INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

GENETIC VARIATION OF MATERNAL AND PATERNAL LINEAGES WITHIN THE HAVASUPAI INDIANS OF NORTHERN ARIZONA

by

Keith Darren Coon

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MOLECULAR AND CELLULAR BIOLOGY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2002

UMI Number: 3073209

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Keith Darren Coon

entitled Genetic Variation Of Maternal And Paternal Lineages Within The Havasupai Indians Of Northern Arizona

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

David W. Mount — 11/5/2002
David W. Mount, Ph.D. — Date

Robert P. Erickson, MD — 11/5/2002
Robert P. Erickson, M.D. — Date

Murray Brilliant — 11/5/2002
Murray Brilliant, Ph.D. — Date

Michael F. Hammer — 11-5-02
Michael F. Hammer, Ph.D. — Date

_____ — _____
Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

David W Mount — 11/5/2002
Dissertation Director: David W. Mount — Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: _____

ACKNOWLEDGEMENTS

First off, I'd like to thank my committee Dr. Robert Erickson, Dr. Murray Brilliant, Dr. Michael Hammer, and Dr. David Mount for their support in the completion of this document and subsequent degree. Their willingness to serve on the dissertation committee of an unknown transfer student was gracious and generous and will not be forgotten. I'd especially like to thank Dr. Mount for his beneficent acceptance of the role of dissertation director and for standing up for me in the final hours of degree completion. I'd also like to express my gratitude to Dr. Hammer, who took me under his wing (and into his lab) upon my arrival at the University of Arizona. His technical assistance and mentoring was at the center of my success at this university. Thank you all.

I'd also like to thank all of the colleagues, at both Arizona State University and University of Arizona, that I have worked with over the many years since I embarked on this journey in 1996. Though there are too many to name, I would like to mention those most special... Josh Johnson, James Jancovich, Lewis Obermiller, Kirt Karl, Joe Austin, Greg Hocutt, Diane Hu, and Phil Young at ASU... Matt Kaplan, Christopher Tillquist, Tasha Altheide, Veronica Kearney, Alan Redd, and Tatiana Karafet at UofA. All of you made the journey, and the destination, a little bit sweeter.

My eternal gratitude goes to Dr. Edmund Stellwag at East Carolina University, a great teacher, scientist, and my very first mentor. I wouldn't be here if it weren't for you. You gave me your time and patience, and taught me the joys of research science. I truly wish I had listened to more of your advice... I would have saved myself a lot of trouble. But, trouble or not, I finally made it. Thank you so much for everything you did for me.

Last, but not least, I'd thank to thank those that were with me for the entire journey and every success and failure that I've seen over the years. To my wife Michele, thanks for sticking by me through the late nights, weekends, and holidays that I had to work while others were out having fun. I know you had to make a lot of sacrifices to see me to the end, I hope one day it will all be worth it. I'd also like to thank my parents, Bill and Jane, and my brother Kevin, without whom no success in my life would be possible. You've always been the most supportive family a person could ask for. I love you. I'd also like to express my gratitude to my other family, Walt and Faye Reynolds... I couldn't ask for a better set of in-laws. Thanks for the support.

All in all, thanks to everyone who helped me to get where I am at. I hope the future is good to all of us.

DEDICATION

I dedicate this document, and the accomplishment that it represents, to my parents, Bill and Jane Coon, without whose undying love and support over the many long years, none of this would be possible. Thank you and I love you always.

TABLE OF CONTENTS

TABLE OF CONTENTS – Continued

TABLE OF CONTENTS – Continued

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The Havasupai Indians are a small (~600 members), Yuman-speaking population that resides on a reservation in the Grand Canyon region of northern Arizona. Due to their location and cultural practices, they are subjected to extreme geographic and reproductive isolation. Additionally, an influenza epidemic at the turn of the century decreased the reproducing population to 43 females and 42 males. These observations suggest that the Havasupai should contain less genetic diversity than other Native American populations. They are also disproportionately affected by disease, having the third largest incidence of Non Insulin-Dependent Diabetes Mellitus (NIDDM) in the world.

An extensive analysis of maternal and paternal variation of the Havasupai was undertaken. Maternal variation was assayed by sequencing the non-coding control region (CR) of mitochondrial DNA (mtDNA), whereas paternal variation was examined using single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs) located on the non-recombining portion of the Y-chromosome (NRY). Due to the availability of familial pedigrees dating back to the mid-1800s and spanning eight generations, precise mutation rates were determined for maternal and paternal lineages. The Havasupai thus offer a unique opportunity to explore genetic variation in a small, homogenous Native American population for which extensive genealogical information is readily available.

Examination of mtDNA sequences from the complete 1127 bp CR of 43 Havasupai individuals along with SNP and STR data from the Y-chromosome of 48 male

Havasupai revealed that contrary to our initial prediction, the Havasupai, historically a small population (as evidenced by the limited number of founder haplotypes and low estimates of π), probably maintained a relatively high level of diversity (as evidenced by the number of rare haplotypes, high haplotype diversity, and high estimate of $E(\nu)$), probably a remnant of their association with the larger Pai population from which they are derived. As the level of diversity displayed by the Havasupai seems to have been maintained since the recent population bottleneck, it must have been too small and/or too short to have any detectable effect on the overall diversity of the tribe. Lastly, there appears to be some association between mtDNA mutations and NIDDM in the Havasupai population.

INTRODUCTION

The Havasupai

The Havasupai are a small Yuman-speaking tribe of Native Americans that currently reside in a reservation within the Grand Canyon of northern Arizona. They are part of a larger group of Native Americans called the Pai, or "people", that inhabited an expanse of land bounded by the lower Colorado river, the Grand Canyon, the Little Colorado River, and the Moencopi Wash on the west, north, east, and south, respectively (Martin, 1986). The Pai had a population of roughly 1200-1500 members, in the early 1860's, that consisted of numerous small, regional bands. Tribal custom usually dictated that unions occur between these local groups to prevent the impropriety of marrying a relative and, generally speaking, the wife would go off and live with the family of the husband (i.e. patrilocality).

Of the many bands of Pai, only two have names have endured; the Hualapai, or Pine Tree People, in the west, and the Havasupai, or People of the Turquoise Waters, to the east (Martin, 1986). Although they spoke a common language, the traditional enemy of both groups, the Yavapai, to the south, was not part of the main group of Pai. Contact with Europeans and subsequent epidemics slowly decimated the Pai population. War between Europeans and the Hualapai (1866-1869) lead to the establishment of Hualapai reservation in 1883. The Havasupai, however, avoided the Hualapai war and retreated to a small range of land centered on Havasu canyon. This land was declared Havasupai reservation in 1880 (Martin, 1986).

Why study the Havasupai?

The Havasupai offer a unique opportunity to examine genetic variation in a Native American population for several reasons. First, as mentioned previously, the already small Pai population from which the Havasupai derive, was diminished by European contact and epidemics resulting in an extremely small initial population (Martin, 1986). Second, the geographic isolation offered by the seclusion of the canyon, alongside culturally mediated reproductive isolation, has resulted in an extremely homogenous population (Markow et al., 1993). Third, the Havasupai display the third largest incidence of Non Insulin-Dependant Diabetes Mellitus (NIDDM) in the world (Zuerlein et al., 1991). Fourth, the Havasupai underwent a population bottleneck at the turn of the century, the evidence of which might appear under the scrutiny of genetic analysis. Lastly, and perhaps most importantly, pedigrees inclusive of eight generations and dating back to the middle 1800s are available for the Havasupai. Thus, the Havasupai provide a unique opportunity to examine genetic variation in a small, homogenous Native American population for which extensive genealogical information is readily available.

The genetics of small populations

In small, reproductively and geographically isolated populations, like the Havasupai, special circumstances exist that can produce rapid changes in gene frequencies that are totally independent of mutation, recombination, and natural selection.

These circumstances, collectively known as genetic drift, are due entirely to chance factors, and the smaller the population, the more susceptible it is to such random changes. The net effect of genetic drift on a small population's gene pool can be rapid evolution, which can have a significant effect on the gene/allele frequencies of subsequent generations.

Another important genetic condition affecting small populations is called founder effect. Founder effect occurs when a small group of individuals has many descendants that have a high survival rate and disproportionately populate future generations. The result for small populations is often high frequencies of specific genetic traits inherited from the few common ancestors who first displayed them and a subsequent loss of genetic diversity.

In addition to genetic drift and founder effect, consanguineous mating, or inbreeding, also has an incommensurate effect on small populations. It is an extreme form of positive assortative mating (i.e. preference for mating between similar individuals) since close relatives usually are genetically more similar than are unrelated people. Thus, when consanguineous mating occurs, the result is significantly less genetic diversity among the descendants than if the parents had mated with someone who was not closely related.

Thus, it can be seen that small populations like the Havasupai, which already display limited genetic diversity as a result of their size, are disproportionately affected by several genetic phenomena, including genetic drift, founder effect, and

consanguineous mating, all of which serve to potentially exacerbate the loss of genetic diversity within the population.

What is known about genetic diversity in the Havasupai?

The unique features of the Havasupai described above have prompted a series of investigations on the Havasupai. Markow and Martin (1993) looked at the level of developmental instability, defined as an increase in morphological variance that has been observed in inbred populations, in the Havasupai. In this inquiry, fluctuating asymmetry, the difference between right and left measurements of paired structures, as measured by dermatoglyphic traits like fingertip pattern and ridge counts, was used as an indicator of developmental stability. This study yielded an estimated average inbreeding coefficient of 1-2% and revealed that the level of fluctuating symmetry, by both measurements, was higher in the Havasupai than in a control group of Caucasians as well as a group of Havasupai of mixed descent (Markow and Martin, 1993). These results suggest that inbreeding has decreased the developmental stability of the Havasupai.

The only analysis of genetic variation in the Havasupai was undertaken in 1993 by Markow et al., who inquired about the level of HLA diversity within the group. They found 20 of a possible 32 HLA-A/B haplotypes, 9 of which showed positive linkage disequilibrium. In addition to the high level of linkage disequilibrium observed in the Havasupai, they also found an excess of heterozygosity. After experimentally eliminating biases that might be due to nonselective causes like non-consanguineous mating and different allele frequencies in males and females, the authors concluded that

balancing selection was responsible for maintaining the level of genetic variation observed at the HLA-A and HLA-B alleles (Markow et al., 1993).

Observations and objectives for present work

The existence of comprehensive pedigree information corresponding to the recent demographic fluctuations undergone by the Havasupai, lead to the initial observation that perhaps the effect(s) of the recent population bottleneck might be observable under the scrutiny of a genetic analysis of Havasupai maternal and paternal lineages. Inspired by the initial observations and the overall dearth of research on this unique and important population, an extensive examination of the genetic variation within the Havasupai was undertaken. Our first objective was to examine maternal lineages by performing a mitochondrial DNA-based population genetic analysis of the Havasupai that would reveal the level of maternal variation, as well as the average mitochondrial DNA mutation rate, for the tribe. Our second objective was to examine paternal lineages by performing a Y chromosome-based population genetic analysis of the Havasupai that would reveal the level of paternal variation, as well as the average Y-chromosome mutation rate, for the tribe. Our third objective was to attempt to gain some insight into the genetic consequences of the recent population bottleneck and the evolutionary processes involved in the loss, or maintenance, of genetic diversity in small, isolated populations.

Maternal genetic variation was ascertained through a DNA sequence analysis of the control region of mitochondrial DNA (mtDNA). Paternal genetic variation was assayed via an analysis of short tandem repeats and single nucleotide polymorphisms

contained within the non-recombining region of the Y-chromosome. The available pedigrees augmented these investigations by permitting the estimation of mutation rates for both of these uniparental genetic markers. Additionally, the pedigrees have clearly demonstrated a bias towards maternal transmission of NIDDM within the Havasupai (Fenger, 1992), suggesting a potential role for mtDNA in the etiology of NIDDM in this population. Thus, a fourth objective was to determine whether or not mitochondrial DNA mutations were associated with the incidence of NIDDM within the Havasupai population. To this end, an analysis of mtDNA mutations associated with the occurrence of NIDDM was undertaken. These analyses were endeavored upon in order to provide much needed insight into the genetics of this small population and its evolutionary relationships to other Native American populations.

What is mitochondrial DNA and why study it?

Mitochondrial DNA (mtDNA) is a circular, double-stranded, extrachromosomal DNA molecule that is 16,569 bp in size. It encodes 2 ribosomal RNAs (rRNAs), 22 transfer RNAs (tRNAs), and 13 polypeptides, which are crucial to the respiratory chain complexes that support oxidative phosphorylation and the generation of ATP. There are several factors that make mtDNA an ideal candidate for use in population genetic studies. First, it is exclusively maternally inherited and does not recombine allowing a precise measure of maternal relatedness (Olivo et al., 1983; Sutovsky et al., 2000). Second, its haploid mode of inheritance makes it more sensitive than nuclear DNA to demographic changes within a population, bottlenecks for example (Wilson et al., 1985; Moore, 1995).

Third, it has a high mutation rate, which results in a magnified view of diversity and evolutionary history (Brown et al., 1979; Wilson et al., 1985). Thus, mtDNA is a valuable tool for the investigation of population demography and genetic variation.

Maternal inheritance of mtDNA

Investigation into the potential role that mtDNA mutations might play in the development of NIDDM obviates the need for a basic understanding of the mode of mtDNA transmission. Although there is still a fair amount of mystery surrounding the inheritance of mtDNA from mother to offspring, several observations appear to garner the general support of the scientific community. First, mtDNA is strictly maternally inherited. Although limited paternal contribution of mtDNA has been argued by some based on the presence of the sperm midpiece mitochondria in human embryos after fertilization (Sathananthan et al., 1996; Ankel-Simons and Cummins, 1996), the dilution hypothesis (Smith and Alcivar, 1993; Ankel-Simons and Cummins, 1996), and a lack of sensitivity of conventional methods of detection (Gyllensten et al., 1991), these proposals have been largely overturned by more recent research.

The dilution hypothesis (Smith and Alcivar, 1993) merely states that the predominance of maternal mtDNA observed in offspring can be explained by the fact that sperm contain ~100 mtDNA (DeMartino et al., 1979; Hecht et al., 1984) and the oocyte contains ~100,000 mtDNA (Micheals et al., 1982; Gyllensten et al. 1985). In other words, paternal contribution of mtDNA is simply diluted beyond the point of detection using conventional analyses. This view was further strengthened by Gyllensten et al.

