, hair seems to be even more variable than other biological samples, such as blood and saliva (Alonso et al. 2002) and exhibits more occurrences of heteroplasmy (Bendall et al. 1997, Carracedo et al. 2000, Wilson et al. 1997). Heteroplasmy may also be difficult for researchers to recognize or detect due to the limitation of automated sequencing equipment sensitivity (Baasner et al. 1998, Calloway et al. 2000, Seo et al. 1998). Moreover, the frequency of heteroplasmy is not well understood and though rare, even two or more events can occur (Budowle et al. 2002, Tully et al. 2000). Generational affects are not fully understood nor can it be fully addressed through cross-sectional sampling strategies alone (Seo et al. 1998). Thus, heteroplasmy is a relatively new phenomenon and there is still much to learn.

Finally, there is the problem of interpreting ancient DNA results due to limited comparative databases from both modern and ancient sources. Ideally, ancient DNA results should be compared to modern samples that are representative on a regional basis in order to test historical relationships. Wilson et al. (1993) reports that geographically appropriate consensus sequences should be considered in order to understand mtDNA diversity within a specific population. For instance, within the small number of available databases, the frequency of mtDNA types may be uncertain as new variants will be observed more often (Tully et al. 2001). Thus, a higher mutation rate, research design, heteroplasmy and inadequate DNA databases have hindered the interpretative value of mtDNA research. It is with these issues in mind that the current study was undertaken.

In 1998, Dr. El Molto was selected to conduct pilot research on the Hei Heru,

Shang Dynasty site in Anyang, China. Ancient DNA analysis was a major component of this research, and lead to completion of a MSc. research paper on the human remains from Hei Heru by Alison Graver (Graver, 1999), and followed by a multi-authored paper (Graver et al. 2000). As addressed in the latter paper, RFLP (restriction fragment length polymorphism) data were generated for the ~3300-year-old remains. RFLP, or mtDNA haplotype, is defined by the presence or absence of a particular marker within the mtDNA genome. In general, mtDNA haplotype data has been used to track the large-scale movement of populations globally through maternal lineages. It was hoped that the aDNA research at Anyang would aid in addressing the genetic relationship of the ancient Shang skeletal remains in comparison to modern RFLP data. From this research, two main problems came to light. First, DNA extraction methodologies were still developing at the time of the aDNA pilot study, which led to difficulties in consistent recovery of DNA from the ancient remains. Second, the data generated in the aDNA pilot study was difficult to interpret in comparison with mtDNA data from modern populations. This was primarily due to the lack of geographically appropriate modern genetic data. Consequently, this led to the study addressed in this thesis.

Since the aDNA pilot study of ancient remains in Anyang (Graver, 1999, Graver et al. 2000) direct nucleotide sequencing technology became more refined and effective at resolving individual molecules of DNA. DNA sequencing allowed individuals to be compared and different genetic lineages determined. Having this level of interpretative resolution of DNA provides an opportunity to test the assumption that mtDNA reflects clonal transmission from mother to child, as it has been recently suggested that paternal

inheritance of mitochondria is a possibility (Bromham et al. 2003). As well, the heteroplasmy phenomenon can be addressed to determine whether it affects the determination of genetic relationships between individuals. As a result, for this study we secured modern hair samples from 30 individuals from three remote villages with a long association in the Anyang region. Moreover, these individuals were selected using a longitudinal research design, where grandmothers, mothers and children from discrete households were sampled. This pilot study, done in "blind", sought to identify maternally related individuals. Moreover, the nature of this study into mtDNA sequences has relevance in eventually aiding the modeling of the ancient DNA from Anyang.

Recently, however, a major paper on mtDNA in Chinese Han populations suggested that this large Asian population is characterized by high variation within populations, but very low differentiation between ethno-linguistically discrete Asian populations (Oota et al. 2002). Examination of how these Anyang data fit relative to other Chinese mtDNA studies (Seo et al. 1998, Tsai et al. 2001) will contribute to assessment of the Chinese Han homogeneity hypothesis. Thus, this thesis is designed to provide an interpretive mtDNA pilot analysis, which can be used as a template for the study of ancient DNA remains from Anyang, China.

Background Context of Anyang and Ancient Chinese History

The discovery of the Anyang archaeological site provided researchers with new insights into Chinese ancient culture. The archaeological site is named for the nearby modern city of Anyang, located north of the Yellow River in Henan Province in Northern

China (Figure 1). The archaeological deposits first came to scientific attention in the late 19th century when hundreds of oracle bones were accidentally discovered and reported (Fagan, 1996). Beginning in 1928, excavations conducted by the Chinese Academy of Social Sciences have slowly uncovered the remains of an ancient city that actually encompasses more than 17 sites covering more than 24 square kilometres (Fagan, 1996). The site complex is attributed to the Late Shang Dynasty, and dates approximately 1300-1046 BC (Latourette, 1967). The city's ruins yielded an abundance of architectural and material remains that provide evidence of a rich and complex civilization in ancient China.

Anyang played a significant role in Chinese history. Research at the site has revealed underlying deposits of the earlier Neolithic culture in the region, and demonstrated the early development of an agricultural economy and pottery. However, the appearance of the Shang Dynasty marked a milestone in Chinese cultural history (Chang, 1977). While certain Neolithic traits continued to be in use during Shang times, demonstrating cultural continuity, new innovations developed that mark the appearance of early civilization in China.

The best known deposits at Anyang consist of the elaborate graves and burial offerings associated with the elite of Shang society. The architectural remains at the ritual centre of the Xiaotun city site include foundations of 53 large structures that were likely palaces and temples, while other residences and workshops indicate the way of life associated with the non-elite sectors of society (Fagan, 1996). These latter zones yield evidence of the commoners, and reveal the development of craft specialization within the

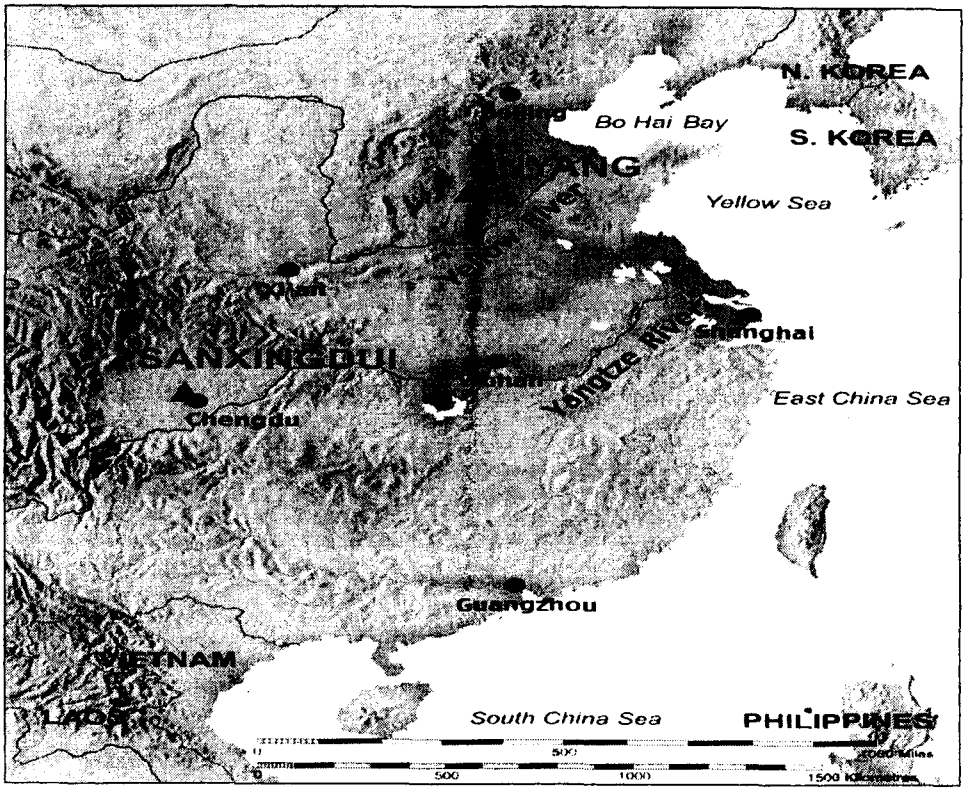
city. This includes bronze-metal working that facilitated the production of a variety of vessels for daily and ceremonial use. Pottery production, jade stone carving and bone working were also important industries.

One of the most important artifact classes recovered at the site consists of the so-called Oracle bones. Of the more than 100,000 inscribed bones recovered, at least 33,500 have been scientifically examined and interpreted (Fagan, 1996). The Oracle bones are a particularly important element of Anyang since they represent the development of a sophisticated writing system that was an important part of ritual divination (Fagan, 1996, Morton, 1980). This literature documents the history of the city and its rulers, and identifies them as the "rulers of Yin", and that the Anyang site is actually Yinxu, the last of about seven capital cities of the Shang Dynasty (Fagan, 1996, Tang et al. 2000). The site clearly reflects the rise of one of the earliest of civilizations, with an elaborate socio-political hierarchy associated with complex burial ritual, massive investments in monumental architecture, intensification of agricultural production, development of metallurgy and a writing system. Clearly, the Shang political elite controlled a large hinterland, and site complex represents one rise of one of the earliest states (Morton, 1980).

Several themes and directions of archaeological research have developed since the first excavations began at Anyang, and have promoted the study of diverse aspects of ancient Chinese culture. While the development of metallurgy and burial ritual are prominent themes in the archaeological literature, attention has increasingly shifted to address all of the status grades making up Shang Dynasty culture. This transformation is

also noted in the nature and orientation of bioarchaeological research at the site. Originally it focused on the elaborate graves of the elites, and the morphological aspects of the skeletal remains recovered in the royal cemeteries at Anyang. In 1998, a collaborative project began between China, the United States and Canada that was designed to undertake a comprehensive analysis of human remains recovered at Hei Heru Road site at Anyang. Under the organizational leadership of Dr. El Molto, the bioarchaeological research included cleaning and sorting of the bones, the development of skeletal inventories, burial reconstructions, vital statistics (age, sex), morphology (metric and non-metric traits), and paleopathology. Samples from the skeletal remains were also collected for analyses of stable isotopes and DNA. Components of the latter research are reflected in Graver's research efforts that are noted above (1999, 2000). The present study seeks to aid in the development of models that will be useful in eventually interpreting the genetic data that will derive from the ancient skeletal remains.

FIGURE 1. Map of China showing the location of Anyang and other major cities (Shen, 2002). In the northern Henan province of China, a late Shang dynasty capital, Yinxu, was discovered in the 1920s near the modern city of Anyang.



CHAPTER 3: RESEARCH DESIGN, MATERIALS AND METHODS

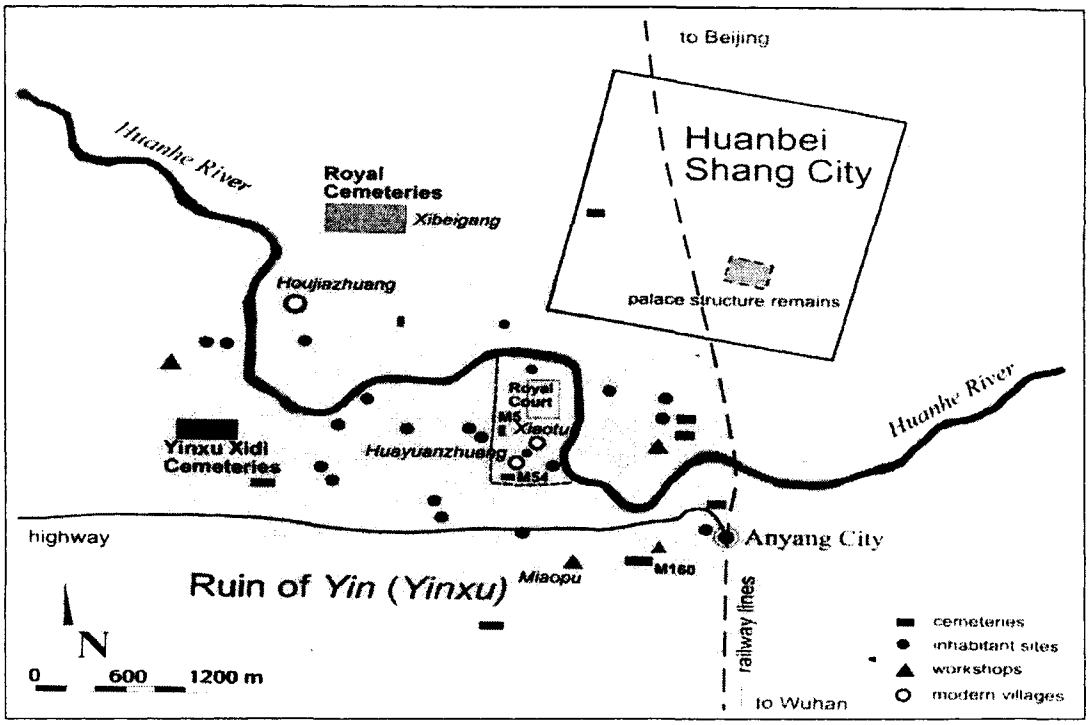
Samples

Head hair samples (n = 30) were collected from three remote villages that presently occupy the ancient Anyang site area. These modern villages have existed for many years and the people trace their ancestry back as far as four hundred years (J. Tang, personal communication). This is important for the present study because, historically, Chinese society was predominantly rural rather than urban (Morton, 1980). The study villages could then represent the genetic profiles of small isolated peasant farming villages with minimal introduction of genes from outside the region. In this way, the genetic information from the study villages can be used to develop a model of rural China in ancient times. Two villages, Huayuanzhuang (HYZ) and Xiaotun (XT) are located in close proximity to one another, and are near the remains of the ancient royal court at Anyang. The third village, Houjiazhuang (HJZ), is located further north in the vicinity of the cemetery grounds (see Figure 2).