(1991) who, employing a more sensitive PCR-based detection method, showed that paternally inherited mtDNA molecules could be detected in mice from *inter*specific crosses at a frequency of 10^{-4} relative to the maternal contributions. This ratio is very close to that which would be expected by simple dilution.

As previously mentioned, however, these proposals have been largely overturned by more recent research. The dilution hypothesis has been disproved by new, more sensitive, PCR-based detection methods that can identify mtDNA in amounts as small as 0.01 fg, which corresponds to as few as two mtDNA molecules (Shitara et al., 1998). Additionally, it has been shown that although paternal mtDNA can be detected in experimentally-derived *inter*specific crosses (Gyllensten et al., 1991; Kaneda et al., 1995), it cannot be detected in naturally occurring *intra*specific offspring (Kaneda et al., 1995; Shitara et al., 1998). Indeed, Kaneda et al. (1995) showed that although paternal mtDNA from the sperm midpiece did exist in the cytoplasm of the oocyte, it was eliminated by the early pronucleus stage. Further studies revealed that even in *inter*specific crosses where paternal mtDNA was identified, it was not distributed to all tissues in the *F1* hybrids, was not transmitted to future generations through the female germ line, and was limited to the first generation of *inter*specific crosses (Shitara et al., 1998). It has also been suggested that the paternal mtDNA found in *inter*specific crosses only eludes elimination from the oocyte because the exclusion mechanism is species-specific (Kaneda et al., 1995; Shitara et al., 1998).

Several mechanisms for the elimination of sperm mitochondria have been proposed. One hypothesis suggests that sperm mitochondria are recognized by oocyte

cytoplasm because of the oxidative damage suffered while traversing the female genital tract (Allen, 1996). This hypothesis, however, does not explain the observation that paternal mtDNA is also eliminated from the oocyte following *in vitro* fertilization, which results in minimal levels of oxidative damage. Sutovsky et al. (1999) proposed a more convincing mechanism for the elimination of sperm mitochondria from the developing embryo. They suggest that sperm mitochondria are eliminated through ubiquitin-mediated proteolysis that occurs within the lysosome or proteosome of the fertilized oocyte (Sutovsky et al., 1999; Sutovsky et al., 2000). The ubiquitin tag is purportedly added to sperm mitochondrial proteins during spermatogenesis, concealed by disulfide bond cross-linking of mitochondrial membranes that occurs during transit through the male reproductive tract, and revealed again in the glutathione-mediated reducing environment of the oocyte cytoplasm (Sutovksy et al., 2000). Species-specificity of this mechanism is probably maintained by the incompatibility across species of ubiquitin-activating and –conjugating enzymes (Sutovsky et al., 2000). It should be noted that the mtDNA, itself, is not targeted for elimination, but rather, some proteinaceous component of the sperm mitochondrial membrane (Kaneda et al., 1995; Shitara et al., 1998; Sutovsky et al., 2000). A 30 kDa integral mitochondrial membrane protein, prohibitin, has been implicated as a potential target for this ubiquitination (Sutovsky et al., 2000), but more research needs to be done to confirm this hypothesis.

mtDNA recombination

Although it contains the enzymes necessary for homologous recombination (Thyagarajan et al., 1996; Tang et al., 2000), it is also commonly agreed that mtDNA does not undergo recombination. There have been several reports of mitochondrial recombination based largely on the presence of linkage disequilibrium (Eyre-Walker et al., 1999; Awadalla et al., 1999) and/or the detection of a rare point mutation that was shared among several mtDNA lineages in a single isolated locale (Hagelberg et al., 1999). However, these reports have been highly contested on the basis of inaccurate sequence data and inappropriate methodology for identification of linkage disequilibrium (Kivisild and Villems, 2000; Jorde and Bamshad, 2000; Kumar et al, 2000; Parsons and Irwin, 2000) and have eventually been largely dismissed, even by their initial proponents (Eyre-Walker and Awadalla, 2001).

Heteroplasmy and mitochondrial disease

As mentioned previously, mtDNA has a high mutation rate, estimated to be 10-100 times that of the nuclear genome. One of the causes of mutation induction in mtDNA is thought to be replication slippage (Howell and Smejkal, 2000). This slippage is probably the result of the low fidelity of mtDNA's gamma DNA polymerase (Kunkel and Loeb, 1981). Other factors that likely contribute to the high rate of mutation include the lack of efficient DNA repair mechanisms, the lack of protective DNA coating proteins such as histones, and a chemical environment that is rich in free radicals (Brown and Wallace, 1994). This high mutation rate is not readily expressed as a biological change due to the high number of mtDNA copies, as a typical somatic cell carries

hundreds of mitochondria, each of which contains anywhere from ten to several thousand copies of mtDNA (Shuster et al., 1988).

Multiple copies of mtDNA per cell permit the existence of a genetic phenomenon known as heteroplasmy, whereby multiple mtDNA genotypes can exist in the same individual, tissue, cell, or organelle. Most etiological mtDNA mutations are heteroplasmic, thus, mtDNA heteroplasmy is often associated with mitochondrial disease (DiMauro et al., 1998; Chinnery and Turnbull, 1999). The accumulation of 'deleterious' mutations in mtDNA is tolerated, often without obvious injurious effects, due to intracytoplasmic complementation among individual mitochondria (i.e. heteroplasmy), which enables defective mitochondria to function by scavenging enzymes and metabolites from wild-type mitochondria co-resident in the cell cytoplasm (Wallace, 1986). These circumstances result in mutational effects that will not be expressed until a certain mutation threshold is crossed. In other words, a certain amount of mutated mtDNA molecules must accumulate before a mutant phenotype can be expressed over the wild-type.

Heteroplasmy for genetically defective mtDNA has been demonstrated in individuals with several different mtDNA-linked diseases including Leber's hereditary optic neuropathy or LHON (Jacobi et al., 2001), chronic progressive external ophthalmoplegia or CPEO (Sudoyo et al., 1993), myoclonus epilepsy with ragged red fibers or MERRF (Szuhai et al., 2001), mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes or MELAS (Cardaioli et al., 2000), a multifactorial version of CPEO called Kearn's-Sayre syndrome or KSS (Byrne, 1991), and most recently

mitochondrial diabetes mellitus or MDM (Zeviani et al., 1991; Reardon et al., 1992; Morten et al., 1993; Moraes et al., 1993; Hao et al., 1995; Suzuki et al., 1996; Suzuki et al., 1997; Kameoka et al., 1998). In the case of LHON, KSS, and some instances of MDM, it has been shown that the severity of disease and age at disease onset are both influenced by the proportion of wild-type relative to mutant mtDNA molecules in the affected tissue of individuals expressing clinical symptoms of the disease (Holt et al., 1990; Moraes et al., 1989; Zeviani et al,. 1991; Reardon et al., 1992; Moraes et al., 1993).

Inheritance of mitochondrial mutation

It should be noted that not all mitochondrial disease-related mutations are heteroplasmic. Most polymorphisms within mtDNA are homoplasmic and are thus considered to have occurred long ago (Wallace, 1994). On the contrary, disease-related mutations, in general, are evolutionarily new heteroplasmic germline mutations (Gerbitz et al., 1996). Varying levels of heteroplasmy have been detected in the offspring of heteroplasmic females in both human (Chinnery et al., 2000) and animal (Hauswirth and Laipis, 1985; Koehler et al., 1991) pedigrees. However, heteroplasmy is known to be unstable (Hayashi et al., 1983) and it has been demonstrated that cells containing a mixture of wild-type and mutant mtDNAs are found to undergo a rapid shift of their genotype toward homoplasmy (Hauswirth and Laipis, 1982; Yoneda et al., 1992). In some cases, mtDNA drifts towards the pure mutant type, suggesting some replicative advantage held by mutated mtDNA (Yoneda et al., 1992). However, this result almost certainly depends on the type/location of the mutation, as some mutations are sure to be

disadvantageous. Thus, it remains unclear whether the maintenance of heteroplasmy within a cell is the result of simple random drift or some type of selection acting on the mutant mtDNAs (Yoneda et al., 1992; Hao et al., 1995).

Although very little is known about the mechanism of transmission of heteroplasmic mutations in humans, an mtDNA "bottleneck", whereby only a small number of mtDNA genomes is transmitted to the progeny with subsequent amplification of those few genomes repopulating the cells of the offspring, has been proposed to explain these rapid changes in genotype (Marchington et al., 1998). This bottleneck occurs between the primordial germ cell and primary oocyte stage (Poulton and Marchington, 2002). There are several factors implicated in the initiation of this bottleneck. First, the mitochondrial organelle copy number is amplified nearly 1000-fold during oogenesis, while mtDNA copy number is only amplified 100-fold, resulting in a dilution condition, where each organelle contains very few, perhaps only one, mtDNA molecule (Hauswirth and Laipis, 1985; Smith and Alcivar, 1993; Poulton et al., 1998). Second, these individual groups of mitochondria and their respective mtDNAs delineate the mitochondrial genotype of each fetal tissue due to the segregation of oocyte cytoplasm (i.e. cytoplasmic partitioning) in early embryogenesis (Chinnery et al., 2000; Poulton and Marchington, 2002). Lastly, embryonic partitioning into the primary germ layers (endo-, meso-, and ectoderm) results in potential, additional segregating events (Hauswirth and Laipis, 1985; Poulton et al., 1998; Poulton and Marchington, 2002).

Thus, dilution of mitochondria during oogenesis, unequal partitioning of mtDNA genomes during cytokinesis, and variable segregation of mutant mtDNA among various

tissues throughout embryonic development, leads to the observed differences between mutation load in daughter cells. It is extremely interesting to note that although the progeny of a single heteroplasmic female can have dramatically different levels of heteroplasmy, the mean level of heteroplasmy among all offspring is approximately equal to the proportion of mutated mtDNA in the mother (Jenuth et al., 1996). The unequal partitioning of mtDNA during cytokinesis is thought to be predominantly a result of random genetic drift (Jenuth et al., 1996; Chinnery et al., 2000; Brown et al., 2001), however, selection cannot be entirely ruled out in certain instances (Yoneda et al., 1992; 't Hart et al., 1996). Additionally, as no animal model yet exists for pathogenic mtDNA mutations, it is possible that pathogenic mutations are inherited differently than the neutral mutations that have been most rigorously investigated.

So, how did this mechanism for mitochondrial inheritance evolve? Over thirty years ago, it was proposed that asexual organisms with high mutation rates will accumulate mildly deleterious mutations over generations that will compromise cellular, and eventually organismal, function (Muller, 1964). Mitochondrial DNA fits Muller's criteria, and it is possible that the mtDNA bottleneck evolved to remedy this course. As previously mentioned, the mtDNA bottleneck leads to rapid genetic drift (Jenuth et al., 1996; Chinnery et al., 2000; Brown et al., 2001), which results in either the loss of the mutation, through drift, or to mutational homoplasmy among offspring (Bergstrom and Pritchard, 1998). Progeny homoplasmic for the mutation are soon exposed to natural selection, such that any severely detrimental phenotypes will be eliminated. Although this causes extreme problems at the individual level, the bottleneck results in the loss of

mildly deleterious mutations before they can accumulate within a population. In conclusion, it can be seen that, despite an abundance of research focusing on the inheritance of mitochondria and mtDNA, we still have only a rudimentary knowledge of the precise mechanisms involved in the transmission of mtDNA and mtDNA mutations.

What is the Y-chromosome and why study it?

The human Y-chromosome is the component of the nuclear DNA package that determines maleness. It is an unusual segment of the nuclear genome since, apart from two small regions in which pairing and exchange take place with the X chromosome, it is male-specific, haploid, and not subject to recombination (Jobling and Tyler-Smith, 1996). These unique properties of the non-recombining part of the Y-chromosome (NRY) have important consequences. Y-chromosomes pass down from father to son largely unchanged, except by the gradual accumulation of mutations, thereby retaining a record of the mutational events that have occurred along particular male lineages throughout evolution (Mitchell and Hammer, 1996). By examining the differences in the patterns of polymorphism between modern Y-chromosomes, a history of human paternal lineages can be reconstructed. In this way, the NRY is the male counterpart to mtDNA and complements maternal lineage investigations as well as studies using biparentally inherited markers in the rest of the genome. Thus, investigation of the Y-chromosome is carried out for essentially the same reasons as for mtDNA.

Population bottlenecks

When a population undergoes a sudden and dramatic decrease in size, a population bottleneck can occur. The forces giving rise to the population decrease can be anything including, but not limited to, natural catastrophe, epidemic disease, and war. Such a bottleneck, while likely eliminating alleles from the population, also has the effect of preserving a random sample of the original population's alleles in the surviving members. Thus, some alleles may increase in frequency, others may decrease, and some rare alleles may be lost altogether. The effects of a population bottleneck can vary depending on both the size to which the population is reduced and the duration of the bottleneck (i.e. the number of generations). The consequence of a bottleneck on any population, particularly one with a small initial size, is usually the loss of genetic variation. Loss of genetic variation is of singular concern in small populations, as it seriously threatens their evolutionary potential.

There are several ways to measure the loss of genetic variation within a population. Two measures of genetic variation that were examined in the Havasupai were haplotype diversity and nucleotide diversity. Haplotype or genetic diversity (h) is the probability that two randomly chosen haplotypes from the sample are different (Nei, 1987). Nucleotide diversity was estimated by two different statistics, π, the mean number of differences between all pairs of haplotypes in the sample (Nei, 1987), and [E(v)] or θ, an estimate of the level of nucleotide polymorphism (based on the number of segregating sites) representing the proportion of nucleotide sites that are expected to be polymorphic in any sample (Watterson, 1975). Using these estimates, among others, the loss, or maintenance, of genetic variation within the Havasupai, was ascertained.

Summary

In sum, an extensive investigation of Havasupai maternal and paternal lineages was undertaken in order to: 1) gain insight into the level of genetic variation displayed by the Havasupai; 2) determine whether or not evidence of a recent population bottleneck can be identified in the genetic history of this population; and 3) see if there is any association between mtDNA mutations and the occurrence of NIDDM within the Havasupai. This work represents a valuable addition to the field of population genetics and provides the only record of maternal and paternal genetic variation within this intriguing, but diminishing, population of Native Americans. In addition, it seeks to further understand the etiology of the diabetes epidemic that plagues this, and many other, Native American populations.

Chapter 1

MITOCHONDRIAL DNA VARIATION IN THE HAVASUPAI

INTRODUCTION

Migratory expansions into the New World appear to have been accompanied by significant reductions in genetic diversity (Wainscoat et al., 1986; Jones and Rouhani, 1986) due to one or more population bottlenecks. These bottlenecks purportedly resulted in reduced numbers of mitochondrial lineages being brought to the Americas. Four major Native American mtDNA founder haplotypes, or haplogroups, designated A, B, C, and D (Schurr et al., 1990; Horai et al., 1993; Torroni et al., 1993) vary in frequency among different New World populations (Lorenz and Smith, 1996). More recent reports propose that additional related (Bailliet et al., 1994) and/or unrelated lineages may also exist (Bailliet et al., 1994; Merriwether, 1994; Merriwether et al., 1995; Merriwether and Ferrell, 1996; Easton et al., 1996).

Superimposed upon the small number of founder haplogroups is the potential loss of diversity in small populations of Native Americans through genetic drift, founder effect (at the population level), and consanguineous mating. For most Native American groups, information on historic population sizes is unavailable, making predictions about relative genetic diversities for specific groups difficult. For one tribe, however, the Havasupai of northern Arizona (figure 1.1), evidence exists to suggest that this population should exhibit less mtDNA variation than other New World groups. First, this tribe, whose present population is approximately 600, suffered a population reduction in the early 1900's due to an influenza epidemic (Markow and Martin, 1993). The band, which numbered approximately 300 people prior to the epidemic, declined over the

course of eight to nine years from 1897 to 1905. Of the 166 surviving the disease, 43 females and 42 males reproduced. Complete genealogies, dating back to the middle-late 1800's and inclusive of eight generations, allowed an average inbreeding coefficient of over 1% for the present population to be calculated (Markow and Martin, 1993). A subsequent study of class I MHC variation revealed that the Havasupai have the lowest HLA variability of any human population examined (Markow et al., 1993).

No studies of mtDNA sequence variation specifically among the Havasupai have been undertaken. Mitochondrial haplogroup frequencies have, however, been surveyed among a number of North American populations (Lorenz and Smith, 1996), including Hokan speakers of the southwestern United States, the language group to which the Havasupai language belongs (Greenberg et al., 1986). Haplogroup B, followed by C, were the most frequent in the southwestern tribes surveyed (Lorenz and Smith, 1996). Haplogroup A was detected in low frequency in the Hokan speaking group that reportedly included several Havasupai samples.

It is predicted that due to the small size of the population and the effect of the recent bottleneck, as well as the geographic and reproductive isolation of this population, that the mtDNA diversity of the Havasupai will be significantly reduced compared to other Native American populations. It is also hoped that some insight into the immediate consequences of a population bottleneck and the evolutionary processes involved in the loss, or maintenance, of genetic diversity in small, isolated populations will be obtained. Finally, we would like to examine potential evolutionary relationships between the Havasupai and other Native American populations for which data are available. In order

to test these hypotheses, we performed DNA sequence analysis of the entire mtDNA control region (CR) for 43 Havasupai samples. The resulting data were compared to other Native American mtDNA data available for hypervariable segment I (HVS-I), hypervariable segment II (HVS-II), as well as a region designated HVS-I+II consisting of the two regions combined.

MATERIALS AND METHODS

Population samples

Following informed consent, blood samples were obtained on Havasupai reservation between 1993 and 1995. Samples were collected in 15ml Vacutainer tubes and transported by helicopter to Arizona State University where they were either used immediately to extract DNA or to create immortalized cell lines by Epstein-Barr virus transformation (Markow et al., 1993). Available genealogies (Markow and Martin, 1993) were utilized to eliminate any individuals for whom admixture was known to be present.

Isolation, amplification, and sequencing of DNA

Total genomic DNA was amplified by PCR (Saiki et al., 1988) using two sets of mtDNA-specific PCR primers that were designed to amplify the complete mtDNA CR in two separate segments. A 725 bp fragment of the CR containing HVS-I was amplified using primers mtL15975 (5' CTCCACCATTAGCACCCAAAG 3') and mtH130 (5' CAGATACTGCGACATAGG 3') and a 619 bp fragment of the CR containing HVS-II

was amplified using primers mtL21 (5' ATTAACCACTCACGGGAGCTC 3') and mtH639 (5' GGGTGATGTGAGCCCGTCTA 3'). PCR products were visualized by agarose gel electrophoresis.

Amplified PCR templates were purified with Microcon100 filtration columns (Amicon, Inc., Beverly, MA 01915, serial # 42413) according to the manufacturer's specifications and resuspended in 15 μL Tris low EDTA (10 mM Tris and 0.1 mM EDTA). The purified samples were analyzed by agarose gel electrophoresis and compared to a known standard to determine the optimum concentration (approximately 0.1 μg of DNA template) for DNA sequence analysis.

Each strand of the purified PCR fragment was sequenced, with the same primers used for PCR, using the PRISM Ready Reaction Dideoxy Terminator Cycle Sequencing Kit (Applied Biosystems Inc.) according to manufacturer's specifications. The sequencing reaction products were analyzed on an automated DNA sequencer (Model 377, Applied Biosystems Inc.). DNA sequences were checked for accuracy by sequencing both strands in both directions and comparing the sequence of each strand to each other in the *Sequence Navigator* program (Applied Biosystems Inc.). Ambiguities were resolved by visual observation of peak heights on each strand. Once resolved, individual sequences were aligned using the published Cambridge sequence (Anderson et al., 1981) and compiled according to sequence similarity to each other by the *GCG pileup* program (SeqLab, 2000).

Diversity, phylogenetic, and statistical analyses

The number of segregating sites, haplotype diversity, *h* (Nei, 1987 equation 8.4), two estimates of nucleotide diversity, π (Nei, 1987 equation 10.5), and E(ν) (Watterson, 1975 equation 1.4a), Tajima's D (Tajima, 1989), and raggedness statistics for the mismatch analysis were all calculated using DnaSP v. 3.51 (Rozas and Rozas, 1999). AMOVA and interpopulation variance analyses were performed using Arlequin v. 2.0 (Schneider et al., 2000) using 1000 permutations to test significance. Fifty-five Native American HVS-I+II haplotypes (from six populations) were used to construct neighbor-joining trees (5000 bootstrap replicates), with an African sequence (accession: AF347015) as an outgroup, in PAUP v. 4.0 (Swofford, 2000), employing the Tamura-Nei model with and without and a gamma correction (α = .17 for HVS-I+II (Yang, 1996)).

Data analyzed

In our analysis of the Havasupai, the entire 1122 bp control region (bp 16024-576) was examined for all 43 individuals in the study. However, the presence of five insertions, two within HVS-I (at bp 16193.1 and 16193.2) and three within HVS-II (at bp 309.1, 309.2, and 315.1), resulted in a total of 1127 sites, rather than the 1122 that were expected. To facilitate comparison with other studies, separate analyses were performed on a region of HVS-I (bp 16050-16383), HVS-II (bp 70-370), and a region, designated HVS-I+II, combining the two (bp 16050-16383 + bp 70-370), that was common in other studies. All sites that created an alignment gap (i.e. insertions and/or deletions designated 'indels') in any of the groups being compared were excluded from the analysis to ensure conformity of examined loci. Hence, a total of 327 bp were analyzed for HVS-I, a total

of 292 bp were analyzed for HVS-II, and a total of 619 bp were analyzed for HVS-I+II. All nucleotide positions are according to Anderson et al. (1981). These same bases were used in all subsequent analyses except for estimation of haplotype frequencies (table 1.3).

In our analyses across cultural groups, haplogroup frequencies were compiled manually, from data extracted from the respective publications, for all Native American populations for which mtDNA CR sequence was available (table 1.3). In order to promote consistent comparisons, only samples whose haplogroup profile was determined using mtDNA CR sequencing, as opposed to RFLP data, were considered in our comparison. On the contrary, all diversity data (table 1.4) was completely re-analyzed from raw sequences obtained from either the MOUSE (Bella Coola, Cayapa, Haida, Huetar, Kuna, Ngöbé, Nuu-Chah-Nulth) or GenBank (Emberá and Wounan) sequence databases. Five populations that have been surveyed for at least part of the CR, the Gavião, Mapuche, Xavante, Yanomami, and the Zoró were omitted from the primary analyses due to missing data points. All available sequences from each respective study were used in the analysis except sequence 35 from the Haida population, which was discarded due to suspected non-Native American origin (Ward et al., 1993). AMOVA analysis, *F*st scores, and mismatch distributions were only generated for the HVS-I region since it is the data set that contains the most populations. Phylogenetic analysis was only performed on the HVS-I+II region, as that is how the Havasupai haplotypes are defined. Thus, populations from the Pacific Northwest (Bella Coola, Haida, and Nuu-Chah-Nulth), that lack HVS-II sequence data, were excluded from this analysis.

RESULTS

Analysis of mtDNA sequence

Automated DNA sequencing of the entire 1127 bp of the mtDNA CR in 43 Havasupai Indians revealed 10 mtDNA haplotypes representing two of the four founding haplogroups initially described in 1990 by Schurr et al. (figure 1.2). Of the 43 individuals in the study, 14 (32.6%) had the most common haplotype (Hav 1), 8 (18.6%) had Hav 2, 10 (23.3%) had Hav 3, while the remaining haplotypes were present in frequencies of less than 10%. Nineteen percent of the sample displayed rare haplotypes (i.e. those that consist of only 1 or 2 individuals).

Forty-two polymorphic sites were found within the region: 11 length polymorphisms (6 deletions and 5 insertions) and 31 substitutions (30 transitions and 1 transversion). The deletion of a dinucleotide CA repeat at bp 522-523 was treated as a single mutational event. The nucleotide composition of the region was A: 29.8%, C: 33.5%, G: 13.7%, and T: 23.0%. All Havasupai individuals carried an A->G transition at two nucleotide positions, 73 and 263. The A->G transition at bp 263, present in all Native American populations, could represent an Amerind-specific substitution (Kolman et al., 1995). The bp 73 polymorphism is common among Native American populations as well. In addition, there were two unique substitutions, an A->T transversion at bp 16146 and a T->C transition at bp 485, (MITOMAP, http://www.gen.emory.edu/mitomap.html, 2002). Comparison of the Havasupai CR

haplotypes (Hav 1-10 from figure 1.2) to those of other Native American populations revealed that all 10 detected CR haplotypes are unique to the Havasupai.

Haplogroup frequencies

Although there is evidence that other founding lineages may exist (Bailliet et al., 1994; Merriwether, 1994; Merriwether et al., 1995; Merriwether and Ferrell, 1996; Easton et al., 1996), the majority of Native American mtDNAs can be placed within four major founding haplogroups A, B, C, and D (Bailliet et al., 1994). The polymorphisms that characterize these haplogroups are summarized in table 1.2. The frequencies of the four founding haplogroups were determined for all Native American populations for which mtDNA CR sequencing data were available (table 1.3). All sequences that did not display a pattern of polymorphism characteristic of one of the four major haplogroups were grouped into the "other" category. The Havasupai display only two of the four founding haplogroups, with haplogroup C far more common (n = 39 or 90.7%) than haplogroup B (n = 4 or 9.3%). There is wide variation in haplogroup frequency among the Central/South American groups and, as has been previously reported, only a very small percentage of Native Americans fall into the "other" category.

Estimations and comparisons of diversity

There were a total of ten mtDNA haplotypes for the Havasupai CR (1127 bp). When CR segments were examined individually, we found nine haplotypes for the 327 bp HVS-I (bp 16050-16383) and six haplotypes for the 292 bp HVS-II (bp 70-370)

regions (table 1.1). There were 29 substitutions in the CR, whereas HVS-I contained 15 substitutions and HVS-II only six. There were 7 substitutions in the non-hypervariable regions of the CR. One substitution present in the complete CR was lost from HVS-I due to the elimination of gap sites. Tajima's D was found to be negative for HVS-I and for the entire CR, while the D value for HVS-II is slightly positive. None of these values was statistically significant.

In addition to the Havasupai data on the complete CR, haplotype and nucleotide diversities for the HVS-I region were determined for the Havasupai as well as eight Amerind (Bella Coola, Cayapa, Emberá, Huetar, Kuna, Ngöbé, Nuu-Chah-Nulth, and Wounan) and one Na-Dene (Haida) group. Table 1.4 shows haplotype and nucleotide diversity values derived from a 327 bp region of HVS-I, a 292 bp region of HVS-II, and a 619 bp region combining the two (HVS-I+II) for which sequence data were available (table 1.4).

Comparisons of HVS-I haplotype and nucleotide diversities across cultural groups reveal that the Havasupai haplotype diversity lies in the middle of the range seen for New World populations of similar size (e.g. Bella Coola and Huetar) as well as when all ten populations are considered. Havasupai haplotype diversity values are most similar to the Cayapa and the Bella Coola. Nucleotide diversity, π, on the other hand, is, along with that of the Haida, among the lowest observed for any Native American population. Long-term nucleotide diversity, E(ν), like haplotype diversity, lies in the middle of the range of diversity when comparing to populations of similar size and in overall comparisons and is most similar to the Haida and the Huetar. For the HVS-II region,

comparisons across cultural groups reveal that only the Ngöbé have lower haplotype diversity. Both measures of HVS-II nucleotide diversity in the Havasupai were the lowest among the Native American groups for which data were available. The HVS-I+II region is similar to that of HVS-I with haplotype diversity in the middle, π among the lowest observed, and a relatively high E(ν).

Mismatch distribution analysis

In order to address the question of whether the Havasupai have a pattern of population expansion/reduction consistent with those of other Native American populations, a pairwise genetic difference, or mismatch distribution, analysis was performed for each population (Harpending et al., 1993). A population's mismatch distribution is proposed to contain an evolutionary "signature" of that population within histograms reflecting the number of pairwise nucleotide differences between individuals within that population (Slatkin and Hudson, 1991; Rogers and Harpending, 1992; Harpending et al., 1993). A unimodal distribution is characteristic of a population growing exponentially, whereas a ragged distribution is indicative of populations that have undergone bottlenecks in the past. A raggedness statistic, *r*, was designed to quantify the smoothness of observed mismatch distributions and is an ad hoc measure of high frequency variation in a number series (Harpending, 1994). Thus, raggedness is low from expanded populations and high in stationary populations (Harpending, 1994). Figure 1.3 shows HVS-I region mismatch distributions and their associated raggedness statistics for ten Native American populations. It can be seen that the Nuu-Chah-Nulth

have the lowest raggedness followed (in ascending order) by the Bella Coola, Emberá, Haida, Havasupai, and Wounan. The highest raggedness score was seen in the Huetar followed (in descending order) by the Kuna, Ngöbé, and Cayapa.