Hair was chosen because it is the least invasive type of sampling technique in comparison to other conventional DNA sample sources such as blood and buccal swabs (Allen et al. 1998), and can be easily transported overseas. Moreover, hair has forensic significance due to its ubiquitous nature and is a common evidentiary sample type (Linch et al. 1998, Melton and Nelson, 2001, Sachs, 1997). The hair samples were stored in Ziploc bags at room temperature for approximately three years prior to DNA extraction. The samples represent a total of 30 specimens collected from individuals within discrete

households in each of the three villages. Within each household, three females (grandmother, mother and child) contributed hair. The three individuals who collected the hair also contributed a sample of their own hair for control purposes. The hair samples have the hair shaft and root end attached. Hair colour was mainly dark brown, 26 individuals, while four individuals contributed white hair samples (blind study #'s 5, 7, 16, and 19).

FIGURE 2. Map showing area of Yinxu ruins near the modern city of Anyang, China. The three villages involved in this study, Houjiazhuang (HJZ), Xiaotun (XT) and Huayuanzhuang (HYZ) are noted. HJZ is located near the royal cemeteries, while XT and HYZ are located within the royal court area (Shen, 2002).



Quality Assurance/Quality Control

The control and monitoring of contamination within the laboratory is the main concern for mtDNA amplification using PCR. Guidelines with stringent precautions to minimise the possibility of contamination were employed as outlined by the ancient DNA (Brown and Brown, 1992, Pääbo et al.1985, Richards et al. 1995) and international forensic communities (Alonso et al. 2002, Capelli et al. 2003, Carracedo et al. 2000). For instance, all disposables such as tubes and tips were autoclaved and UV irradiated prior to use. Reagents were UV irradiated prior to use, if appropriate, as it has been shown to reduce contamination in PCR (Ou et al. 1991). Forceps and racks were autoclaved and UV irradiated. All extractions and PCR amplifications included blank reactions (no DNA added) that were run in parallel to ensure reagent quality. Workspaces and dedicated hoods were cleaned before and after each sample was prepared. Gloves were changed between each sample extraction. The extraction procedure took place in a clean room facility physically separated from any post PCR analysis. The mtDNA profiles and the sequence for all laboratory staff are on file for comparative reference to check for intra-laboratory sources of contamination.

Finally, sequencing results were analysed and interpreted according to established standards (Röhl et al. 2001, Tully et al. 2001, Wilson et al. 1993). The standardisation of DNA sequencing nomenclature and the interpretation of results are important for proper identification of maternal relatives. Recommendations from the forensic community were used in this study. The Revised Cambridge Reference Sequence (RCRS, Anderson et al. 1981, Andrews et al. 1999) was used for comparative purposes and polymorphism

calls (Tully et al. 2001). Criteria for determining heteroplasmy were taken from Tully et al. (2001). For instance, two bases present at a single site should be at approximately equal intensities, although 20% peak height may be enough if the peak can be differentiated from background noise. The sequencing of both strands ensures high quality sequence to differentiate heteroplasmy from sequence noise/artefacts. "Hot spots" for heteroplasmy, such as C-stretch regions, are noted.

DNA Extraction: Modified Wilson et al. (1995)

Two centimetres of the shaft portion of a single hair from each sample were placed in a 1.5mL tube containing 5% Terg-a-zyme™ detergent (Alconox Ltd., Jersey City, N.J.). The tubes containing the hair were sonicated for 20 minutes in a water bath to remove surface debris. The hair was then rinsed once in a 95% ethanol solution followed by ddH_20.

Before the hair was transferred to a sterile 0.2mL micro-tissue grinder (Kontes Glass, Vineland, N.J.), a reagent blank was prepared for each tube. A sterile pestle was placed in the tube with 200μL of extraction buffer (TE: 10mM Tris-HCl, 0.1mM EDTA pH 8.0) and grinding simulated by moving the pestle up and down. The 200μL of extraction buffer was transferred to a new tube, and this blank reaction was run in parallel along with the actual hair sample. Then the 2cm hair sample was placed in the glass tube with another 200μL of extraction buffer and ground up until no fragments could be seen. The DNA containing supernatant was transferred to a new sterile tube and 1μL (20mg/mL) of Proteinase K (QIAGEN) was added. The tubes were vortexed and

centrifuged before being placed in an Eppendorf Thermomixer® for overnight incubation at 56°C, under medium agitation.

Following overnight incubation, the tubes were briefly centrifuged to bring down the condensate. 200μL of phenol/chloroform/iso-amyl alcohol (PCIA, 25:24:1, Sigma Chemicals) was added to each tube and vortexed for 30 seconds. The top aqueous layer (approximately 200μL) was transferred to a prepared Microcon-100YM column (Millipore) for purification following manufacturer's recommendations. The purified product was then ready for PCR.

PCR Amplification

The mitochondrial DNA area subjected to analysis was Hypervariable Region 1 and 2 of the control region. Hypervariable Region 1 (HV1) and Hypervariable Region 2 (HV2), according to Anderson et al. (1981, GenBank, Accession #J01415), are nucleotide positions 15997-16401 for HV1 and 73-340 for HV2. One set of primers was used to amplify 415bp of HV1, 15996f (5'CTCCACCATTAGCACCCAAAGC3') and 16410r (5' TGATTTCACGGAGGATGGTG 3'). Another set of primers was used to amplify 375bp of HV2, 15f (5' CACCCTATTAACCACTCACG 3') and 389r (5' CTGGTTAGGCTGGTGTTAGG 3'). These hypervariable regions have been used successfully to identify maternally related individuals and the primers have been previously reported in the literature (Vigilant et al. 1989, Wilson et al. 1995).

PCR amplification was performed in 50μL reactions using an Eppendorf Mastercycler®. Separate PCR reactions were done for HV1 and HV2. Following the

company's recommendations for HotStarTaq™ (QIAGEN), each reaction contained 5μL of 10x PCR Buffer, 1μL of 10mM dNTP's (Stratagene), 2μL of each 10mM primer (Operon Technologies), 0.25μL of HotStarTaq (5 units/μL), 5-10μL of template DNA and ddH_20 to volume. The thermocycling conditions were as follows: hot start at 95°C for 15 minutes, followed by 35 cycles of 94°C for 30 seconds, 60°C (HV1) or 56°C (HV2) for 30 seconds, and 72°C for 1 minute and a final extension of 72°C for 10 minutes. PCR reaction products were visualized by gel electrophoresis on 5% polyacrylamide gel electrophoresis (PAGE) with ethidium bromide staining.

Sequencing

In preparation for sequencing, PCR products were first purified using Performa™ DTR Gel Filtration Cartridges (Edge Biosystems, Gaithersburg, MD) to remove unincorporated dNTP's and other low molecular weight materials from the sequencing reactions. Sequencing reactions were accomplished with a Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's recommendations. The same primer sets were used for sequencing of HV1 and HV2 with sequencing accomplished on an ABI 310 and 3100 (Applied Biosystems). Sequencing was processed in both forward and reverse directions in order to reduce electropherogram ambiguities, such as high background signal and heteroplasmy. Finally, mtDNA variability was reported according to standardized nomenclature for base calling (Carracedo et al. 2000, Röhl et al. 2001, Wilson et al 1993).

CHAPTER 4: RESEARCH RESULTS

DNA was successfully extracted and sequenced for all 30 individual hair samples for HV1 and 28 samples yielded data for HV2. Each genetic profile was confirmed by both a forward and reverse sequence. The forward and reverse sequences for each sample were then assembled together and edited for base calls. Polymorphic sites were determined for each hair sample by direct comparison with the rCRS (Accession #J01415, Andrews et al. 1999) obtained from GenBank (Benson et al. 2000) using Sequencher Software Version 4.0.5. (Gene Codes Corp., Ann Arbor, MI). Tables 1 and 2 respectively list the polymorphic data for HV1 and HV2 for the individuals in the study.

For HV1 (see Table 1), there were 46 mutations noted among the mtDNA haplotypes or lineages. The majority of these (n=32) were unique to a specific haplotype, while 14 were shared polymorphisms among the haplotypes. The most common polymorphism, nucleotide position (np) 16223T occurred in 11 of 16 lineages (Lineages 1-5, 8-10, 12, 13, and 16). The next most common mutation, at np 16362C, occurred in 7 of 16 lineages (3, 4, 8-10, 13, and 16). The two mutations, which result in a homopolymer C-stretch in HV1, np 16183C and 16189C, were shared between the same five lineages (2, 3, 6, 7, and 11). Np 16304C was shared by four lineages (4, 6, 14, and 15). Four polymorphisms were shared in three lineages. Np 16234T was found in lineages 3, 4, and 7; 16298C in lineages 1, 5, and 12; and np 16311C in lineages 3, 6, and 8. Np 16319A was present in lineages 1, 9, and 13. The following five polymorphisms were shared between two lineages: np 16217C (Lineages 7, and 11); 16248T (Lineages 4, and 6); 16249C (Lineages 6, and 7); 16290T (Lineages 9, and 13) and 16291T (Lineages

8, and 14).

Mitochondrial sequence variation was assessed for HV1 only, since some data were missing for HV2. HV1 is considered the most variable area of the control region (Anderson et al. 1981) and therefore is the most useful for differentiation of matrilineages. From the 30 individuals, a total of 146 polymorphisms were recorded, which included 132 transitions, 12 transversions, two deletions and no insertions. Consistent with other mtDNA studies (Budowle et al. 1999, Lee et al. 1997, Vigilant et al. 1991), transitions outnumbered transversions. The respective relative frequencies of transitions, transversions, and deletions were 90.4%, 8.2% and 1.3%. Transitions were categorized by base type, namely purine (G/A) and pyrimidine (C/T). Pyrimidine transitions were more common than purine transitions. Of the 132 transitions, 114 were pyrimidines, while only 18 were purines. Of the 12 transversions, seven occurred at 16183C, three at 16092A, one at 16265T, and one at 16362A. Two deletions were recorded at np 16182 and 16293.

HV2 sequencing of the 30 samples was more problematic than that for of HV1. Forward and reverse sequences were successful for 28 samples, but samples (5, and 21) repeatedly provided poor quality data and were consequently unreadable (see Table 2). As well, sequencing of all of HV2, from np 73 to 340 was successful for 25 of 28 samples. Three samples (8, 9, and 10) repeatedly had unresolved individual bases due to high background noise before nucleotide position 100 and consequently, the first 35 bases of were omitted in the results.

Table 3 lists the results of the blind study in the identification of single

individuals, pairs and sets of maternally related individuals. Maternally related individuals were identified when the mtDNA sequences matched. For all hair samples, it was possible to generate mtDNA sequences. The data show six individuals who had unique mtDNA profiles compared to all other samples. As well, six pairs of individuals were identified as maternal relatives, and four groups, consisting of three individuals, were considered maternal relatives because of matching mtDNA sequences.

Upon completion of the blind study, the identification of the samples was provided by Dr. Molto and witnessed by Dr. Carney Matheson. This information allowed the blind study samples to be identified according to village (i.e. XT, HYZ, HJZ), household (i.e. A, B, C) and generation (i.e. grandmother, mother, child). Tables 4, 5, and 6 show the correlation of the blind study samples with exact identifications.

Table 3 and Table 4 facilitate a more complete understanding of the genetic relationship of each individual to their household and village. In the blind study a total of six individuals could not be maternally matched to anyone in the composite sample (Table 4). These unmatched individuals likely represented five grandmothers from different villages (5, 7, 12, 16, and 19) and one collector (3) from the village of HYZ. Six pairs of maternally related individuals were matched (Table 5). Four of six were mother-child pairs (28-27, 10-18, 20-23, and 24-6), one of six was a grandmother-mother pair (2-22), and one of six was a grandmother-collector pair (21-30). Four sets of three maternally related individuals were identified (Table 6). The first set was grandmother-mother-child (25-17-29) from the same household (C) in village HYZ. The second set was mother-child-collector (9-13-11) from village XT. The third set was a grandmother

(1) from household B related to a mother-child pair (8-26) from household C from village XT. The fourth set was mother-child pair (4-14) from HJZ village related to a child (15) from HYZ village.

Of interest was that there were no unequivocal cases of heteroplasmy in this study sample. Direct sequencing of forward and reverse mtDNA strands cleared up any ambiguities in the sample set. According to standards outlined by the international forensic community, heteroplasmy could not be reported.

In summary, seven of nine grandmothers did not match mothers or children from the same household. Eight of nine mothers and children from the same household were maternally related. Only one of the nine households contained a matrilineage consisting of all three generations. There was one instance where the grandmother-mother pair (2-22) was not maternally related to the child (15) in the same household. Two of the collectors were maternally related to one of the households from which they sampled. A total of 16 distinct lineages within nine households from three villages were identified.

TABLE 1. Variable sites of mtDNA hypervariable region 1 in 30 individuals from three villages in the Anyang region of China, with respect to nucleotide position 16000-16401 of the reference sequence (rCRS, Anderson et al. 1981, Andrews et al. 1999). Differences from consensus sequence are noted. "d" refers to deletions, "." indicates no change from rCRS. Nucleotide positions with a decimal (i.e. 15465.1) refer to insertions.