AMOVA and interpopulation variance

Several hierarchies of population substructure tested demonstrated significant levels of interpopulation variance in the AMOVA analysis of the HVS-I region (table 1.5). Five separate clusters of population substructure were tested: one cluster with one group containing all ten populations, one cluster with three groups separated according to geographic location (Pacific Northwest, Southwest US, and Central/South America), one cluster with six groups separated according to spoken language (Wakashan, Salishan, Na-Dene, Hokan, Chibcha, and Chocó), one cluster separated according to language phylum (Na-Dene and Amerind), and one cluster with the Havasupai grouped separately from all other populations. Approximately 75% of variation was found within populations in the cluster consisting of one group containing all populations ($p < 0.0001$). When the populations were grouped according to geographic criteria, ~16% of the variation was found among groups ($p = 0.03$). Significant results were also obtained using looser linguistic groupings with ~17% of variation found among groups ($p = 0.04$). When populations were grouped according to language phylum (Na-Dene for the Haida and Amerind for all other populations) no significant interpopulation variation was detected, as variation among groups in this cluster was 4.7% ($p = 0.10$). Additionally, no statistically significant interpopulation variation was seen when the Havasupai were

grouped separately from the other Native American populations ($p = 0.09$), where 31.9% variation was found among groups.

Genetic divergence between populations can be estimated by pairwise interpopulation variances. Significant levels of divergence were revealed in all possible pairwise groupings (table 1.6). All population comparisons except the Bella Coola/Ngöbé ($p = 0.045$), Bella Coola/Nuu-Chah-Nulth ($p = 0.009$), Cayapa/Nuu-Chah-Nulth ($p = 0.009$), Emberá/Wounan ($p = 0.027$), Huetar/Kuna ($p = 0.009$), Huetar/Ngöbé ($p = 0.018$), and Ngöbé/Nuu-Chah-Nulth ($p = 0.009$) comparisons were significant to the $p < 0.0001$ level. Interpopulation variances ranged from 6% between the Emberá and Wounan to 70% between the Haida and Havasupai. It is interesting to note that largest variances found between any population pairs were generally found in Havasupai comparisons. The Havasupai show an average 50% variance from all other populations. They are most closely related to the Wounan (23%) and most distantly related to the Haida (70%).

Phylogenetic analysis

Figure 1.4 represents the bootstrap consensus neighbor-joining tree found by PAUP v. 4.0 (Swofford, 2000) using the Tamura-Nei model. Bootstrap values from 5000 bootstrap replicates are indicated at the nodes of the tree. Fifty-five Native American HVS-I+II haplotypes as well as an African (Yoruban) outgroup sequence were included. This tree implies that haplogroup D, represented by Huetar haplotype ht09, and the "other" haplogroup, represented by the Cayapa haplotype cy10, appear to be ancestral to

the other haplogroups A, B, and C. It further suggests that haplogroup A is ancestral to B and C and that haplogroup C is the most recently diverged of the major founding haplogroups. All haplotypes appear to assemble correctly into haplogroups (i.e. B haplotypes cluster together, C haplotypes cluster together, etc.) and haplogroups A and C are supported by relatively high bootstrap values (63 and 82 respectively). Haplogroup B appears to be the most diverse haplogroup and demonstrates a low bootstrap value of 39. All Havasupai haplotypes aggregate correctly within their respective groups (hv01-07 cluster in haplogroup C and hv08-10 cluster in haplogroup B). Additionally, many of the Havasupai haplotypes group with each other, where haplotypes differ by only a few substitutions (see the hv04/ hv05/ hv06 cluster and the hv03/hv07 cluster within haplogroup C), demonstrating high bootstrap values. The Havasupai haplogroup B haplotypes don't appear to cluster as tightly.

DISCUSSION

Analysis of mtDNA sequence

Analysis of mtDNA sequences from the complete 1127 bp CR of 43 Havasupai revealed the presence of ten haplotypes representing only two (B and C) of the four (A, B, C, and D) major Native American founder haplogroups (figure 1.2). The existence of only the B and C haplogroups has been observed in restriction site data among Quechan/Cocopa tribes, local to Yuma, Arizona and Baja, California respectively (Lorenz and Smith, 1996). None of the ten Havasupai CR haplotypes (Hav 1 – Hav 10)

have been detected in any of the other Native American populations surveyed. It is possible, however, that these lineages do exist at low frequencies in other populations that have not been adequately sampled or are potentially frequent in other Native American populations of the southwest United States that have not been characterized using CR sequencing and, hence, not included in this study.

Rare haplotypes

The majority of Havasupai (~74%) belong to either haplotype Hav 1 (32.6%), Hav 2 (18.6%), or Hav 3 (23.3%), however, 19% of Havasupai belong to rare haplotypes. The number of rare haplotypes suggests that the level of diversity in the Havasupai prior to the bottleneck might have been substantial despite their small population size, probably a remnant of a greater genetic landscape associated with the larger Pai population from which the Havasupai are derived (Martin, 1986). Although the one mutation that was found in the Havasupai mtDNA pedigrees (see figure 4.8 in chapter 4) does potentially create a rare haplotype (Hav 5), there is no way to determine where/when in the pedigree the mutation occurred (i.e. whether or not the mutation occurred before or after the bottleneck). Regardless, either all, or all but one, of the rare haplotypes observed in the Havasupai were present before the population bottleneck.

The presence (i.e. maintenance) of rare haplotypes after a population bottleneck is unusual, as bottlenecks are known to reduce sequence diversity. Stone and Stoneking (1998) have reported that the average percentage of rare haplotypes in Native American populations is 45.7%. Surprisingly, 60% of Havasupai haplotypes are rare. In

comparison, two other populations purported to have undergone recent population bottlenecks, the Kuna and the Ngöbé of Panama (Batista et al., 1995), contain 43% and 60% rare haplotypes, respectively. However, it should be noted that the Kuna population is over 100 times larger, and the Ngöbé population is over 200 times larger than the Havasupai and, thus, both were expected to display significantly higher levels of diversity than the Havasupai.

As mentioned previously, the loss of diversity seen after a population bottleneck is dependent on the size and duration of that bottleneck. Although the size and duration, and indeed the existence, of the Kuna and Ngöbé population bottlenecks are unknown, this is not the case for the Havasupai. The Havasupai bottleneck lasted approximately ten years and reduced the population by less than half (from ~300 to 166 individuals). Perhaps a bottleneck of this size and duration is not sufficient to cause a substantial reduction in genetic diversity. If so, the number of rare haplotypes contained within the population before the bottleneck, would be largely maintained during and after the bottleneck, and this appears to be the case with the Havasupai.

Haplogroup frequencies

Our analysis of haplogroup frequencies (table 1.3) revealed that in the Havasupai, neither haplogroup A nor D were present, and that haplogroup C (90.7%) was far more prevalent that haplogroup B (9.3%). This contrasts earlier studies in which haplogroup B was found to be predominant over haplogroup C in a mixed sample of 18 Hokan speakers (the linguistic group to which the Havasupai belong), including several Havasupai of

unverified descent (Lorenz and Smith, 1996). This same investigation also detected a small percentage (11%) of haplogroup A in their sample (Lorenz and Smith, 1996). Given the absence of haplogroup A in the present study on a much greater number of known Havasupai, it is likely that a few individuals from one or more of the other tribes in the mixed sample accounted for the haplogroup A lineages detected in the earlier survey. This discrepancy also emphasizes the necessity for comprehensive sampling when attempting to characterize human populations.

The finding that only two of the four major founding lineages are represented in the Havasupai is consistent with the small initial population size and isolated geography of the Havasupai. The existence of a limited number of founder haplotypes has been previously observed in another population that has been subjected to severe isolation, the Xavante of Brazil (Ward et al., 1996). A limited number of founder haplotypes has also been observed in populations speculated to have undergone population bottlenecks, like the Kuna and Ngöbé of Panama (Batista et al, 1995; Kolman et al. 1995). However, due to the maintenance of rare haplotypes since the bottleneck, it appears more likely that the limited number of founding haplotypes observed in the Havasupai is a reflection of the small initial population size rather than the effect of the recent bottleneck.

Estimations and comparisons of diversity

The level of genetic variation in the Havasupai was estimated using two standard measures of diversity: haplotype diversity and nucleotide diversity. Haplotype or genetic diversity (h) estimates the probability that two randomly chosen haplotypes from

the sample are different (Nei, 1987). Nucleotide diversity was estimated by two different statistics, π, the mean number of differences between all pairs of haplotypes in the sample (Nei, 1987), and E(ν), an estimate of the level of nucleotide polymorphism (based on the number of segregating sites) representing the proportion of nucleotide sites that are expected to be polymorphic in any sample (Watterson, 1975). Using these indices, a quantitative measure of the loss, or maintenance, of genetic variation within the Havasupai, was ascertained.

Comparisons of genetic diversity across CR segments revealed several interesting trends. First, haplotype diversity in the Havasupai for all three regions analyzed (HVS-I, HVS-II, and HVS-I+I) was considerably higher than was initially expected (table 1.4). This was likely due to the high incidence of rare haplotypes observed. Another interesting observation was the dichotomy seen in the two estimates of nucleotide diversity (π and E(ν)). Estimates of E(ν) were higher than expected in the HVS-I and HVS-I+II regions, whereas estimates of π were predictably low in all regions analyzed. As π is not affected by rare polymorphisms, it serves principally as an indicator of population size. Hence, the low estimates of π observed in the Havasupai probably represent the low long-term effective population size as opposed to the loss of diversity. This, of course, is supported by historical evidence, which purports that the entire Pai population numbered less than 1500 in the 1850's and the Havasupai, once separated from the rest of the Pai, had an initial population of only 250-300 members (Martin, 1986). In contrast, E(ν), while affected by rare polymorphisms, is independent of haplotype frequency and measures a long-term average diversity rather than that of a

current generation (Watterson, 1975). It is, thus, more reflective of changes in demography (e.g. expansion, reduction, and/or selection) and likely represents the higher levels of diversity that were present in the Havasupai in the past. It can be seen from table 1.4 that diversity indices for HVS-I+II are more similar to HVS-I than HVS-II, suggesting that HVS-I might be a better indicator of overall mtDNA CR diversity than HVS-II in this population. Hence, subsequent analyses, excluding phylogenetic analysis, have focused solely on the HVS-I region of the mtDNA CR.

Mismatch distribution analysis

It has been suggested that a smooth, unimodal distribution of pairwise genetic differences (i.e. mismatch distribution) is characteristic of a population growing exponentially, whereas a ragged mismatch distribution can be indicative of stationary populations that have potentially undergone bottlenecks in the past (Harpending, 1994). The raggedness statistic, *r*, designed by Harpending (1994), represents a way to quantify the raggedness of observed mismatch distributions. HVS-I region pairwise genetic difference distributions and their associated raggedness statistics for ten Native American populations are displayed in figure 1.3. The Nuu-Chah-Nulth distribution reveals the lowest raggedness followed (in ascending order) by the Bella Coola, Emberá, Haida, Havasupai, and Wounan, whereas the highest raggedness was seen in the Huetar followed (in descending order) by the Kuna, Ngöbé, and Cayapa. Three of the four populations (Huetar, Kuna, and Ngöbé) showing high *r* scores have proposed some population bottleneck in their past and all four belong to the same linguistic family (i.e. Chibcha)

and probably share some common ancestry. The relatively smooth mismatch distribution and corresponding low raggedness index observed for the Havasupai, appears to further support the observation that the recent population bottleneck had little or no detectable effect on the genetic diversity of the tribe.

AMOVA and interpopulation variance

The results of the AMOVA test revealed several significant levels of population hierarchy existing among the ten populations examined for the HVS-I region (table 1.5). The one group with all populations cluster revealed the genetic distinctiveness of all populations in the study. Clustering the populations by geographic criteria also revealed the genetic distinctiveness of each geographic region (Pacific Northwest, Southwest US, and Central/South America) at a significance level of 97%. Previous reports that have shown a pattern of linguistic clustering that is consistent with those made by genetic clustering, at least in Europe, has led to the hypothesis that human linguistic evolution has paralleled genetic differentiation (Cavalli-Sforza et al., 1988; Barbujani and Sokal, 1990; Cavalli-Sforza et al., 1992). This phenomena has not been routinely observed in the Americas, however, where only Central American tribes show good correlation between the linguistic classifications of Greenberg (1987) and genetic affiliation (Barrantes et al., 1989). Somewhat surprisingly, linguistic clustering by spoken language also showed significant interpopulation variance ($p = 0.04$). In contrast, linguistic clustering by language phylum (Amerind vs. Na-Dene) did not yield significant results. This was not unexpected, however, and has been previously documented with a smaller

number of populations (Chakraborty, 1976; Spuhler, 1979; Black et al., 1983; Batista et al., 1995). Clustering the Havasupai as one group, compared with another group containing all other populations, did not reveal significant interpopulation variance (p = 0.09).

Another test of genetic divergence, based on pairwise interpopulation variance, supported the AMOVA finding of significant genetic distinctiveness between each population in this analysis (table 1.6). Interpopulation variances ranged from 6% (Emberá/Wounan) to 70% (Haida/Havasupai) with an average 50% variance between the Havasupai and all other populations. The Havasupai are most closely related to the Wounan (23%) and most distantly related to the Haida (70%). Geographic distance appeared to have little relationship to genetic variance in individual populations. The Bella Coola (Pacific Northwest) and Ngöbé (Panama), though geographically very distant, show a variance of only 5% (the lowest observed), whereas the Emberá (Panama) and Huetar (Costa Rica) are geographically quite close, even speaking the same language (Chibcha), yet reveal a genetic variance of 29%. It is interesting that the Ngöbé population shows extremely low levels of interpopulation variation with both the Bella Coola (5%) and Nuu-Chah-Nulth (8%) of the Pacific Northwest, which suggests a relatively significant amount of gene flow between these populations. This level of gene flow is not mirrored in the other Chibcha populations, although the levels of variation between Amerind Pacific Northwest populations and Central/South American Chibcha populations generally show some moderate level of gene flow.

Wright (1978) has proposed that 5-15% variance indicates moderate genetic differentiation, while 15-25% variance indicates great genetic differentiation, and above 25% variance indicates very great genetic differentiation. It can be seen that, by far, the highest levels of interpopulation variance are found in the Havasupai, which is to be expected considering that no other populations from the same geographic region were analyzed. Under Wright's assumptions, the Wounan are the only population that do not demonstrate very great genetic differentiation from the Havasupai and even they are on the upper bounds of great amounts of interpopulation variance. With such high levels of genetic differentiation between the Havasupai and all other populations, it is somewhat surprising that the AMOVA did not yield significant results when the Havasupai are grouped against all other populations. All results from these analyses have to be interpreted carefully though, as there are no geographically, linguistically, or culturally related groups to compare to the Havasupai. It would be interesting to look at the Havasupai in comparison with other Native American populations from the southwest United States to determine the level of interpopulation variance between more closely related populations.

Phylogenetic analysis

The phylogenetic tree in figure 1.4 suggests that haplogroup D, represented by Huetar haplotype ht09, and the "other" haplogroup, represented by the Cayapa haplotype cy10, appear to be ancestral to the other haplogroups A, B, and C. However, the lack of multiple constituents for each of these potentially ancestral haplogroups renders any

potential assumptions unreliable until further substantiated with more data. The tree further indicates that haplogroup A is ancestral to B and C and that haplogroup C is the most recently diverged of the major founding haplogroups. Many of the Havasupai haplotypes group closely within their respective haplogroups (see the hv04/ hv05/ hv06 cluster and the hv03/hv07 cluster within haplogroup C) with high bootstrap values as is expected among haplotypes that differ by only a few substitutions. This observation is more pronounced among the haplogroup C haplotypes than those of haplogroup B. Additionally, it appears that all ten unique Havasupai haplotypes are derived from ten discrete maternal lineages that predate the population bottleneck at the turn of the century. Reconstruction of the tree in figure 1.4 employing a gamma correction ($\alpha = .17$ for HVS-I+II (Yang, 1996)) did not significantly alter the topology of the tree or the observed bootstrap values (data not shown).

Summary

In conclusion, the Havasupai Indians of northern Arizona are an interesting and unique population. The low levels of HLA variability reported previously (Markow et al., 1993) were not mirrored, as was predicted, by the mitochondrial DNA data discussed herein. Indeed, despite high levels of inbreeding (Markow and Martin, 1993) and a population bottleneck at the turn of the century that reduced the number of reproducing females and males to 43 and 42, respectively, the tribe maintains a relatively high level of genetic variation. Thus, our original hypothesis, that the Havasupai would exhibit less mtDNA variation than other New World groups, must be rejected. Examination of

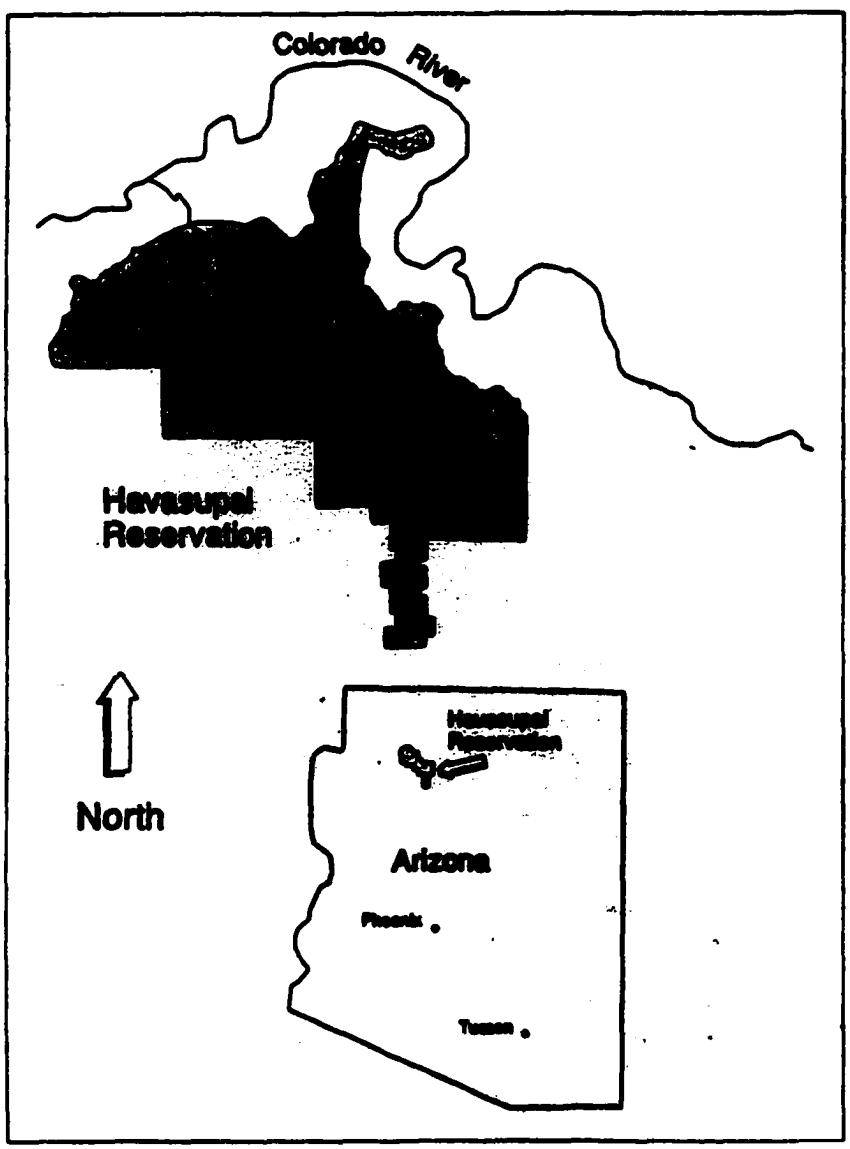
mitochondrial diversity in the Havasupai seems to suggest a picture whereby the Havasupai, initially a small population (as evidenced by the limited representation of founder haplogroups and the low estimate of π), underwent a population expansion at some point that generated a relatively large amount of diversity (as evidenced by the number of rare haplotypes, the high haplotype diversity, and the high estimates of [E(ν)]). This relatively high level of diversity appears to have been unaffected by the recent population bottleneck that occurred at the turn of the century.

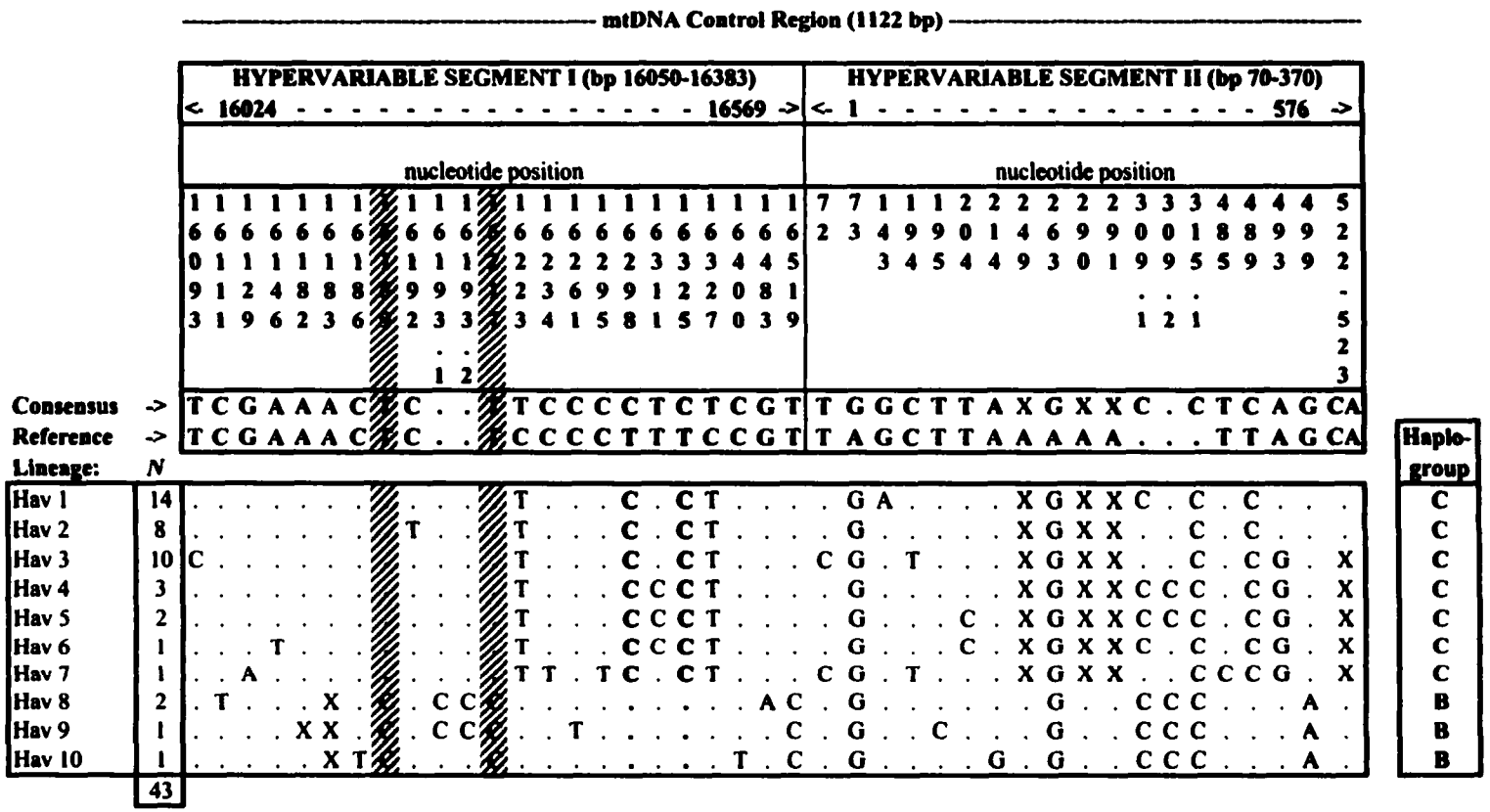
Under this supposition, it would appear that the Havasupai, in a historical perspective, probably maintained relatively high levels of genetic diversity, with periods of expansion that could have resulted in an increased number of rare haplotypes. This increased diversity was probably a result of their association with a larger group of Pai Indians consisting of twelve regional bands, including the Hualapai and the Yavapai, which was disbanded after losing a war to Europeans in 1866 (Martin, 1986). This assumption is supported by the negative values of Tajima's D, though not statistically significant, in the CR. As the level of diversity displayed by the Havasupai seems to have been maintained within the population since the bottleneck at the turn of the century, it must have been too small and/or short to have any detectable effect on the genetic variability of the population.

Investigation into evolutionary relationships between the Havasupai and other Native American populations for which data are available reveals that the Havasupai are quite distinct from the other populations in the sample, showing high levels of interpopulation variance. However, this is to be expected due to the lack of

geographically and/or culturally related populations that have been surveyed for mtDNA sequence variation. Hence, it becomes especially important to investigate more closely related populations, such as the Hualapai and Yavapai, that were part of the larger Pai population to see if they carry any of the ten unique Havasupai haplotypes and at what frequency these haplotypes might exist in other closely related populations.

Figure 1.1: Map of Havasupai Reservation





Reference sequence is that of Anderson et al., 1981.
Consensus sequence is that of the majority of Havasupai without comparison to reference sequence.
An X in any sequence denotes a deletion.
▨ indicates a CR polymorphism characteristic of Haplogroup B (Torroni et al., 1993).
□ indicates a CR polymorphism characteristic of Haplogroup C (Torroni et al., 1993).

Figure 1.2: Polymorphic Nucleotide Positions
Polymorphic nucleotide positions within the mtDNA control region of Havasupai Indians

Table 1.1: Summary of CR Polymorphism in the Havasupai

Summary of nucleotide polymorphism in the Havasupai: A comparison of the three different noncoding regions of the control region (CR) as well as the entire CR

	HVS-I	HVS-II	non-HVS	CR
	bp 16050-16383	bp 70-370	remaining bp	bp 16024-576
Total Number of Sites Expected	334	301	487	1122
Total Number of Insertions	2	3	0	5
Total Number of Sites	336	304	487	1127
Sites with Alignment Gaps (excluded from analysis)	9	12	2	11
Total Number of Sites Analyzed	327	292	485	1116
Total Number of Polymorphic (segregating) Sites	15	6	7	29
Number of Haplotypes	9	6	6	10
Haplotype Diversity ($\hat{h}$) (Nei, 1987)	0.81 ± 0.03	0.75 ± 0.03	0.61 ± 0.05	0.81 ± 0.04
Nucleotide Diversity ($\hat{\pi}$) (Nei, 1987)	0.007 ± 0.002	0.005 ±0.000	0.003 ± 0.000	0.005 ± 0.000
Nucleotide Diversity [$E(v)$] (Watterson, 1975)	3.5 ± 1.3	1.4 ± 0.7	1.6 ± 0.7	6.7 ± 0.7
Tajima's Neutrality Test (Tajima, 1989)	D = -1.07610	D = 0.12962	D = -0.71856	D= -0.87482
Significance	p > 0.10	p > 0.10	p > 0.10	p > 0.10

Table 1.2: Founding Native American Haplogroups

A summary of the nucleotide polymorphisms that classify the four founding Native American haplogroups, as characterized by control region sequencing. The designations used herein are shown in bold.