NUCLEOTIDE POSITIONS

	16086	16092	16093	16129	16136	16140	16150	16154	16162	16164	16172	16182	16183	16184	16185	16186	16189	16192	16201	16217	16218	16223	16232	16234	16248	16249	16260	16265	16266	16270	16274	16278	16290	16291	16293	16294	16295	16297	16298	16299	16304	16311	16316	16319	16327	16362
RCRS	T	T	T	G	T	T	C	T	A	A	T	A	A	C	C	C	T	C	C	T	C	C	C	C	C	T	C	A	C	C	G	C	C	C	A	C	C	T	T	T	A	G	C	T
1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	T	.	.	.	.	.	.	A	.	C
2	.	.	.	.	.	C	T	.	.	.	.	.	C	.	.	.	C	.	.	C	.	.	.	T	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	d	.	.	C	C	.	.	.	.	A	.	.
4	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.
5	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
6	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.	.	C
7	.	.	.	.	.	.	.	.	.	G	.	d	C	.	.	T	C	.	.	.	.	T	.	T	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	C
8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	A	.	C
9	.	.	.	A	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	A	.	C
11	.	.	.	A	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	C	.	G	.	.	C
13	.	.	.	A	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
14	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.
15	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.
16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	A	.	C
19	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	.	.	A	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.
20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
21	.	.	.	.	C	.	.	C	.	.	.	.	C	.	.	.	C	.	.	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
22	.	.	.	.	.	C	T	.	.	.	.	.	C	.	.	.	C	.	.	C	.	.	.	T	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
24	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.	.	C
25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
26	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	T	.	.	.	.	.	.	A	.	C
27	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
28	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
30	.	.	.	.	C	.	.	C	.	.	.	.	C	.	.	.	C	.	.	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

TABLE 2. Variable sites of mtDNA hypervariable region 2 in 30 individuals from three villages in the Anyang region of China, with respect to nucleotide position 73-340 of the reference sequence (rCRS, Anderson et al. 1981, Andrews et al. 1999). Differences from consensus sequence are noted. "*" refer to areas where sequencing data was not achieved, "d" refers to deletions, "n" refers to no base call, "." indicates no change from rCRS. Nucleotide positions with a decimal (i.e. 309.1) refer to insertions.

NUCLEOTIDE POSTIONS

	73	143	146	150	151	152	153	182	194	204	235	249	263	309.1	309.2	315.1
RCRS	A	G	T	C	C	T	A	C	C	T	A	A	A	.	.	.
1	G	.	.	.	.	C	.	T	.	.	G	.	G	C	.	C
2	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C
3	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C
4	G	.	.	.	.	.	.	.	.	.	.	d	G	.	.	C
5	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
6	G	.	.	.	T	C	.	.	.	.	.	.	G	.	.	C
7	G	.	C	T	.	.	.	.	.	.	.	.	G	C	.	C
8	*	.	.	.	.	C	.	T	.	.	G	.	G	.	.	.
9	G	.	.	.	.	.	.	.	.	.	.	d	G	C	C	C
10	*	.	.	T	N	C	.	.	.	.	G	.	G	T	.	.
11	*	.	.	.	.	.	.	.	.	.	.	d	G	C	C	C
12	*	.	.	.	.	.	G	.	.	.	.	.	G	C	T	C
13	G	.	.	.	.	.	.	.	.	.	.	d	G	C	.	C
14	G	.	.	.	.	.	.	.	.	.	.	d	G	.	.	C
15	G	.	.	.	.	.	.	.	.	.	.	d	G	.	.	C
16	G	.	.	.	.	C	.	.	.	.	.	d	G	.	.	C
17	G	.	.	.	.	.	.	.	T	.	.	.	G	.	.	C
18	G	.	.	T	T	C	.	.	.	.	G	.	G	C	.	C
19	G	.	.	.	.	.	.	.	.	C	.	d	G	C	C	C
20	G	A	.	.	.	C	.	.	.	.	.	.	G	.	.	C
21	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
22	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C
23	G	A	.	.	.	C	.	.	.	.	.	.	G	.	.	C
24	G	.	.	.	T	C	.	.	.	.	.	.	G	.	.	C
25	G	.	.	.	.	.	.	.	T	.	.	.	G	.	.	C
26	G	.	.	.	.	C	.	T	.	.	G	.	G	C	.	C
27	G	.	.	.	T	C	.	.	.	.	.	d	G	C	.	C
28	G	.	.	.	T	C	.	.	.	.	.	d	G	C	.	C
29	G	.	.	.	.	.	.	.	T	.	.	.	G	.	.	C
30	G	.	.	T	.	.	.	.	.	.	.	d	G	C	C	C

TABLE 3. A summary of maternally related individuals in the blind study of 30 individual hair samples from individuals in remote villages near Anyang, China. Blind study samples are categorized according to single individuals, and sets of two or three maternally related (MR) individuals. The Maternal Lineages are numbered sequentially to represent the different mtDNA profiles. The number of HV1 and HV2 shared and unique sites are listed.

Number of MR	Blind Study Numbers	Maternal Lineage	HV1 Polymorphisms		HV2 Polymorphisms	
			Shared (14)	Unique (32)	Shared (10)	Unique (6)
1	3	1	3	4	3	0
	5	2	3	2	N/A	N/A
	7	3	6	5	5	1
	12	4	5	3	4	1
	16	5	2	1	5	0
	19	6	6	1	6	1
2	2-22	7	5	2	3	0
	6-24	8	4	1	5	0
	10-18	9	4	0	8	0
	20-23	10	2	2	4	1
	21-30	11	3	3	7	0
	27-28	12	2	2	7	0
3	1-8-26	13	4	1	6	1
	4-14-15	14	2	2	4	0
	9-11-13	15	1	3	6	0
	17-25-29	16	2	0	3	1

TABLE 4. Blind study results of unmatched individuals compared with confirmed identifications according to village, household and generation. Six individuals were identified that could not be matched to anyone in the study group. These individuals, 19, 7, 5, 16, 12, and 3 are highlighted.

	VILLAGES									
	Xiaotun (XT)			Huayuanzhuang (HYZ)				Houjiazhuang (HJZ)		
HOUSEHOLD	A	B	C	A	B	C	Other	A	B	C
Grandmother	19	1	7	5	2	25		21	16	12
Mother	9	24	8	20	22	17		28	10	4
Child	13	6	26	23	15	29		27	18	14
Collector	11						3	30		

TABLE 5. Blind study results of pairs of maternally related individuals compared with confirmed identifications according to village, household and generation. Six pairs of maternal relatives were identified: 6-24, 20-23, 2-22, 278-28, 21-30, and 10-18. The maternal pairs are highlighted.

	VILLAGES									
	Xiaotun (XT)			Huayuanzhuang (HYZ)				Houjiazhuang (HJZ)		
HOUSEHOLD	A	B	C	A	B	C	Other	A	B	C
Grandmother	19	1	7	5	2	25			16	12
Mother	9	24	8	20	22	17		28	10	4
Child	13	6	26	23	15	29		27	18	14
Collector	11						3			

TABLE 6. Blind study results of sets of three maternally related individuals compared with confirmed identifications according to village, household and generation. In total, four sets of three maternal relatives were identified: 9-11-13, 1-8-26, 17-25-29, and 4-14-15 are highlighted.

	VILLAGES									
	Xiaotun (XT)			Huayuanzhuang (HYZ)				Houjiazhuang (HJZ)		
HOUSEHOLD	A	B	C	A	B	C	Other	A	B	C
Grandmother	19	1	7	5	2	25		21	16	12
Mother		24	8	20	22	17		28	10	
Child		6	26	23		29		27	18	
Collector							3	30		

CHAPTER 5: DISCUSSION AND CONCLUSION

Discussion

This thesis utilizes a proven mtDNA extraction protocol for hair developed by the FBI (Wilson et al. 1995). A total of 30 head hair samples from modern Chinese from three villages in the Anyang region, Henan Province, constituted the composite sample. These data generated in a 'blind' design to ensure unbiased sample analysis and international forensic standards were implemented for the interpretation of mtDNA data (Wilson et al. 1993). The non-invasive sampling of hair is an effective source of DNA, but some studies have shown that hair tends to be more variable in terms of mtDNA mutations than other tissues (Alonso et al. 2002, Bendall et al. 1997, Carracedo et al. 2000, Wilson et al. 1997). Included in this hypothesized variability for hair is a higher rate of heteroplasmic loci. As noted above, heteroplasmy was, however, not found in the 30 samples.

To assess whether this inter-sample variation would be detectable over generations, this study followed a longitudinal sampling design. The study group included grandmothers, mothers and children from nine different households, equally drawn from three villages. As well, three of the samples represented collectors who were from the villages where the samples were obtained. The main objectives were to test mtDNA lineage integrity from hair in terms of assigning individuals to defined mtDNA lineages and to provide a baseline model for interpreting ancient DNA results from the Anyang region. A corollary to the latter objective was to determine if the haplotype variability found in these modern Anyang samples supports the recent 'homogeneity hypothesis'

advanced by Oota et al. (2002).

The technical aspects of mtDNA extraction from hair have been standardized enough that a high rate of success was obtained in generating quality data in this study. Only minor difficulties arose where PCR amplification and sequencing data were not obtained. Of methodological significance, when working with hair samples in particular, the analyst should note hair characteristics, specifically colour and structure. Lack of hair pigmentation (melanin = colour) may compromise mtDNA extraction from such hair. In my view this problem arose because when white hair is wet in an extraction buffer, it is difficult to visualize and consequently, the piece of hair can be lost or not be sufficiently ground up. Grinding hair is a particularly important step in the extraction process. The major structure of hair is keratin and keratin fibers, which are thick and chemically resistant (Hashimoto, 1998). As well, melanin increases rigidity to hair and proteins may protect it from degradation (Riley, 1997). Sufficiently isolating DNA from keratinized hair cells increases the chance of a successful amplification (Hellmann et al. 2001). Consequently, insufficient grinding of the hair may not release DNA bound in the shaft and may result in poor DNA yield. This was found for one white hair sample (5) in this study. It had a lower DNA yield or didn't amplify at all after several PCR attempts. An explanation for the difficulty with the white hair sample is that unpigmented hair (i.e. white hair colour) lacks melanin, and therefore, theoretically, the DNA is less protected from environmental degradation, such as solar radiation, than pigmented hair. White hair lacks the absorbent protection of melanin and is subsequently unprotected against UV light that is known to cross-link DNA (Riley, 1997). Damage caused by the generation of

Reactive Oxygen Species (ROS) in aging melanocytes may also cause mutations to mtDNA (Tobin and Paus, 2001). The impact of this methodological problem appears to be limited since it (Sample 5) was the only white hair sample that was problematic in terms of DNA yield. Despite this, HVI data were still obtained. This is a precautionary methodological point, which could have a bearing on sample selection in forensic cases or in other longitudinal mtDNA research.

The interpretation of mtDNA data generated with automated sequencers has become the standard in forensic laboratory casework. Most laboratories carry out sequencing in forward and reverse directions, and report data according to accepted nomenclature (Alonso et al. 2002, Carracedo et al. 2000, Wilson et al. 1993). This standardization has allowed inter-laboratory testing of samples done in "blind" to check the level of quality assurance and the proper interpretation of genetic data. Results from inter-laboratory DNA testing have shown that overall the error rate in participating labs was low, and that sources of the error were identified during sequencing analysis and interpretation, such as transcription errors, methodological problems (i.e. sequence artifacts) and heteroplasmic variability (Alonso et al. 2002). In this study, these types of errors were taken into consideration for sequencing analysis and interpretation of the 30 samples. Presumptive identification of maternally related individuals was not confounded by any difficulties listed previously, but without following the standards of the international forensic community, sequencing interpretation may become easily compromised if the standards of the international forensic community are not followed. For example, at np 248A and 249A, an adenine deletion at either site could be improperly

recorded. Guidelines in Röhl et al. (2001) suggest recording the deletion at np 249 rather than np 248. In this study, deletions originally recorded at np 248 were corrected to np 249 in order to maintain continuity within these published guidelines. This brief example of how errors can be generated in sequencing analysis and interpretation show how important it is to be current with accepted guidelines in the DNA community, especially when contributing or comparing to available DNA databases (Dennis, 2003)

Conforming to international guidelines for reporting mtDNA data is not only useful for recognizing where recording errors can occur, but also improves interpretation of control region variation. The control region of the mitochondrial genome has been utilized extensively in anthropology and forensic science (Bender et al. 2000) to determine maternal relatedness between individuals. The control region contains two principal areas, HV1 and HV2 that include sufficient variation to make presumptive identification of related individuals possible (Parsons et al. 1997). HV1 is reportedly more variable than HV2, in that HV1 generally has more polymorphisms than HV2. This study supports this finding, as HV1 had more polymorphisms than HV2: 46 versus 19, respectively.