Haplogroup	Characteristic Polymorphism	Position	Haplotype Equivalence(s)	Reference
A	C -> T	bp 16290	Cluster II	Ward et al. 1991
	G -> A	bp 16319	**Haplogroup A**	**Torroni et al. 1993**
			Cluster III	Horai et al. 1993
			Cluster IV	Ginther et al. 1993
			Cluster III	Bailliet et al. 1994
B	T -> C	bp 16189	Cluster IV	Ward et al. 1991
	T -> C	bp 16217	**Haplogroup B**	**Torroni et al. 1993**
			Cluster I	Horai et al. 1993
			Cluster I	Ginther et al. 1993
			Cluster I	Bailliet et al. 1994
C	T -> C	bp 16298	**Haplogroup C**	**Torroni et al. 1993**
	C -> T	bp 16327	Cluster IV	Horai et al. 1993
			Cluster II	Ginther et al. 1993
			Cluster IV	Bailliet et al. 1994
D	C -> T	bp 16187	Cluster III	Ward et al. 1991
	T -> C	bp 16325	**Haplotype D**	**Torroni et al. 1993**
			Cluster II	Horai et al. 1993
			Cluster III	Ginther et al. 1993
			Cluster II	Bailliet et al. 1994

Table 1.3: Haplogroup Frequencies

Frequencies of the four founding Native American haplogroups among the Havasupai and fourteen other Native American populations

Population	N	Haplogroups A	B	C	D	Other	Region Analyzed for Haplotype Analysis	Reference(s)
Havasupai	43	0.0%	9.3%	90.7%	0.0%	0.0%	complete CR	present study
Bella Coola	40	62.5%	5.0%	7.5%	25.0%	0.0%	HVS-I	Ward et al. (1993)
Cayapa	30	29.0%	40.0%	9.0%	0.0%	22.0%	HVS-I and HVS-II	Rickards et al. (1999)
Embera	44	23.0%	52.3%	25.0%	0.0%	0.0%	HVS-I	Kolman and Bernningham (1997)
Gaviao	27	14.8%	14.8%	0.0%	70.4%	0.0%	HVS-I	Ward et al. (1996)
Haida	40	87.5%	0.0%	7.5%	5.0%	0.0%	HVS-I	Ward et al. (1993)
Huetar	27	70.4%	3.7%	0.0%	25.9%	0.0%	HVS-I and HVS-II	Santos et al. (1994)
Kuna	63	71.4%	28.6%	0.0%	0.0%	0.0%	HVS-I	Batista et al. (1995)
Mapuche	39	15.4%	38.5%	20.5%	25.6%	0.0%	HVS-I and HVS-II	Ginther et al. (1993)
Ngobe	46	67.0%	33.0%	0.0%	0.0%	0.0%	HVS-I and HVS-II	Kolman et al. (1995)
Nuu-Chah-Nulth	63	44.4%	3.2%	19.0%	22.2%	11.1%	HVS-I	Ward et al. (1991)
Wounan	31	29.0%	19.4%	48.4%	3.3%	0.0%	HVS-I	Kolman and Bermingham (1997)
Xavante	25	16.0%	84.0%	0.0%	0.0%	0.0%	HVS-I	Ward et al. (1996)
Yanomami	50	0.0%	4.0%	62.0%	14.0%	20.0%	HVS-I and HVS-II	Easton et al. (1996)
Zoro	30	20.0%	6.6%	13.3%	63.3%	0.0%	HVS-I	Ward et al. (1996)

Table 1.4: Diversity Indices

An estimate of mtDNA sequence diversity for the HVS-I, HVS-II, and HVS-I+II regions of the CR in the Havasupai and other Native American populations

HVS-I Diversity Indices (bp 16050-16383)

Group	Geographic Region	Population Size	Number of Individuals	Number of Mitochondrial Haplotypes	Number of Nucleotides Surveyed	Number of Polymorphic Sites	Haplotype Diversity (Nei, 1987) h	Nucleotide Diversity (Nei, 1987) π	Nucleotide Diversity (Watterson, 1975) $E(\nu)$	(Tajima, 1989) D	
Havasupai	Southwest US	~600	43	9	327	15	0.81 ± 0.03	0.007 ± 0.002	3.5 ± 1.3	-1.076	p > 0.10
Bella Coola	Pacific Northwest	~600	40	11	327	19	0.90 ± 0.02	0.014 ± 0.001	4.5 ± 1.6	0.074	p > 0.10
Cayapa	Ecuador	~3600	30	8	327	18	0.84 ± 0.04	0.019 ± 0.001	4.5 ± 1.7	1.155	p > 0.10
Embera	Panama	~30000	44	20	327	23	0.94 ± 0.02	0.018 ± 0.001	5.3 ± 1.8	0.364	p > 0.10
Haida	Pacific Northwest	~1200	40	9	327	15	0.69 ± 0.06	0.007 ± 0.002	3.5 ± 1.3	-1.118	p > 0.10
Huetar	Costa Rica	~650	27	7	327	11	0.71 ± 0.07	0.010 ± 0.001	2.9 ± 1.2	0.507	p > 0.10
Kuna	Panama	~65000	63	7	327	10	0.59 ± 0.06	0.010 ± 0.001	2.1 ± 0.9	1.519	p > 0.10
Nuu-Chah-Nulth	Pacific Northwest	~2400	63	27	327	25	0.95 ± 0.01	0.016 ± 0.001	5.3 ± 1.7	1.492	p > 0.10
Ngobe	Panama	~125000	46	6	327	11	0.62 ± 0.06	0.012 ± 0.001	2.5 ± 1.0	-0.029	p > 0.10
Wounan	Panama	~5000	31	14	327	27	0.91 ± 0.03	0.020 ± 0.001	6.8 ± 2.4	-0.103	p > 0.10

HVS-II Diversity Indices (bp 70-370)

Group	Geographic Region	Population Size	Number of Individuals	Number of Mitochondrial Haplotypes	Number of Nucleotides Surveyed	Number of Polymorphic Sites	Haplotype Diversity (Nei, 1987) h	Nucleotide Diversity (Nei, 1987) π	Nucleotide Diversity (Watterson, 1975) $E(\nu)$	(Tajima, 1989) D	
Havasupai	Southwest US	~600	43	6	292	6	0.75 ± 0.03	0.005 ± 0.000	1.4 ± 0.7	0.130	p > 0.10
Cayapa	Ecuador	~3600	30	6	292	8	0.77 ± 0.04	0.007 ± 0.001	2.0 ± 0.9	0.216	p > 0.10
Embera	Panama	~30000	15	8	292	10	0.90 ± 0.05	0.008 ± 0.002	3.1 ± 1.4	-0.742	p > 0.10
Huetar	Costa Rica	~650	27	7	292	9	0.83 ± 0.04	0.011 ± 0.008	2.3 ± 1.0	1.183	p > 0.10
Ngobe	Panama	~125000	46	6	292	7	0.60 ± 0.06	0.006 ± 0.001	1.6 ± 0.7	0.166	p > 0.10
Wounan	Panama	~5000	10	6	292	6	0.84 ± 0.10	0.008 ± 0.001	2.1 ± 1.2	0.544	p > 0.10

HVS-I+II Diversity Indices (bp 16050-16383 + 70-370)

Group	Geographic Region	Population Size	Number of Individuals	Number of Mitochondrial Haplotypes	Number of Nucleotides Surveyed	Number of Polymorphic Sites	Haplotype Diversity (Nei, 1987) h	Nucleotide Diversity (Nei, 1987) π	Nucleotide Diversity (Watterson, 1975) $E(\nu)$	(Tajima, 1989) D	
Havasupai	Southwest US	~600	43	10	619	21	0.81 ± 0.04	0.006 ± 0.001	4.9 ± 1.7	-0.756	p > 0.10
Cayapa	Ecuador	~3600	30	10	619	26	0.86 ± 0.04	0.013 ± 0.001	6.6 ± 2.3	0.907	p > 0.10
Embera	Panama	~30000	15	10	619	27	0.94 ± 0.04	0.014 ± 0.001	8.3 ± 3.3	0.260	p > 0.10
Huetar	Costa Rica	~650	27	9	619	20	0.86 ± 0.04	0.010 ± 0.001	5.2 ± 2.0	0.896	p > 0.10
Ngobe	Panama	~125000	46	9	619	18	0.66 ± 0.07	0.009 ± 0.001	4.1 ± 1.5	1.060	p > 0.10
Wounan	Panama	~5000	10	7	619	23	0.91 ± 0.08	0.014 ± 0.002	8.1 ± 3.6	0.044	p > 0.10

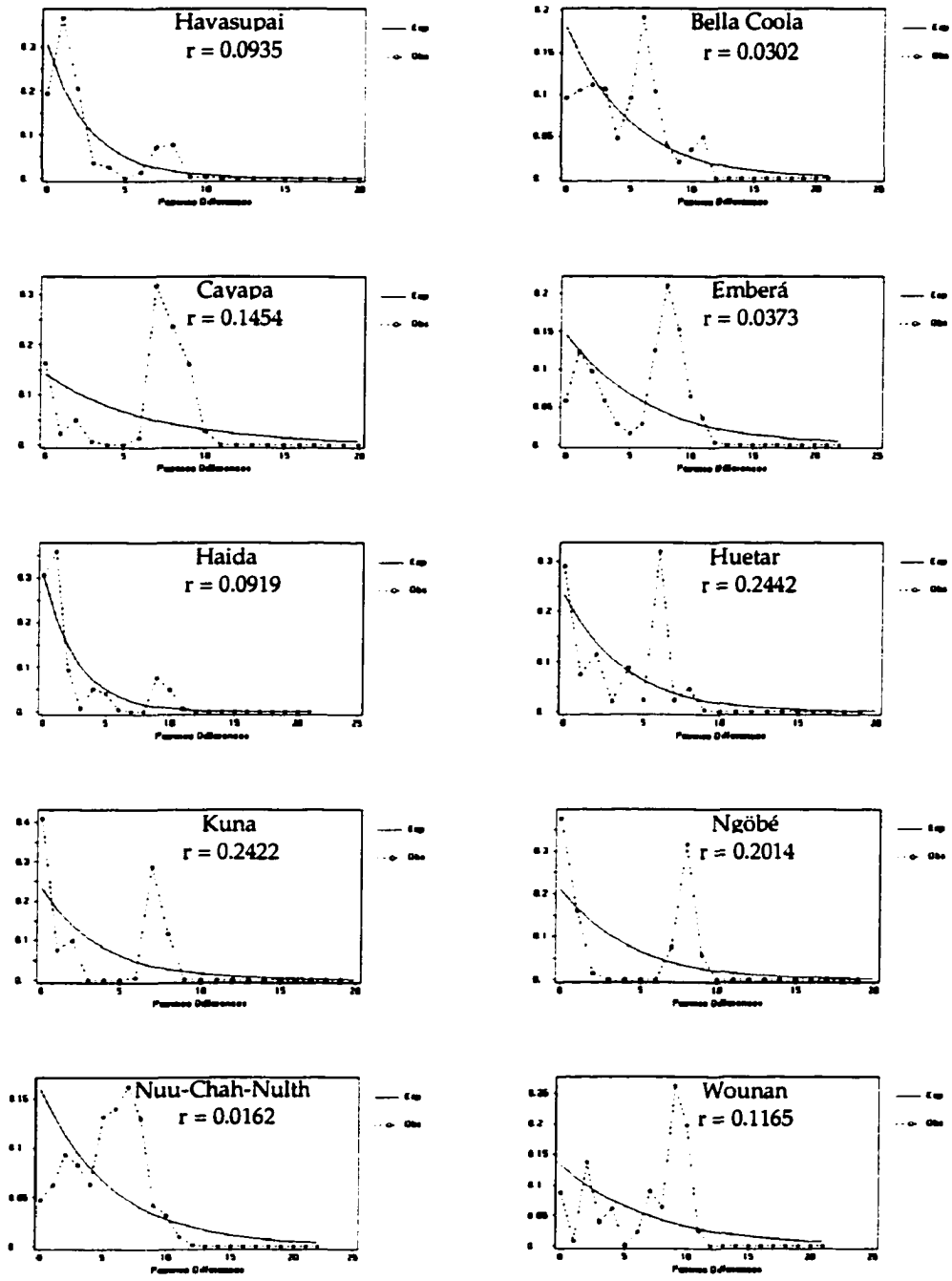


Figure 1.3: Mismatch Distributions
Mismatch distributions for the HVS-1 Region of the mtDNA control region in the Havasapai and other Native American Populations (under the infinite sites model)

Table 1.5: AMOVA
Analysis of MOlecular VAriance (AMOVA) of Native American population substructure

	Within Populations		Among Populations Within Groups		Among Groups	
	Variance		Variance		Variance	
Group	%	*p*	%	*p*	%	*p*
One group with all populations	74.1	< 0.0001				
Geographic clustering	69.6	< 0.0001	14.3	< 0.0001	16.0	0.03
Linguistic clustering	72.3	< 0.0001	10.5	< 0.0001	17.3	0.04
Language phylum clustering	71.5	< 0.0001	23.9	< 0.0001	4.7	0.10
Havasupai clustering	55.4	< 0.0001	12.7	< 0.0001	31.9	0.09

Clusters

One group cluster
- Bella Coola, Cayapa, Embera, Haida, Havasupai, Huetar, Kuna, Ngobe, Nuu-Chah-Nulth, and Wounan

Geographic clusters
- Pacific Northwest: Bella Coola, Haida, and Nuu-Chah-Nulth
- Southwest United States: Havasupai
- South/Central America: Cayapa, Embera, Huetar, Kuna, Ngobe, and Wounan

Linguistic clusters
- Wakashan: Nuu-Chah-Nulth
- Salishan: Bella Coola
- Na-Dene: Haida
- Hokan: Havasupai
- Chibcha: Cayapa, Huetar, Kuna, and Ngobe
- Choco: Embera and Wounan

Language Phylum clusters
- Na-Dene: Haida
- Amerind: Bella Coola, Cayapa, Embera, Havasupai, Huetar, Kuna, Ngobe, Nuu-Chah-Nulth, and Wounan

Havasupai clusters
- Group #1: Havasupai
- Group #2: Bella Coola, Cayapa, Embera, Haida, Huetar, Kuna, Ngobe, Nuu-Chah-Nulth, and Wounan