Another level of variation that has been described pertains to the homopolymer stretch, also known as the C-stretch in HV1 (np 16184-16193) and HV2 (np 303-315). These C-stretches are associated with one or more cytosine insertions that produce a length heteroplasmy (Bendall and Sykes, 1995, Stewart et al. 2001, Stoneking, 2000). In HV1, the transversion at np 16183C and transition at 16189C occurred in the same five mtDNA lineages from this study creating a C-stretch of 11 cytosines. Though readable

sequence data following a C-stretch were not obtained, replication and priming inside the C-stretch produced reliable data, and identifications were not affected. On the other hand, the C-stretch in HV2 was more problematic. For instance, all maternally related individuals matched at all sites within HV1, while maternal relatives (identifiable by HV1) had different polymorphisms between these relatives in HV2. In particular, the homopolymer stretch in HV2 has a significant amount of variation between maternally related individuals. There were three lineages that had differences in the HV2 homopolymer stretch between np 303 – 315 (rCRS, GenBank, Accession #J01415). Lineage 9 made up by a mother (10) and a child (18) differed in the length polymorphism. The mother had a T insertion at np 310 while the child had C insertions at np 309.1 and 315.1. This same type of variation is seen in Lineage 13, where the mother (8) had a T at np 310, while the grandmother (1) and child (26) had C insertions at 309.1 and 315.1. Another related variation within the homopolymer stretch occurred in Lineage 15. The mother (9) and collector (11) had C insertions at 309.1 and 309.2, while the child lacked the insertion at np 309.2. These types of C-stretch variations have been seen in other studies (Stewart et al. 2001, Tagliabracci et al. 2001). Though there were sequence variations, such as differences in number of C's in the C-stretch or variation in the relative amounts of heteroplasmic length variants between presumptively identified maternal relatives, the individuals were not excluded from the lineage (Stewart et al. 2001). Minor sequence variation, including one or more site differences and levels of heteroplasmy, may not necessarily exclude individuals from a maternal lineage (Alonso et al. 2002). Cali et al. (2000) suggest that due to the accelerated mutation rate and

consequent increased genetic drift of mtDNA control region, precise matches may not survive long between related samples.

The longitudinal mtDNA analysis of three generations within nine households from three remote villages in China was successful at two levels: it successfully identified maternally related individuals, and also enabled interpretation of marriage practices. When all the data are considered, a genetic pattern emerges within the households. As noted, I found that a majority of grandmothers were not genetically related to maternally related mothers and children in the same household.

Not until the genetic information is evaluated in a cultural context does the pattern make sense. Oota et al (2001) reported that human mtDNA genetic diversity is strongly correlated with residence patterns. As culture determines societal patterns of belief and economy, a possible explanation for this genetic pattern can be related to the Chinese practice of patrilineality and patrilocality. In Chinese culture, ancestry is primarily traced through the male line back to a common ancestor. Thus, sons and daughters trace their descent through their father, grandfather and great grandfather (Haviland, 1996). Furthermore, there is a general practice, especially in rural China, for a patrilocal system of residency. Patrilocality is defined by the movement of newly married women from their place of birth to the house of her husband and his family. It is common for men to live with their wives and children in the house of his mother and father, and this can even include his brothers and their families (Haviland, 1996). Thus, these rural villages with a long history in the region of Anyang appear to continue a patrilineal system of descent with individual households following a patrilocal pattern of residency.

Taking into consideration the mtDNA data and residence patterns, I developed a model of matrilineal relationships within a patrilocal household and matrilineal relationships within a matrilocal household. In a patrilocal residence (Figure 3), it could be expected that at least three maternal lineages would exist within a three generational household. First, the grandfather would not be maternally related to anyone else in the house. Second, the grandmother would be maternally related to her children (i.e. sons and unmarried daughters). Third, the son's wife would have a different maternal lineage because she comes from outside the household, and her children would be maternally related to her. If more than one son resided with the grandparents, then this pattern would be further complicated by the addition of daughters-in-law and her children. This maternal genetic pattern is what is observed seven of nine households from the remote villages in China. That is, grandmothers were not related to mothers and children from the same household. The pattern in villages XT and HJZ was observed for all households tested, while only one household in village HYZ appeared to follow a patrilocal pattern.

The two remaining households in village HYZ differed in their household makeup and may appear to follow a matrilocal residency pattern (see Figure 3). A matrilocal residence pattern will differ from a patrilocal one in that a mtDNA lineage is continuously represented throughout generations. The preservation of mtDNA lineages is maintained when females remain in their place of birth, while males move to their wives' houses (Haviland, 1996). In household C, HYZ, all three female generations had the same mtDNA type. The other household, B (HYZ), the grandmother and mother (2-22) were related. Although the child (15) was not related to the grandmother-mother pair, and this

will be addressed below, this household conforms to a matrilocal residency pattern. Although these two families appear to follow a matrilineal residency pattern, without further sampling of men from the villages and households, it is impossible to say for certain the type of residency. There are other explanations may explain their household situations. For instance, household C (HYZ), a situation where a daughter lost her husband returns to her mother's household with her child, or has a child out of wedlock could show three generations of maternal relatives. Thus, the mtDNA analysis of matrilineal relationships in different residences can facilitate interpretation of human genetic variation under the influence of sex-specific roles among groups.

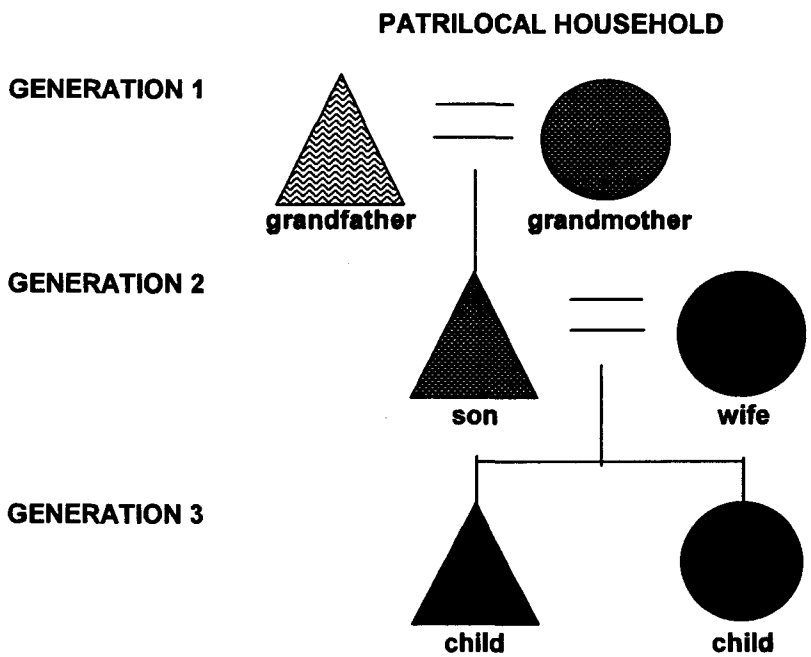
In summary, a total of 16 distinct matrilineages were identified from nine households from the three villages. This unexpected mtDNA lineage diversity arose because of a design error, namely that the sample collectors did not differentiate between maternal and paternal grandmothers. Ultimately, the error was made during the sample arrangements as Dr. Molto and Dr. Jigen Tang did not clearly discuss the complexities of mtDNA transmission, and how it is affected by Chinese kinship and household organization (E. Molto, personal communication, March, 2002). Chinese society is highly patrilineal and characterized by patrilocal residence (Haviland, 1996). Seven mismatches occurred between mothers and grandmothers because, of course, the grandmothers were from different mtDNA lineages than the mothers (i.e. daughters-in-law) and grandchildren. This result alone supports the use of mtDNA to identify relationships and constitutes a significant research finding of this thesis. Oota et al. (2001) reports more mtDNA variation in a patrilineal and patrilocal society, such as China. The pattern of

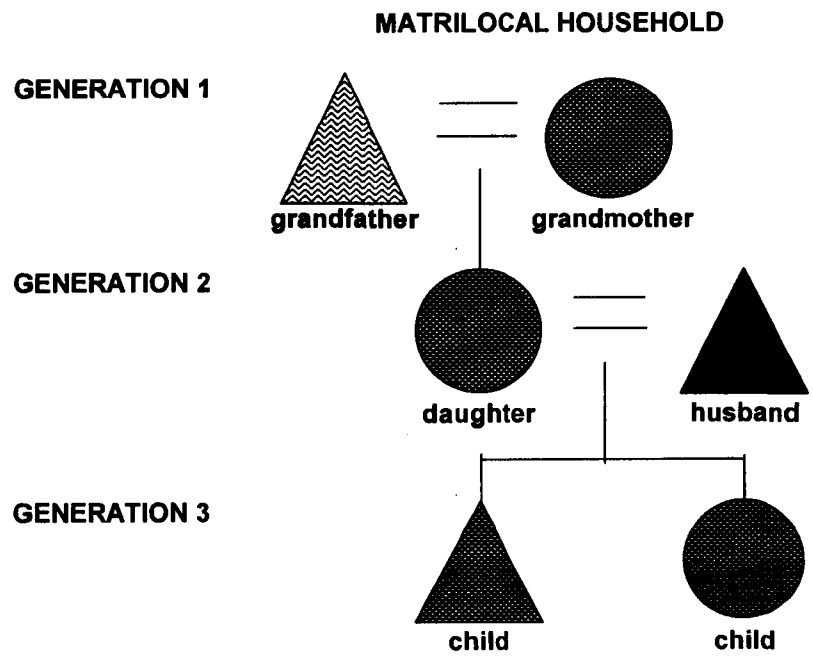
lineage diversity in this study appears to fit with the Oota et al. (2001) model.

The circumstance involving a child (15) from one village who is maternally related to a mother-child pair (4-14) in another village is unusual. Taking everything into consideration to try and account for this anomaly, sequencing data was re-evaluated. Upon closer inspection, these three individuals had no electropherogram or sequence ambiguities, and all profiles matched at every nucleotide position. In fact, the individuals shared a unique transversion polymorphism at np 16092A, a type of mutation that is less common and more conservative. Two alternative explanations may offer plausible explanations for this situation. Firstly, adoption in Chinese culture is common. This child living in another village (HYZ) resides with a grandmother-mother pair (2-22). This maternally related grandmother and mother is one of two cases where a grandmother and mother in the same household were maternally related. It is possible that the mother was unable to have children and adopted the child. Adoption was not uncommon in Chinese culture, but it was more frequent for a childless family to adopt a boy in order to continue on the patrilineage (Latourette, 1967). Secondly, sampling error may have contributed to this unusual situation. Collector 3 from HYZ was not maternally related to any of the families from which he sampled, unlike collectors 11 and 30 who were related to families in their villages. Due to his possible unfamiliarity with the village residencies, perhaps the collector took a sample from the child (15) from that household (HYZ-B) without knowing the relationship of that child to the rest of the family members. It is possible that the child (15) was visiting relatives in the neighbouring village and was accidentally included in the study. As well, sampling error could be attributed to incorrect recording

of the initial samples, whereby they were mixed up. Consequently, due to insufficient data about the sample collection, or particulars related to individuals in the study, I can only speculate about the causes of unexpected or ambiguous

FIGURE 3. Model of maternal relationships in patrilineal and matrilineal households with three generations. Circles represent males and triangles represent females. Different patterns of colour represent distinct mtDNA lineages within each household.





results. That being said, the quality controls built into the DNA extraction and analysis process make it clear that the genetic profiles accurately reflect the nature of the biological materials submitted for analysis.

The ancient DNA pilot study by Graver (1999) and Graver et al. (2000) produced sequencing results for a small number of skeletal remains from the ancient site at Anyang, China. Three individuals, 610A, 610B and a chariot driver were analyzed for a portion of the HV1 region of the mitochondrial genome. Due to the degraded nature of the mtDNA, only 50 to 150 bases were resolved in HV1 for genetic profiling. Individuals 610A and 610B shared a polymorphism at np 16223T. The chariot driver sample had three polymorphisms, np 16224C, 16311C, and 16324C. When the sequencing data from the ancient project is compared to the modern Anyang samples, we find that two polymorphisms are unique to the ancient data, while two polymorphisms also occurred in the modern data. Np 16224C and 16324C are unique to the ancient study and are not reported in the modern mtDNA data set. The common polymorphisms, np 16223T and 16311C are shared with 11 and 3 lineages respectively. This mtDNA variation in Chinese ethnic populations may reflect this area's ethnohistory, where genetic traces are preserved over time (Yao et al. 2002). Interestingly, when the ancient mtDNA data from Anyang was compared to other modern population studies, we find that some polymorphisms are still present in modern mtDNA lineages. For instance, three polymorphisms, np 16223T, 16311C and 16324C were in common with data reported by Yao et al. (2002), but np 16224C was not represented in their Chinese and Thai data set at all. In contrast, a study of a modern Korean population by Lee et al. (1997) reported all

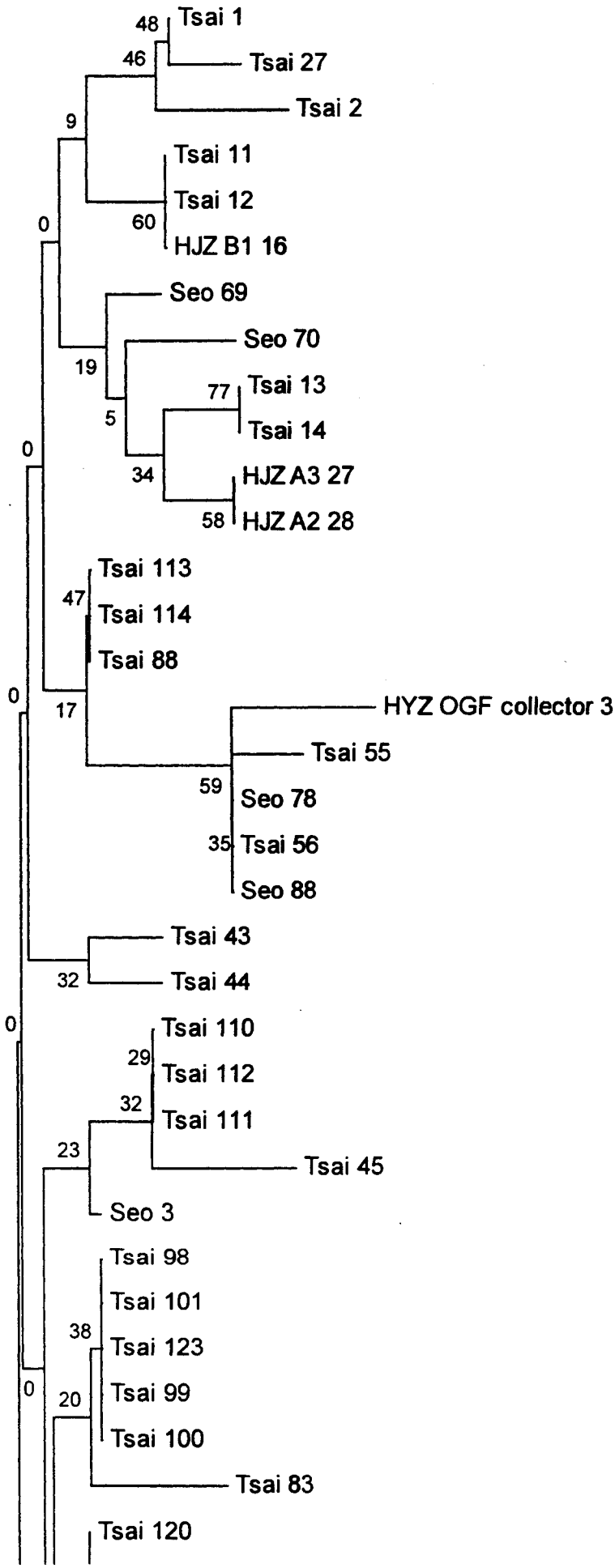
four polymorphisms that occurred in the ancient DNA study. Regardless, further mtDNA testing of ancient samples from Anyang is necessary in order to understand the loss or preservation of polymorphisms from past to present populations. Thus, we see that some polymorphisms may be conserved over time within the Anyang region, while other ones may be lost over time.

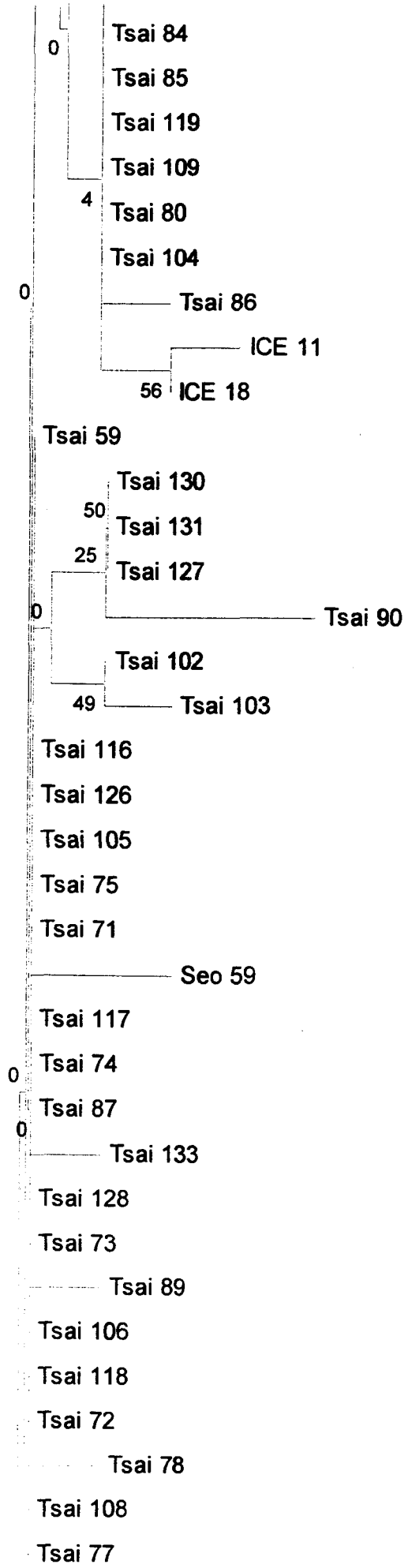
In order to evaluate the Anyang data set within a larger geographic and population area, the Anyang mtDNA data was analyzed along with two other Asian mtDNA data sets (see Appendix). These data sets included 155 individuals of Taiwanese Han descent (Tsai et al. 2001), 149 mtDNA individuals of Japanese descent (Seo et al. 1998), and 30 individuals from Anyang, China. An out-group was chosen from Helgason et al. 2000 which consisted of 30 individuals self-identified as Icelandic (NCBI, Accession AF236888-AF236917). Genetic analyses were conducted using MEGA version 2.1 (Kumar et al. 2001). The mtDNA region for phylogenetic analyses was HV1, np 16021 to 16401 (RCRS, Anderson et al. 1981, Andrews et al. 1999). An unrooted phylogenetic tree was generated with a Neighbor-joining method (1000 bootstrap replications). In general, the Anyang data is randomly dispersed among the larger Asian data sets. There are no lineages from Anyang grouping together by village, other than those individuals who were identified as being maternally related. The out-group of Icelandic sequences cluster together off one main branch, although some Asian mtDNA types do fall into the out-group main cluster. An explanation may be that mtDNA division of major population groups is less specific when only HV1 data is compared. Ideally, the use of HV1 and HV2 may clarify the overlap of major populations. Unfortunately, limited available

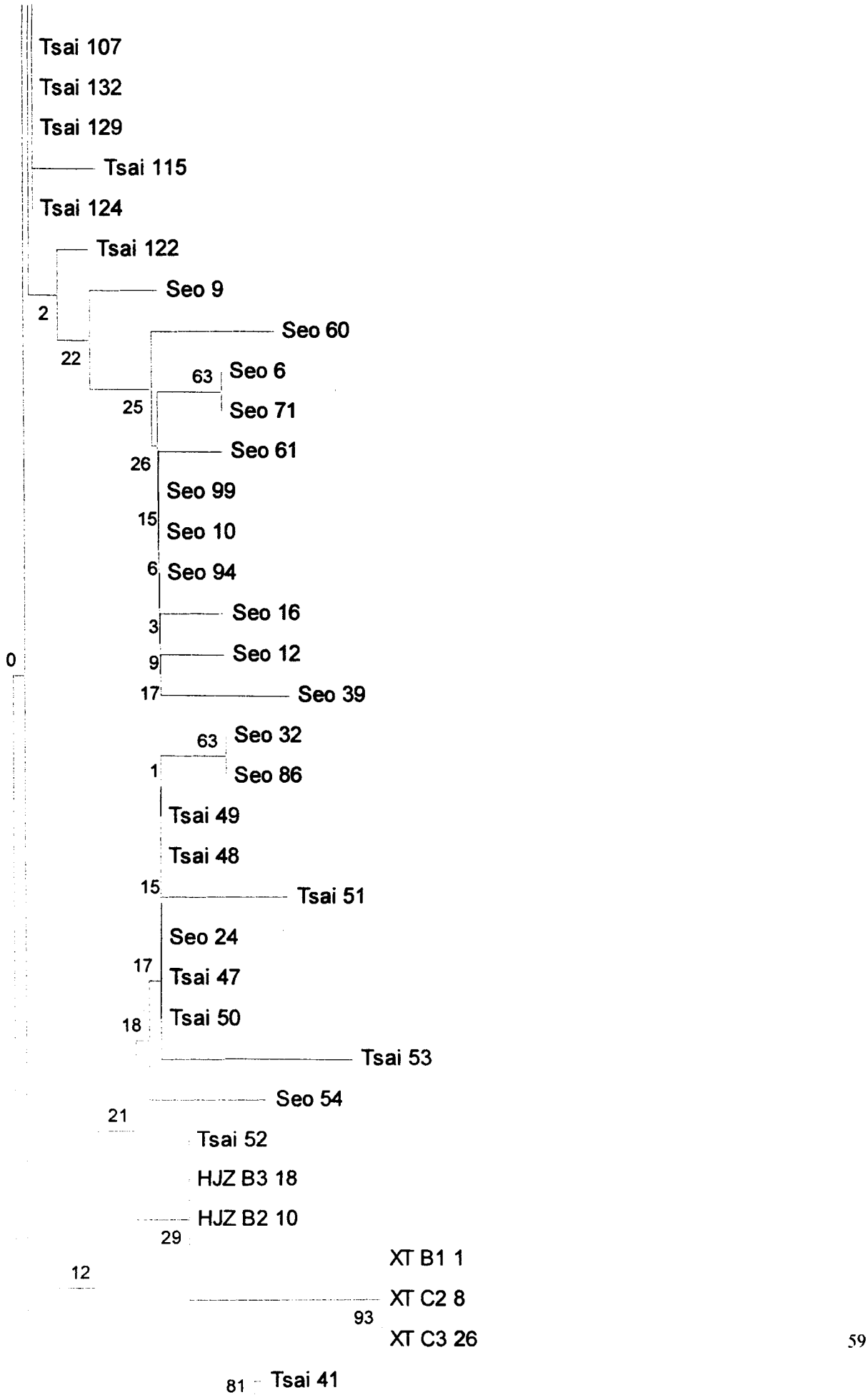
mtDNA databases can make it difficult to find and include HV2 data in phylogenetic analyses. Overall though, the Anyang mtDNA data when compared to other Asian population data, fits well and supports overall population homogeneity hypothesis (Oota et al. 2002).