Table 1.6: Genetic Divergence

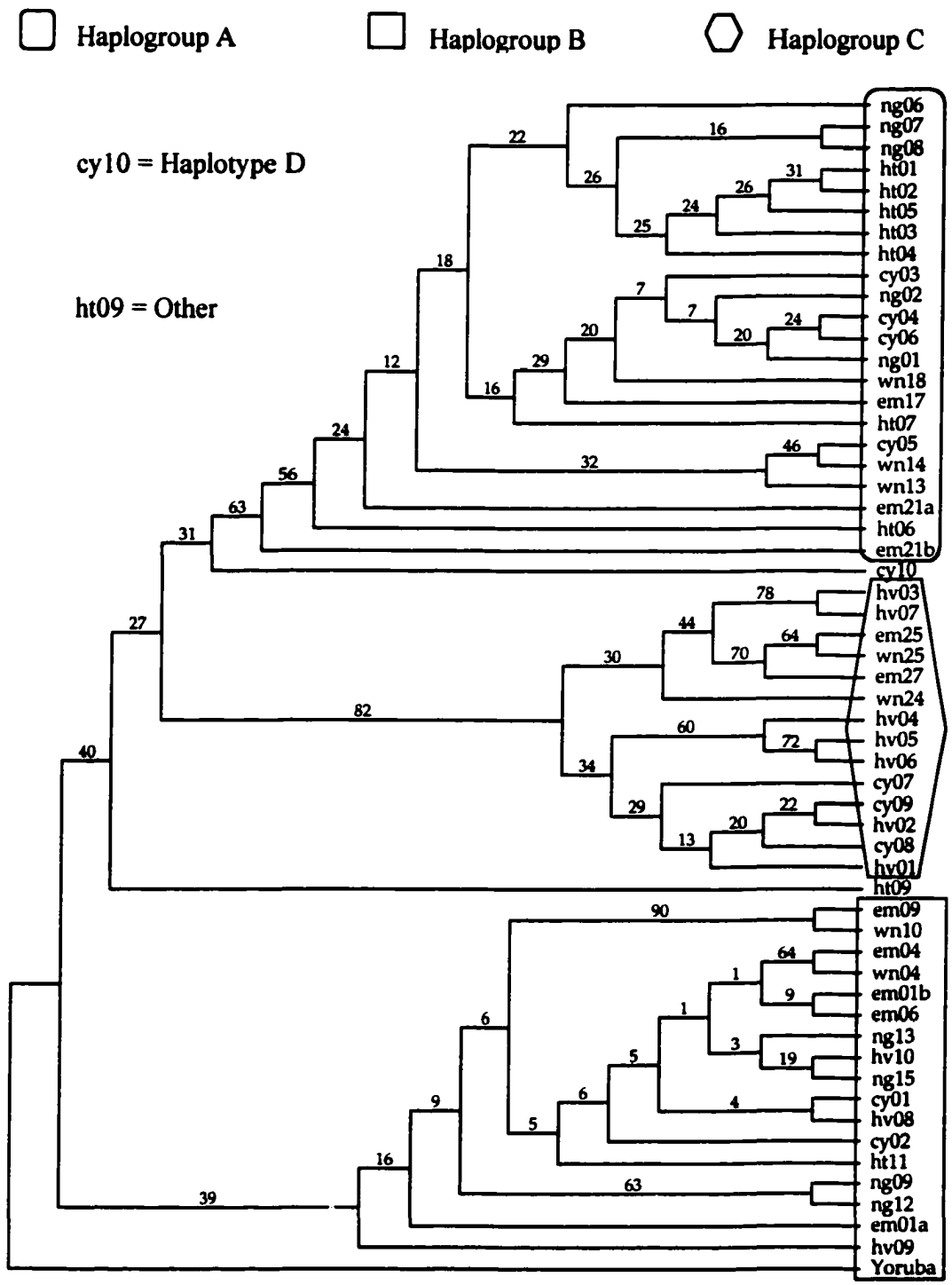
Genetic divergence between ten Native American populations as measured by the percent of interpopulation variance between their HVS-I control region mtDNA sequences

	Bella Coola	Cayapa	Embera	Haida	Havasupai	Huetar	Kuna	Ngobe	Nuu-Chah-Nulth	Wounan
Bella Coola	0%									
Cayapa	13%	0%								
Embera	22%	11%	0%							
Haida	9%	27%	40%	0%						
Havasupai	54%	44%	36%	70%	0%					
Huetar	8%	18%	29%	17%	65%	0%				
Kuna	21%	20%	24%	32%	62%	13% (p = 0.009)	0%			
Ngobe	5% (p = 0.045)	12%	17%	16%	57%	9% (p = 0.018)	14%	0%		
Nuu-Chah-Nulth	7% (p = 0.009)	7% (p = 0.009)	15%	18%	38%	13%	19%	8% (p = 0.009)	0%	
Wounan	20%	11%	6% (p = 0.027)	37%	23%	27%	29%	21%	10%	0%

Note: All values are statistically significant to the $p < 0.0001$ level except where noted above.

Figure 1.4: Phylogenetic Tree

Consensus neighbor-joining tree of all 55 Native American control region haplotypes obtained using the Tamura-Nei correction and 5000 bootstrap replicates.



Chapter 2

MITOCHONDRIAL DNA MUTATIONS ASSOCIATED WITH THE OCCURRENCE OF NIDDM

INTRODUCTION

Background

The heterogeneity of non-insulin-dependent diabetes mellitus (NIDDM) and NIDDM-related complications is so vast that studies have increasingly focused on patient populations manifesting a specific sub-type of NIDDM that may be more homogenous than unselected clinical populations. For example, the elucidation of a defective glucokinase gene in maturity-onset diabetes of the young (MODY) has dramatically emphasized the importance of striving for homogeneity in future genetic studies (Froguel et al., 1992; Vionnet et al., 1992; Permutt et al., 1992; Froguel and Velho, 1993; Saker et al., 1996).

Mutations in mitochondrial DNA (mtDNA) have been associated with NIDDM and NIDDM-related retinopathies in some affected populations (Reardon et al., 1992; Morten et al., 1993; Massin et al., 1995; Suzuki et al., 1997; Vialettes et al., 1997). A subtype of NIDDM that is characterized by maternal transmission and a progressive impairment of insulin secretion is associated with mtDNA mutations and has, thus, been collectively termed mitochondrial diabetes mellitus, or MDM (Gerbitz et al. 1995).

Several loci in mtDNA have been associated with MDM and some of these loci are associated with other diseases as well. Most of these diseases are associated with point mutations, however, there are some disease-related deletions and/or duplications as well. It has been suggested that point mutations in mitochondrial tRNA genes are especially frequent in neuromuscular disorders because of their generalized effect on

mitochondrial protein synthesis and consequent impairment of multiple OXPHOS enzyme complexes (Moraes et al., 1993b). The mitochondrial tRNA Leu(UUR) gene appears to be an etiological hot spot for mtDNA mutations. It contains the primary mitochondrial mutation associated with MDM located at bp 3243 (Reardon et al., 1992; van den Ouweland et al., 1992; Alcolado et al., 1994; Kadowaki et al., 1994; 't Hart et al., 1996; Shigemoto et al., 1998), as well as five other sites (all bp locations based on Anderson et al., 1981): bp 3252 (Morten et al., 1993), bp 3256 (Moraes et al., 1993), bp 3260 (Zeviani et al., 1991), bp 3264 (Suzuki et al., 1997), and bp 3271 (Suzuki et al., 1996), which have been associated with MDM in different populations. Two other tRNA point mutations associated with MDM in selected populations include bp 8296 (Kameoka et al., 1998) and bp 8344 (Suzuki et al., 1994) within the tRNA Lys gene. Mutations in non-coding regions of mtDNA, including a 9 bp direct triplicate repeat at bp 8277 in the intergenic region between cytochrome c oxidase II (COII) and tRNA Lys (Thomas et al., 1996) and a T->C transition at bp 16189 of the control region (Poulton et al., 1998), have also been implicated in MDM in different populations.

tRNA mutations associated with MDM

Results of numerous studies on the 3243 tRNA Leu(UUR) has led to a proposed pathogenic mechanism for this mutation. This mutation occurs within the binding site of mitochondrial termination factor (mTERF) (Kruse et al., 1989). mTERF, when bound, results in the termination of mitochondrial transcription at the 16S rRNA-tRNA Leu(UUR) gene boundary (Christianson and Clayton, 1986), which leads to transcription

of rRNAs that is 50-100 times as frequent as that of more distal genes (Gelfand and Attardi, 1981). The 3243 mutation strongly decreases the affinity of mTERF for its target sequence leading to severe protein synthesis and OXPHOS defects including myopathy, ataxia, and hyperalaninaemia (Shoffner, 1996) which are symptomatic of mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) (King et al., 1992). Although reduced transcription termination resulting from defective mTERF binding could theoretically lead to a lack of 12S and 16S rRNAs, no change in steady state levels of transcripts has been detected either *in vitro* (Montoya et al., 1983; King et al., 1992; Chomyn et al., 1992) or *in vivo* (Moraes et al., 1992a; Hammans et al., 1992; Moraes et al., 1992b).

The exact mechanism whereby the tRNA Leu(UUR) mutations cause protein synthesis defects such as those mentioned previously remains obscure, but it has been hypothesized that they affect tRNA Leu(UUR) synthetase charging specificity resulting in the insertion of the incorrect amino acid (i.e. missense substitutions) and subsequent instability of the irregular polypeptides (Normanly et al., 1989). Another potential mechanism of pathogenesis might involve the formation of 19S RNA, a RNA species consisting of 16S rRNA, tRNA Leu(UUR), and ND1 transcripts, which has elevated levels of expression in another mitochondrial disease, MELAS, resulting from tRNA Leu(UUR) mutations (King et al., 1992).

The tRNA Lys A->G mutation at bp 8296 is located at the start point of the aminoacyl acceptor stem of the tRNA and replaces a highly evolutionarily conserved nucleotide (Kameoka et al., 1998), whereas the A->G bp 8344 mutation occurs at an

unpaired position within the T-loop of the tRNA (Suzuki et al., 1994). It should be noted, however, that the bp 8344 mutation has been widely implicated in the etiology of a mitochondrial myopathy known as myoclonic epilepsy with ragged red fibers, or MERRF (Shoffner et al., 1990; Noer et al., 1991; Lauber et al., 1991; Lertrit et al., 1992; Hammans et al., 1993; Graf et al., 1993; de Vries et al, 1993; Calabresi et al., 1994; Chomyn et al., 1994).

In summary, there is an elevated incidence of mtDNA mutations in MDM. The lack of complete association between these mutants and the manifestation of disease may reflect genetic heterogeneity in the population studied. Studying a homogenous population with a sufficiently high incidence of the disease can reduce genetic heterogeneity. In order to minimize genetic heterogeneity, this study will focus on a small population of Native Americans, the Havasupai, among which the incidence of NIDDM is disproportionately high.

Study population

The Havasupai are a small, Hokan-speaking tribe of approximately 600 members that live on a secluded reservation within the Grand Canyon of northern Arizona (Martin, 1986). Geographical and reproductive isolation (80% of Havasupai marriages are endogamous) have resulted in an extremely genetically homogeneous population (Markow et al., 1993). The tribe has the third largest incidence of NIDDM in the world with 55% of women and 38% of men over the age of 35 diagnosed with the disease (Zuerlein et al., 1991) and a clear maternal bias in disease transmission (Fenger, 1992).

The Havasupai, by virtue of their small numbers and genetic homogeneity, provide an ideal study population in which to overcome the problems of genetic heterogeneity encountered in previous studies. A clear set of affected and unaffected familial lineages has already been identified using detailed genealogies dating back to the mid-1800's. Furthermore, certain Havasupai families show clear evidence of NIDDM-related retinopathy, a condition that has been associated with defined mtDNA mutations in other ethnic groups (Suzuki, 1997).

The maternal transmission of NIDDM in the Havasupai (Fenger, 1992) and the existence of NIDDM-related retinopathies in some pedigrees lead to the hypothesis that members of these affected lineages should exhibit mtDNA mutations associated with NIDDM and NIDDM-related retinopathy in other studies. In order to test this hypothesis, several potentially etiological loci including the tRNA Leu(UUR) gene, the tRNA Lys gene, and the intergenic region between COII and tRNA Lys were examined in 52 Havasupai individuals using automated DNA sequencing. In addition, 43 Havasupai control region (CR) sequences from a previous investigation (Coon and Markow, in preparation) were assayed for potential MDM-associated mutations as well.

MATERIALS AND METHODS

Population samples

Following informed consent, blood samples were obtained on Havasupai reservation between 1993 and 1995. Samples were collected in 15ml Vacutainer tubes

and transported by helicopter to Arizona State University where they were either used immediately to extract DNA or to create immortalized cell lines by Epstein-Barr virus transformation (Markow et al., 1993).

Isolation, amplification and sequencing of DNA

mtDNA-specific PCR primers were designed to amplify candidate regions of the mitochondrial genome including the tRNA Leu (UUR) gene, the tRNA Lys gene, and the intergenic region located between the cytochrome oxidase II gene (COXII) and tRNA Lys. A 288 bp fragment containing the 75 bp tRNA Leu (UUR) gene was amplified using primers mtL3162 (5' CGCCTTCCCCCGTAAATG 3') and mtH3449 (5' GGTTGTAGTAGCCCGTAG 3'). A 258 bp fragment containing the 70 bp tRNA Lys gene and the 28 bp intergenic region was amplified using primers mtL8176 (5' TGAAATCTGTGGAGCAAACC 3') and mtH8433 (5' ATGAGGAATAGTGTAAGGAG 3'). All PCR products were visualized by agarose gel electrophoresis. Subsequent sequence analysis was performed according to Coon and Markow (in preparation).

Statistical analysis

Chi-square tests were performed in order to test the statistical significance of several variables (gender, Native American founder haplogroup, CR haplotype, and specific CR substitutions) associated with the occurrence of NIDDM in the Havasupai population. Although the *G*-statistic is thought to be better than the chi-square in many

instances where experimental design is more complex (e.g. where frequencies are divided into more than two classes), the chi-square test is generally numerically similar to that obtained with the *G*-statistic (Sokal and Rohlf, 1981). Thus, in single classification tests such as those performed herein, chi-square is the preferred method of testing statistical significance.

RESULTS

Automated DNA sequencing of the tRNA Leu(UUR) and tRNA Lys genes, as well as the COII-tRNA Lys intergenic region, from 52 Havasupai individuals revealed none of the MDM-related polymorphisms that had been identified in other populations. There was no variation whatsoever in either of the tRNA genes and the only polymorphism in the COII-tRNA Lys intergenic region, a 9 bp deletion, has not been associated with disease in any known populations.