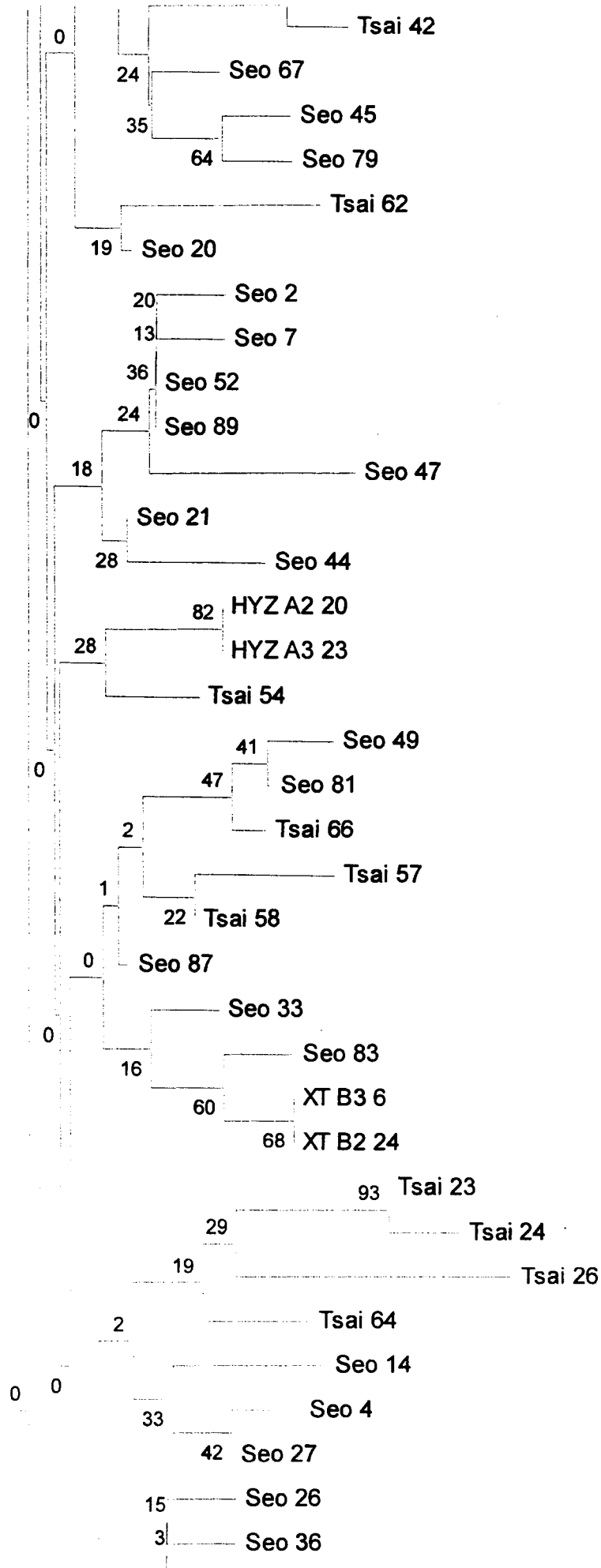
Conclusion

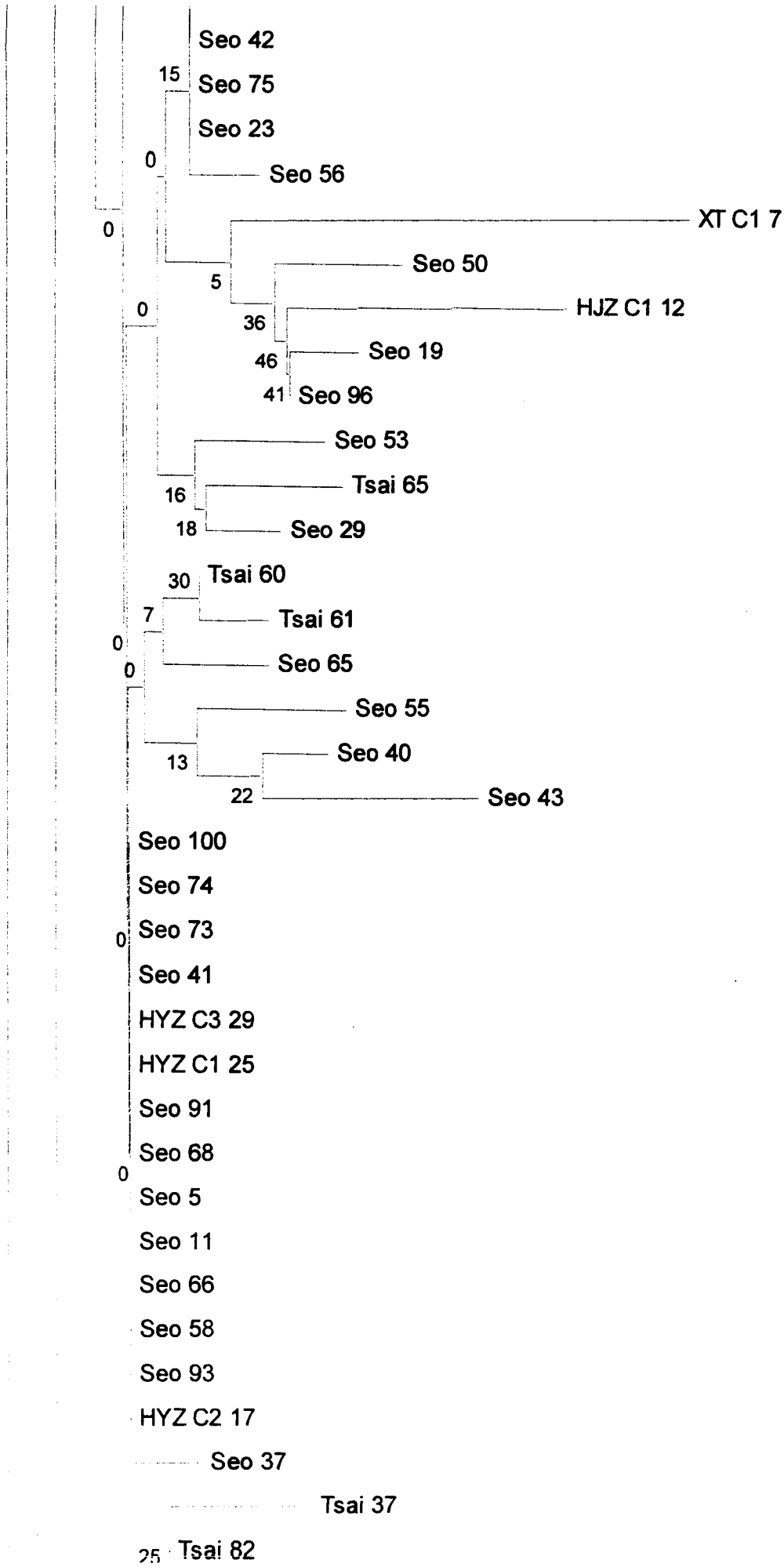
The longitudinal study concerning the identification of maternal lineages using mtDNA from hair samples was successful in determining related persons from one of three villages in remote central China. The 'blind study' research design was useful to validate the methodology, reduce researcher bias, and importantly, show the value of mtDNA for addressing questions about household make-up and residency type. I should admit that upon release of the actual individual identifications, which showed that seven of nine grandmothers did not match mothers and daughters in the same household, that I was concerned that something had gone wrong with the analyses. It wasn't until Chinese residency patterns were taken into consideration that the mtDNA longitudinal data made sense. This shows the importance of looking beyond the genetic data and putting it into a cultural, spatial or temporal context so that the true value of the genetic data can be understood. Finally, the phylogenetic analyses of mtDNA from Anyang showed that these people fit well within the larger Asian population, but that the inclusion of HV2 is necessary for deeper understanding of mtDNA in Asia.

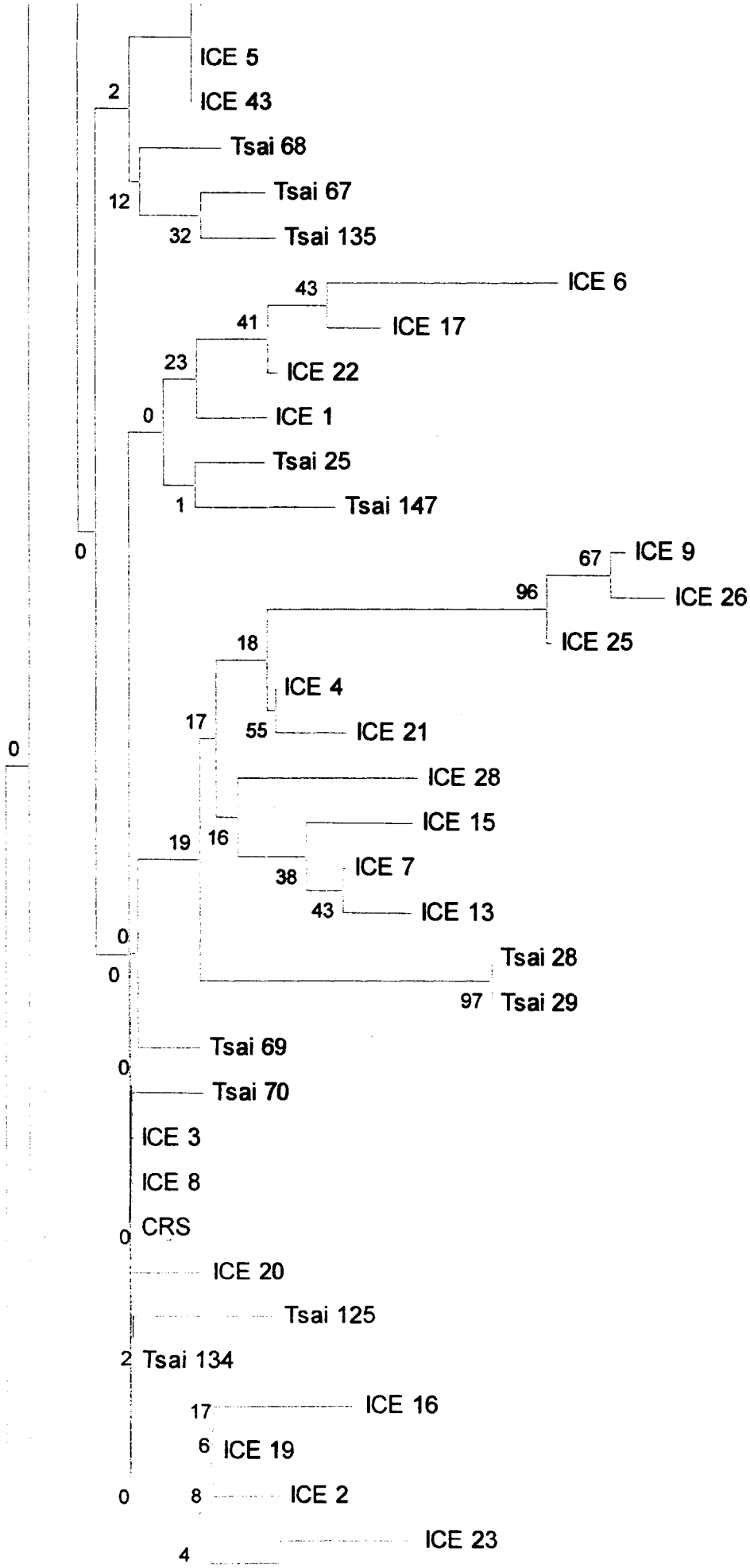


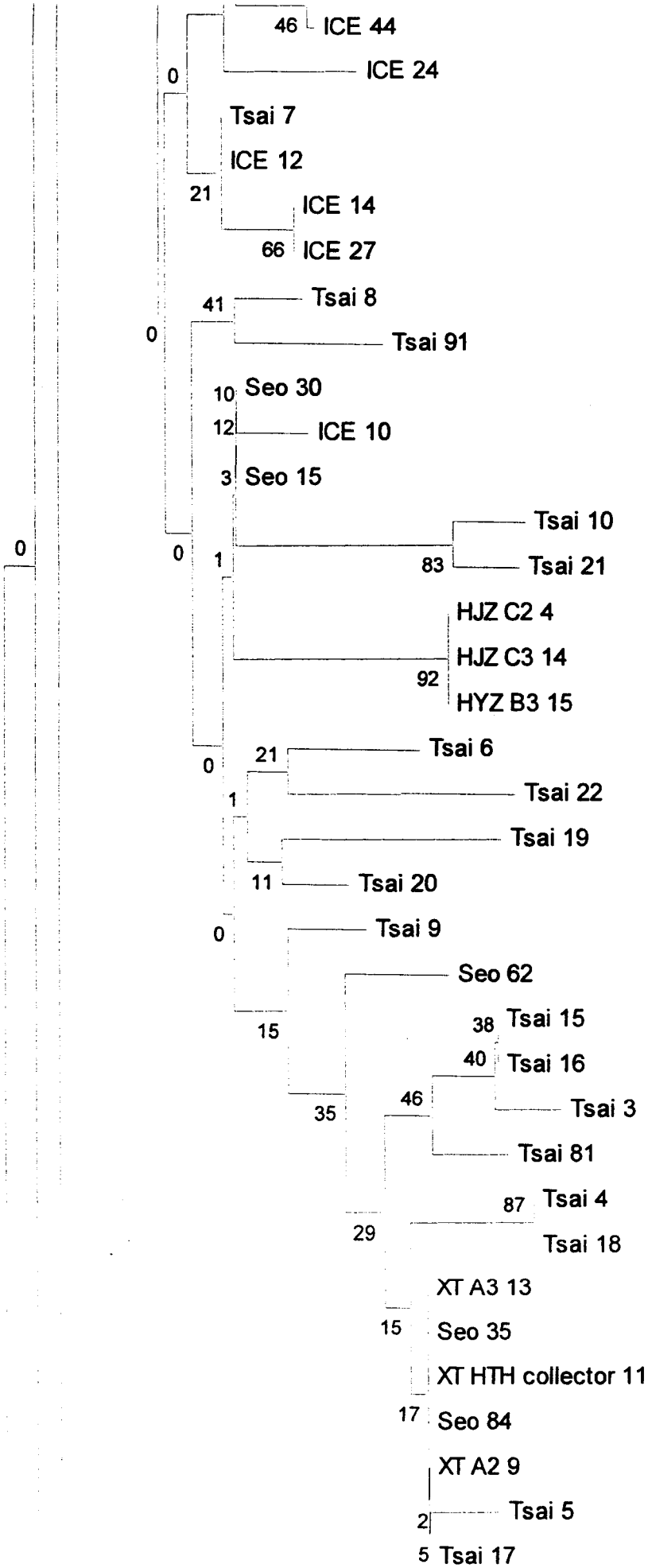


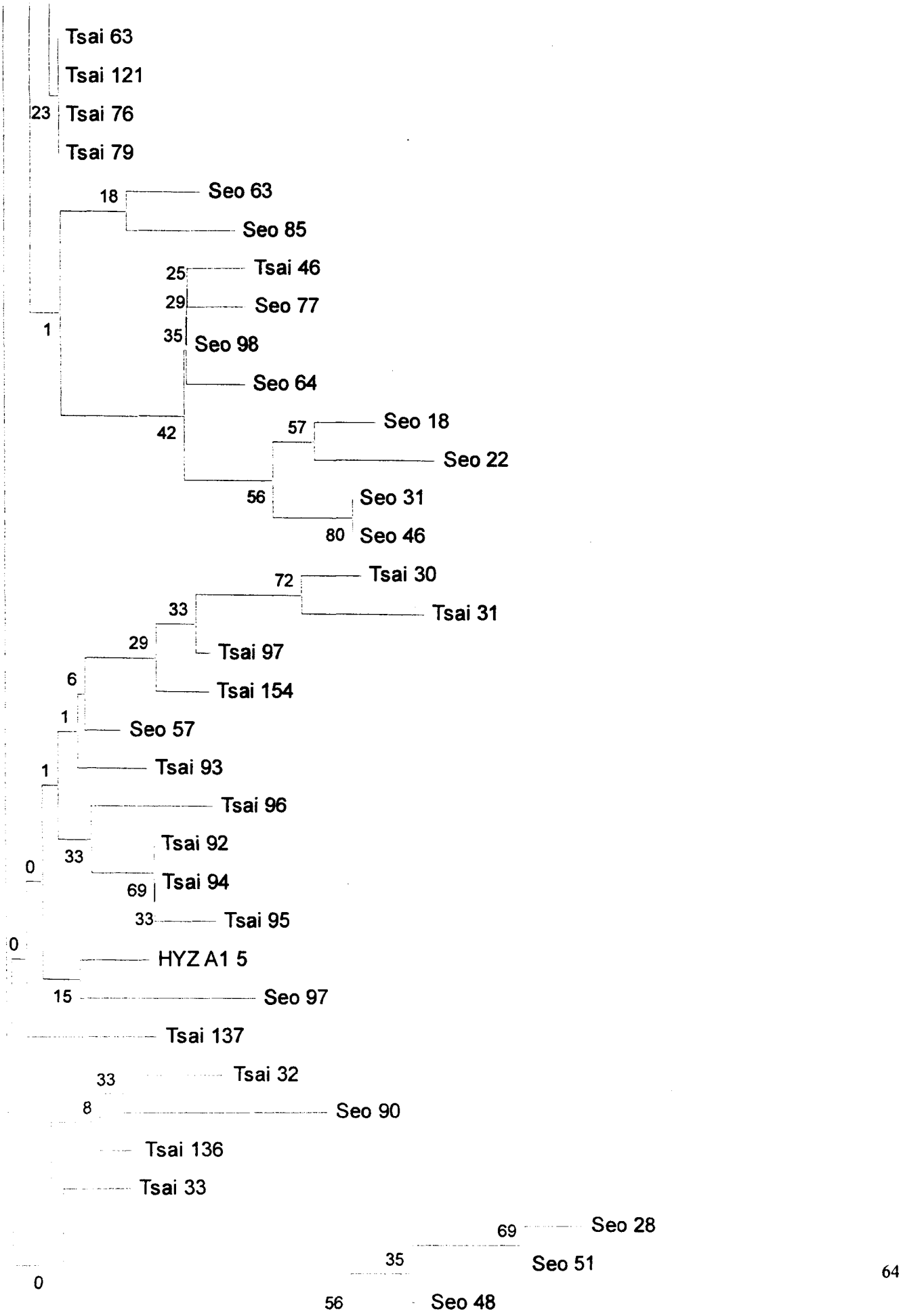


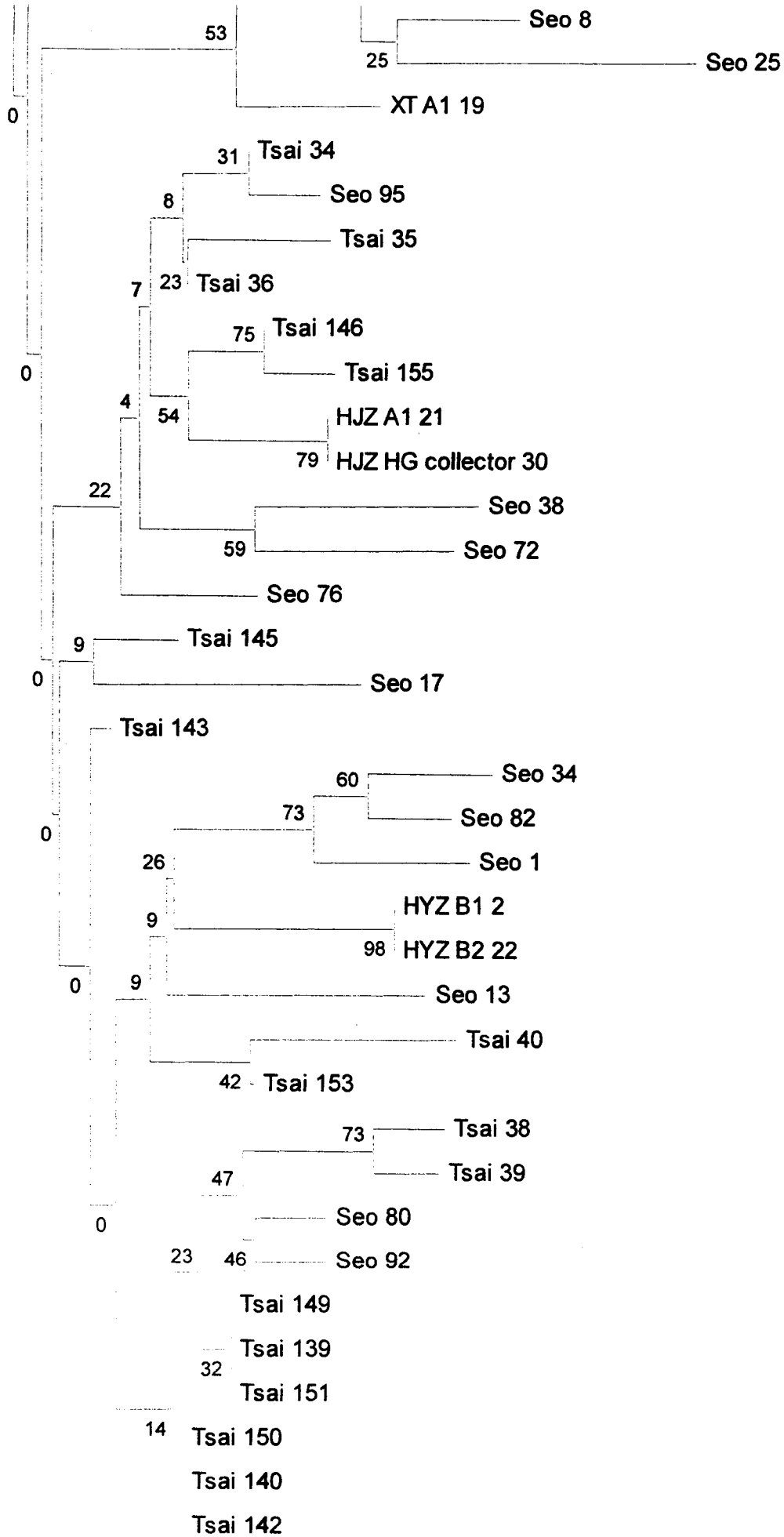


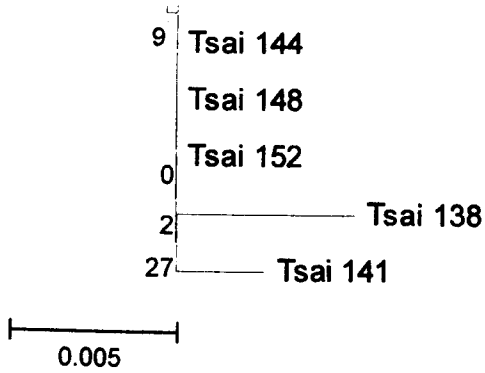


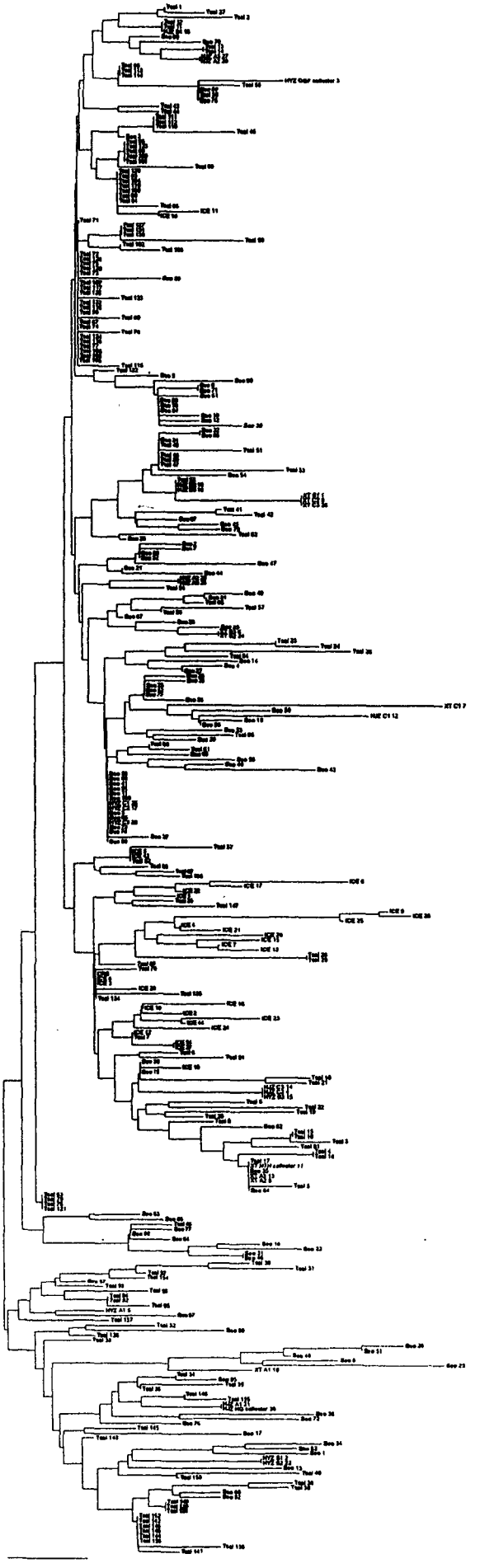












LITERATURE CITED

ALLEN, M., ENGSTROM, A-S., MEYERS, S., HANDT, O., SALDEEN, T., VON HAESELER, A., PÄÄBO, S., AND U. GYLLENSTEN. 1998. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. Journal of Forensic Science 43(3):453-464.

ALONSO, A., SALAS, A., ALBARRN, C., ARROYO, E., CASTRO, A., CRESPILLO, M., DI LONARDO, A.M., LAREU, M.V., CUBRIA, C.L., SOTO, M.L., LORENTE, J.A., SEMPER, M.M., PALACIO, A., PAREDES, M., PEREIRA, L., LEZAUN, A.P., BRITO, J.P., SALA, A., VIDE, M.C., WHITTLE, M., YUNIS, J.J., AND J. GOMEZ. 2002. Results of the 1999-2000 collaborative exercise and proficiency testing program on mitochondrial DNA of the GEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor. Forensic Science International 125:1-7.

ANDERSON, S., BANKIER, A.T., BARRELL, B.G., DE BRUIJN, M.H.L., COULSON, A.R., DROUIN, J., EPERON, I.C., NIERLICH, D.P., ROE, B.A., SANGER, F., SCHREIER, P.H., SMITH, A.J.H., STANDEN, R., AND I.G. YOUNG. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457-465.

ANDREWS, R.M., KUBACKA, I., CHINNERY, P.F., LIGHTOWLERS, R.N., TURNBULL, D.M., AND N. HOWELL. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics 23:147.

BAASNER, A., SCHAFER, C., JUNGE, A., AND B. MADEA. 1998. Polymorphic sites in the human mitochondrial DNA control region sequences: population data and maternal inheritance. Forensic Science International 98:169-178.

BAILLIET, G., ROTHHAMMER, F., CARNESE, F.R., BRAVI, C.M., AND N.O. BIANCHI. 1994. Founder mitochondrial haplotypes in Amerindian populations. American Journal of Human Genetics 54:27-33.

BAKER, L.E., MCCORMICK, W.F., AND K.J. MATTESON. 2001. A silica-based mitochondrial DNA extraction method applied to forensic hair shafts and teeth. Journal of Forensic Science 46(1): 126-130.

BENDALL, K.E., MACAULAY, V.A., AND B.C. SYKES. 1997. Variable levels of a heteroplasmic point mutation in individual hair roots. American Journal of Human Genetics 61:1303-1308.

BENDALL, K.E., AND B.C. SYKES. 1995. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. American Journal of Human Genetics 57:248-256.

BENDER, K., SCHNEIDER, P.M., AND C. RITTNER. 2000. Application of mtDNA sequence analysis in forensic casework for the identification of human remains. Forensic Science International 113:103-107.

BENSON, D.A., KARSCH-MIZRACHI, I., LIPMAN, D.J., OSTELL, J., RAPP, B.A., AND D.L. WHEELER. 2000. GenBank. Nucleic Acids Research 28(1):15-18.

BONNICHSEN, R., HODGES, L., REAM, W., FIELD, K.G., KIRNER, D.L., SELSOR, K., AND R.E. TAYLOR. 2001. Methods for the study of ancient hair: radiocarbon dates and gene sequences from individual hairs. Journal of Archaeological Science 28:775-785.

BOOM, R., SOL, C.J.A., SALIMANS, M.M.M., JANSEN, C.L., WERTHEIM-VAN DILLEN, P.M.E., AND J. VAN DER NOORDAA. 1990. Rapid and simple method for purification of nucleic acids. Journal of Clinical Microbiology 28(3):495-503.

BROMHAM, L., EYRE-WALKER, A., SMITH, N.H., AND J.M. SMITH. 2003. Mitochondrial Steve: paternal inheritance of mitochondria in humans. Trends in Ecology and Evolution 18(1):2-4.

BROWN, T.A., AND K.A. BROWN. 1992. Ancient DNA and the archaeologist. Antiquity 66:10-23.

BUDOWLE, B., ALLARD, M.W., AND M.R. WILSON. 2002. Critique of interpretation of high levels of heteroplasmy in the human mitochondrial DNA hypervariable region I from hair. Forensic Science International 126:30-33.

CALLOWAY, C.D., REYNOLDS, R.L., HERRIN JR., G.L., AND W.W. ANDERSON. 2000. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. American Journal of Human Genetics 66:1384-1397.

CANN, R.L., BROWN, W.M., AND A. WILSON. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106:479-499.

CANN, R.L., STONEKING, M., AND A.C. WILSON. 1987. Mitochondrial DNA and human evolution. Nature 325:31-36.

CANO, R.J., POINAR, H.N, PIENIAZEK, N.J., ACRA, A., AND G.O. POINAR. 1993. Amplification and sequencing of DNA from a 120-135-million-year-old weevil. Nature 363:536-538.

CAPELLI, C., TSCHENTSCHER, F., AND V.L. PASCALI. 2003. "Ancient" protocols for the crime scene? Similarities and differences between forensic genetics and ancient

DNA analysis. Forensic Science International 131:59-64.

CARRACEDO, A., BÄR, W., LINCOLN, P., MAYR, W., MORLING, N., OLAISEN, B., SCHNEIDER, P., BUDOWLE, B., BRINKMANN, B., GILL, P., HOLLAND, M., TULLY, G., AND M. WILSON. 2000. DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Science International 110:79-85.

CHANG, K.C. 1977. The continuing quest for China's origins. Archaeology 30(3):187-193.

COMAS, D., PÄÄBO, S., AND J. BERTRANPETIT. 1995. Heteroplasmy in the control region of human mitochondrial DNA. Genome Research 5(1): 89-90.

DENNIS, C. 2003. Error reports threaten to unravel databases of mitochondrial DNA. Nature 421:773-774.

FAGAN, B.M. (editor).1996. The Oxford Companion to Archaeology Oxford University Press, New York.

GILES, R.E., BLANC, H., CANN, H.M. AND D.J. WALLACE. 1980. Maternal inheritance of human mitochondrial DNA. Proceedings of the National Academy of Science USA 77:6715-6719.

GOCKE, C.D., BENKO, F.A., AND P.K. ROGAN. 1998. Transmission of mitochondrial DNA heteroplasmy in normal pedigrees. Human Genetics 102(2):182-186.

GOLENBERG, E.M., GIANNASI, D.E., CLEGG, M.T., SMILEY, C.J., DURBIN, M., HENDERSON, D., AND G. ZURAWSKI. 1990. Chloroplast DNA sequence from a Miocene Magnolia species. Nature 344:656-658.

GRAVER, A. 1999. Ancient DNA: A biomolecular investigation of human remains from the Shang Dynasty Hei Heru site at Anyang, China. Research paper on file, Department of Anthropology, University of Toronto.

GRAVER, A.M., MOLTO, J.E., RAPP JR. G., AND W. EN. 2000. Ancient DNA analysis of human remains from a late Shang Dynasty site at Anyang, China. 5th International Ancient DNA Conference (Manchester, England) poster presentation.

HANDT, O., HOSS, M., KRINGS, M., AND S. PÄÄBO. 1994. Ancient DNA: Methodological challenges. Experientia 50:524-528.

HASHIMOTO, K. 1988. The structure of human hair. Clinics in Dermatology 6(4): 7-21.

HAVILAND, W.A. 1996. Cultural Anthropology (Eighth Edition). Harcourt Brace & Company 277-278.

HELGASON, A., SIGURETH ARDOTTIR, S., GULCHER, J.R., WARD, R. AND K. STEFANSSON. 2000. MtDNA and the origin of the Icelanders: deciphering signals of recent population history American Journal of Human Genetics 66 (3):999-1016.

HELLMANN, A., ROHLEDER, U., SCHMITTER, H., AND M. WITTIG. 2001. STR typing of human telogen hairs – a new approach. International Journal of Legal Medicine 114:269-273.

HIGUCHI, R., VON BEROLDINGEN, C.H., SENSABAUGH, G.F., AND H.A. ERLICH. 1988. DNA typing from single hairs. Nature 332:543-546.

HIGUCHI, R., BOWMAN, B., FRIEBERGER, M., RYDER, O.A., AND A.C. WILSON. 1984. DNA sequences from the quagga, an extinct member of the horse family. Nature (London) 312:282-284.

HIGUCHI, R.G., WRISCHNIK, L.A., OAKES, E., GEORGE, M., TONG, B., A.C. WILSON. 1987. Mitochondrial DNA of the Extinct Quagga: Relatedness and extent of postmortem change. Journal of Molecular Evolution 25:283-287.

HORAI, S., AND K. HAYASAKA. 1990. Intraspecific nucleotide sequence differences in the major non-coding region of human mitochondrial DNA. American Journal of Human Genetics 46:828-842.

HORAI, S., KONDO, R., NAKAGAWA-HATTORI, Y., HAYASHI, S., SONODA, S., AND K. TAJIMA. 1993. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Molecular Biological Evolution 10(1):23-47.

IVANOV, P.L., WADHAMS, M.J., ROBY, R.K., HOLLAND, M.M., WEEDN, V.W., AND T.J. PARSONS. 1996. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nature Genetics 12:417-420.

KRINGS, M., STONE, A., SCHMITZ, R.W., KRAINITZKI, H. STONEKING, M., AND S. PÄÄBO. 1997. Neandertal DNA sequences and the origin of modern humans. Cell 90:19-30.

KUMAR, S., TAMURA, K., JAKOBSEN, I.B., AND N. MASATOSHI. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17(12):1244-1245.

LATOURETTE, K.S. 1967. The Chinese: Their History and Culture. The Macmillian Company, New York:26-33.

LEE, S.D.L., SHIN, C.H., KIM, K.B., LEE, Y.S., AND J.B. LEE. 1997. Sequence variation of mitochondrial DNA control region in Koreans. Forensic Science International 87:99-116.

LINCH, C.A., SMITH, S.L., AND J.A. PRAHLOW. 1998. Evaluation of the human hair root for DNA typing subsequent to microscopic comparison. Journal of Forensic Science 43(2):305-314.

LINCH, C.A., WHITING, D.A., AND M.M. HOLLAND. 2001. Human hair histogenesis for the mitochondrial DNA forensic scientist. Journal of Forensic Science 46(4):844-853.

LUTZ, S., WITTIG, H., WEISSER, H-J., HEIZMANN, J., JUNGE, A., DIMO-SIMONIN, N., PARSON, W., EDELMANN, J., ANSLINGER, K., JUNG, S., AND C. AUGUSTIN. 2000. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Science International 113:97-101.

MELTON, T., CLIFFORD, S., KAYSER, M., NASIDZE, I., BATZER, M., AND M. STONEKING. 2001. Diversity and heterogeneity in mitochondrial DNA of North American populations. Journal of Forensic Science 46(1):46-52.

MELTON, T., AND K. NELSON. 2001. Forensic mitochondrial DNA analysis: Two years of commercial casework experience in the United States. Croatian Medical Journal 42(3):298-303.

MERRIWETHER, D.A., ROTHHAMMER, F., AND R.E. FERRELL. 1994. Genetic variation in the New World: ancient teeth, bone, and tissue as sources of DNA. Experientia 50:592-601.

MERRIWETHER, D.A., ROTHHAMMER, F., AND R.E. FERRELL. 1995. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. American Journal of Physical Anthropology 98:411-430.

MORTON, W.S. 1980. China: Its History and Culture. McGraw-Hill Book Company, New York: 14-21.

MOURIER, T., HANSEN, A.J., WILLERSLEV, E., AND P. ARCTANDER. 2001. The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Letter to the Editor: Molecular and Biological Evolution 18(9):1833-1837.

MULLIS, K.B., and F. FALOONA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology 155:335-350.

OOTA, H., KITANO, T., JIN, F., YUASA, I., WANG, L., UEDA, S., SAITOU, N., AND M. STONEKING. 2002. Extreme mtDNA homogeneity in continental Asian populations. American Journal of Physical Anthropology 118:146-153.

OOTA, H., SETTHEETHAM-ISHIDA, W., TIWAWECH, D., ISHIDA, T., AND M. STONEKING. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nature Genetics 29:20-21.

OU, C-Y., MOORE, J.L., AND G. SCHOCHETMAN. 1991. Use of UV irradiation to reduce false positivity in polymerase chain reaction. Biotechniques 10(4):442-444.

PÄÄBO, S. 1984. Uber den Nachweiss von DNA in altagyptischen Mumien. Das Altertum 30:213-218.

PÄÄBO, S. 1985a. Preservation of DNA in ancient Egyptian mummies. Journal of Archaeological Science 12:411-417.

PÄÄBO, S. 1985b. Molecular cloning of ancient Egyptian mummy DNA. Nature (London) 314:644-645.

PÄÄBO, S. 1986. Molecular genetic investigations of ancient human remains. Cold Spring Harbor Symposia on Quantitative Biology, Volume L1:441-446.

PÄÄBO, S., IRWIN, D.M., AND A.C. WILSON. 1988. Mitochondrial DNA sequences from a 7,000-year-old brain. Nucleic Acid Research 16:9775-9787.

PARSONS, T.J., MUNIEC, D.S., SULLIVAN, K., WOODYATT, N., ALLISTON-GREINER, R., WILSON, M.R., BERRY, D.L., HOLLAND, K.A., WEEDN, V.W., GILL, P. AND M.M. HOLLAND. 1997. A high observed substitution rate in the human mitochondrial DNA control region. Nature Genetics 15: 363-367.

RICHARDS, M., SYKES, B.C., AND R.E.M. HEDGES. 1995. Authenticating DNA extracted from ancient skeletal remains. Journal of Archaeological Science 22:291-299.

RILEY, P.A. 1997. Melanin. International Journal of Biochemistry and Cell Biology 29(11):1235-1239.

RÖHL, A., BRINKMANN, B., FORSTER, L., AND P. FORSTER. 2001. An annotated mtDNA database. International Journal of Legal Medicine 155:29-39.

SACHS, H. History of hair analysis. 1997. Forensic Science International 84:7-16.

SARICH, V.M., AND A.C. WILSON. 1967. Immunological time scale for hominid evolution. Science 158:1200-1203.

SEO, Y., STRADMANN-BELLINGHAUSEN, B., RITTNER, C., TAKAHAMA, K., AND P.M. SCHNEIDER. 1998. Sequence polymorphism of mitochondrial DNA control region in Japanese. Forensic Science International 97:155-164.

SHEN, C. 2002. Anyang and Sanxingdui: unveiling the mysteries of ancient Chinese civilizations. Royal Ontario Museum, Kingsweb Inc., Mississauga, Ontario.

STARIKOVSKAYA, Y.B., SUKERNIK, R.I., SCHURR, T.G., KOGELNIK, A.M., AND D.C. WALLACE. 1998. MtDNA diversity in Chukchi and Siberian Eskimos: Implications for the genetic history of ancient Beringia and the peopling of the New World. American Journal of Human Genetics 63:1473-1491.

STEWART, J.E.B., FISHER, C.L., AAGAARD, P.J., WILSON, M.R., ISENBERG, A.R., POLANSKEY, D., POKORAK, E., DIZINNO, J.A., AND B. BUDOWLE. 2001. Length variation in HV2 of the human mitochondrial DNA control region. Journal of Forensic Science 46(4):862-870.

STONE, A.C., STARRS, J.E., AND M. STONEKING. 2001. Mitochondrial DNA analysis of the presumptive remains of Jesse James. Journal of Forensic Science 46(1):173-173.

STONEKING, M. 2000. Hypervariable sites in the mtDNA control region are mutational hotspots. American Journal of Human Genetics 67:1029-1032.

TAGLIABRACCI, A., TURCHI, C., BUSCEMI, L., AND C. SASSAROLI. 2001. Polymorphism of the mitochondrial DNA control region in Italians. International Journal of Legal Medicine 114:224-228.

TAMARIN, R.H. 1996. Principles of Genetics, 5th Edition. Wm. C. Publishers Dubuque Iowa. pp 492.

TOBIN, D.J., AND R. PAUS. 2001. Graying: gerontobiology of the hair follicle pigmentary unit. Experimental Gerontology 36:29-54.

TSAI, L.C., LIN, C.Y., LEE, J.C.I., CHANG, J.G., LINACRE, A., AND W. GOODWIN. 2001. Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. Forensic Science International 119:239-247.

TULLY, G., BÄR, W., BRINKMANN, B., CARRACEDO, A., GILL, P., MORLING, N., PARSON, W., AND P. SCHNEIDER. 2001. Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Science International 124:83-91.

TULLY, L.A., PARSONS, T.J., STEIGHNER, R.J., HOLLAND, M.M., MARINO, M.A., AND V.L. PRENGER. 2000. A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. American Journal of Human Genetics 67:432-443.

TZEN, C-Y., WU, T-Y., AND H-F. LIU. 2001. Sequence polymorphism in the coding region of mitochondrial genome encompassing position 8389-8865. Forensic Science International 120:204-209.

VIGILANT, L., PENNINGTON, R., HARPENDING, H. KOCHER, T.D., AND A.C. WILSON. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proceedings of the National Academy of Science USA 86:9350-9354.

VIGILANT, L., STONEKING, M., HARPENDING, H., HAWKES, K., AND A.C. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. Science 253:1503-1507.

VON HAESLER, A., SAJANTILA, A., AND S. PÄÄBO. 1995. The genetical archaeology of the human genome. Nature Genetics 14:135-140.

WALLACE, D.C. 1995. Mitochondrial DNA variation in human evolution, degenerative diseases, and aging. American Journal of Human Genetics 57:201-223.

WALLACE, D.C. 1999. Mitochondrial diseases in man and mouse. Science 283:1482-1488.

WILSON, M.R., HOLLAND, M.M., STONEKING, M., DIZINNO, J.A., AND B. BUDOWLE. 1993. Guidelines for the use of mitochondrial DNA sequencing in forensic science. Crime Lab Digest 20(4):68-77.

WILSON, M.R., POLANSKEY, D., BUTLER, J., DIZINNO, J.A., REPLOGLE, J., AND B. BUDOWLE. 1995. Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. Biotechniques 18:662-669.

WILSON, M.R., POLANSKEY, D., REPLOGLE, J., DIZINNO, J.A., AND B. BUDOWLE. 1997. A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes. Human Genetics 100:167-171.

WOLPOFF, M.H., HAWKS, J., FRAYER, D.W., AND K. HUNLEY. 2001. Modern human ancestry at the peripheries: a test of the replacement theory. Science 291:293-297.

YAO, Y-G., NIE, L., HARPENDING, H., FU, Y-X., YUAN, Z-G., AND Y-P. ZHANG. 2002. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence

diversity. American Journal of Physical Anthropology 118:63-76.

YAO, Y-G., WATKINS, W.S., AND Y-P. ZHANG. 2000. Evolutionary history of mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of east and southeast Asia. Human Genetics 107:504-512.