Fourteen of the 52 individuals surveyed (26.9%) contained the 9 bp deletion. This 9 bp deletion is characteristic of individuals belonging to the founder Native American haplogroup B (Toroni et al., 1993). Pooling the tRNA gene and intergenic region sequence data from this investigation with our CR sequence data from a previous investigation (Coon and Markow, in preparation), we end up with a total of 55 individuals (28 males and 27 females) that have been characterized for inclusion into haplogroups. Fourteen of the 27 (51.9%) females in the study are affected with NIDDM while only 13 of the 28 (46.4%) males are affected (table 2.1A). Of these 55 individuals,

which are 25.5% haplogroup B and 74.5% haplogroup C, 27 (49.1%) are affected by NIDDM and 28 (50.9%) are unaffected (table 2.1B). Haplogroup comparisons reveal that 9 of 14 (64.3%) haplogroup B individuals are affected as opposed to 18 of 41 (43.9%) of individuals belonging to haplogroup C.

Ten mtDNA haplotypes (Hav 1 – Hav 10) have been previously characterized using CR sequencing of 43 Havasupai individuals (Coon and Markow, in preparation). Table 2.1C displays the results of an analysis comparing these 43 individuals, for which haplotypes have been determined. It can be observed that haplotype Hav 2, Hav 7, Hav 9, and Hav 10 have high percentages of individuals affected by NIDDM. However, not much information can be deduced from this observation in most groups (Hav 7, Hav 9, and Hav 10) as they each have only one representative sample. In contrast, 7 of 8 (87.5%) individuals displaying the Hav 2 haplotype are affected by NIDDM. Chi-square analysis demonstrated that there is a 95% probability that this substitution is non-randomly associated with the occurrence of NIDDM ($p < 0.5$).

DNA sequence of the mitochondrial CR for each of the 43 individuals analyzed for this region, as well as diagnosis relative to NIDDM, is displayed in figure 2.1. To further explore the association, if any, between CR mutations and NIDDM, we conducted an analysis of each polymorphism on a site by site basis (table 2.2). The only position that associated positively with the occurrence of NIDDM is a C -> T transition at bp 16192. Seven affected individuals (16.3%) were found to have this substitution, whereas only one individual (2.3%) with the substitution was found to be unaffected. Chi-square analysis of this substitution reveals that the C->T transition at bp 16192 associates non-

randomly with the occurrence of NIDDM in the Havasupai ($p < 0.05$). None of the other loci associate positively with the occurrence of NIDDM to any significant degree.

DISCUSSION

Examination of the mitochondrial tRNA Leu(UUR) and tRNA Lys genes in 52 Havasupai by DNA sequencing revealed no observable mutation. Thus, the high incidence of NIDDM observed in the Havasupai must be the result of mutation at some other unexamined mtDNA loci, or is, perhaps not related to mtDNA mutations at all. The COII-tRNA Lys intergenic region demonstrated the presence of a 9 bp deletion that is characteristic of Native American founder haplogroup B, but not known to be associated with NIDDM in any population. The prevalence of haplogroup B, as characterized by the 9 bp deletion (26.9%) is almost three times the percentage of haplotype B individuals found in our earlier study, as characterized by CR sequencing, which suggested that the Havasupai contain only 9.3% haplogroup B (Coon and Markow, in preparation). When the two data sets are pooled (55 individuals total), 25.5% of individuals belong to haplogroup B. Although there is no evidence to suggest that the 9 bp deletion designating haplogroup B has any association with NIDDM, it is interesting to note that there is an increase, though not statistically significant, in the occurrence of NIDDM in haplogroup B individuals (64.3%) compared with those belonging to haplogroup C (43.9%) in this population. Another polymorphism characteristic of haplogroup B, a T -> C transition at bp 16189 in the CR, has, however, been associated with NIDDM in

several populations. More specifically, the 16189 polymorphism has been associated with mutations within the tRNA Leu(UUR) gene that are often implicated in MDM (Poulton et al., 1998). The Havasupai contain 14 individuals that carry the 16189 variant and 9 of those 14 are affected by NIDDM.

The 16189 T->C transition results in the creation of a heteroplasmic poly-cytosine (poly-C) tract of varying length. The 16189 length variants are not exclusively associated with disease mutations. However, the frequency of this mutation is estimated to be 15% in the normal population (Bendall and Sykes, 1995) compared with 45% (Marchington et al., 1996) to 55% (Poulton et al., 1998) reported in disease populations. Note also that the 3243 mutation (within the tRNA Leu gene) does not segregate with any particular length variant (Marchington et al., 1996) and that the same proportions of length variation are found within all related individuals (Bendall and Sykes, 1995) and within all tissues from the same individual (Poulton et al., 1998).

Further support for an association between the 16189 T->C transition and MDM comes from a report that found that 11% of diabetic men studied had the 16189 variant and that the prevalence of the polymorphism was positively correlated with fasting insulin concentration (Poulton et al. 1998). Thus, it has been suggested that the 16189 variant may indicate a propensity for the occurrence or fixation of tRNA Leu(UUR) mutations that could result in MDM. It should be noted, however, that the T->C mutation at bp 16189 is considered common in some ethnic groups and is associated with insulin resistance rather than NIDDM proper, and is thus not considered a true causal mutation (Poulton et al., 1998).

Examination of the ten mtDNA haplotypes (Hav 1 – Hav 10) characterized by CR sequencing (table 2.1C) revealed that 7 of 8 (87.5%) individuals displaying the Hav 2 haplotype are affected by NIDDM, whereas only 1 is unaffected ($p < 0.05$). Analysis of the CR polymorphisms on a site-by-site basis demonstrated only one site with a statistically significant association with NIDDM. A C -> T transition at bp 16192 was found in 7 individuals (16.3%) affected by NIDDM compared to only one unaffected individual (2.3%) ($p < 0.05$). It is interesting to note that the presence of this substitution is what predominantly characterizes haplotype Hav 2. In other words, haplotype Hav 2 is distinguished from its most closely related haplotype, Hav 1, mainly by the presence of the bp 16192 substitution. The mechanism by which this substitution could affect a NIDDM phenotype is unknown. It is not located in any region of the CR that is known to have specific function, replication initiation for example. However, it does lie in close proximity to a 12-15 bp sequence (at bp 16158-16172) that is associated with termination of mtDNA replication (Doda et al., 1981). Additionally, it lies very close to the bp 16189 substitution, which has been associated with MDM in other studies (Marchington et al., 1996; Poulton et al., 1998). Perhaps it could affect the mtDNA molecule in the same, albeit unknown, way that the bp 16189 substitution does. If the relationship between the bp 16189 substitution and the occurrence of NIDDM is somehow locus and/or sequence specific, the close proximity of the bp 16192 substitution could render a similar result.

In summary, there appears to be some association between mitochondrial DNA mutations and NIDDM in the Havasupai population. There is a slightly higher incidence of NIDDM in individuals belonging to founder haplogroup B than that found in

haplogroup C individuals. In addition, the incidence of haplogroup B in the Havasupai is higher than was previously reported (Coon and Markow, in preparation), though the predominance of haplogroup C was clearly supported. One Havasupai lineage (Hav 2), which is defined by the presence of a C -> T transition at bp 16192 in the CR, was found to have a statistically significant increase in the incidence of NIDDM compared with other lineages. However, the mechanism (if any) by which this polymorphism contributes to the etiology of NIDDM remains to be elucidated. Further research, examining more of the mitochondrial molecule, might prove fruitful in illuminating the potential association between NIDDM and mtDNA mutations in this and other populations.

Figure 2.1: Control Region Mutations
Control region mutations found in 43 Havasupai, sorted by incidence of NIDDM

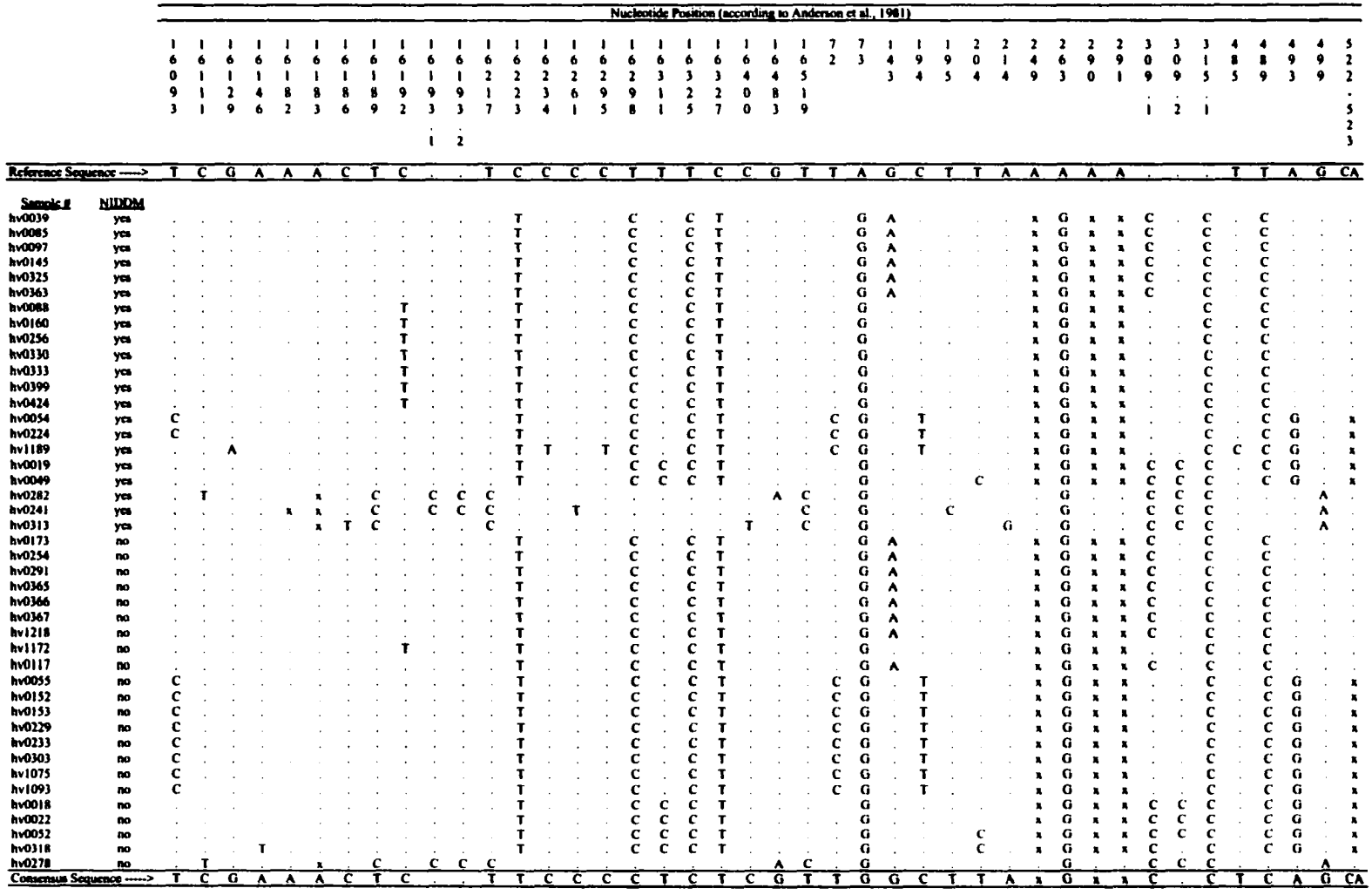


Table 2.1: Association of Havasupai Subgroups to NIDDM

Association of gender, Native American founder haplogroup, and control region haplotype to the occurrence of NIDDM in the Havasupai

		Affected		Unaffected		Total	Expected Value	chi-		Statistical
	Group	n	%	n	%	N	(N/2)	square	df	Significance
A.	Male	13	23.6	15	27.3	28	14	0.14	1	$p > 0.05$
	Female	14	25.5	13	23.6	27	13.5	0.04	1	$p > 0.05$
	Total	27		28		55	27.5	0.02	1	$p > 0.05$
B.	Haplogroup B	9	16.4	5	9.1	14	7	1.14	1	$p > 0.05$
	Haplogroup C	18	32.7	23	41.8	41	20.5	0.61	1	$p > 0.05$
	Total	27		28		55	27.5	0.02	1	$p > 0.05$
C.	Hav #1	6	14	8	18.6	14	7	0.29	1	$p > 0.05$
	Hav #2	**7**	**16**	**1**	**2.3**	**8**	**4**	**4.50**	**1**	**$p < 0.05$**
	Hav #3	2	5	8	18.6	10	5	3.60	1	$p > 0.05$
	Hav #4	1	2	2	4.7	3	1.5	0.33	1	$p > 0.05$
	Hav #5	1	2	1	2.3	2	1	0.00	1	$p > 0.05$
	Hav #6	0	0	1	2.3	1	0.5	1.00	1	$p > 0.05$
	Hav #7	1	2	0	0.0	1	0.5	1.00	1	$p > 0.05$
	Hav #8	1	2	1	2.3	2	1	0.00	1	$p > 0.05$
	Hav #9	1	2	0	0.0	1	0.5	1.00	1	$p > 0.05$
	Hav #10	1	2	0	0.0	1	0.5	1.00	1	$p > 0.05$
	Total	21		22		43	21.5	0.02	1	$p > 0.05$

This table shows the association of gender (1A), Native American founder haplogroup (1B), and control region haplotype (1C) on the occurrence of NIDDM in the Havasupai, along with statistical significance as determined by the chi-square test. Percent affected/unaffected is calculated by dividing affected/unaffected sample size (n) by the total sample size (N = 55 for 1A and 1B and N =43 for 1C). The haplotype Hav #2 appears to correlate significantly with the occurance of NIDDM.

Table 2.2: Association of Individual CR Mutations to NIDDM

Nucleotide Position	Polymorphism	Affected n	Affected %	Unaffected n	Unaffected %	Total N	Expected Value (N/2)	chi-square	df	Statistical Significance
16093	T -> C	2	4.7	8	18.6	10	5.0	3.60	1	p > 0.05
16111	C -> T	1	2.3	1	2.3	2	1.0	0.00	1	p > 0.05
16129	G -> A	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
16146	A -> T	0	0.0	1	2.3	1	0.5	1.00	1	p > 0.05
16182	delete A	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
16183	delete A	3	7.0	1	2.3	4	2.0	1.00	1	p > 0.05
16186	C -> T	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
16189	T -> C	3	7.0	1	2.3	4	2.0	1.00	1	p > 0.05
16192	**C -> T**	**7**	**16.3**	**1**	**2.3**	**8**	**4.0**	**4.50**	**1**	**p < 0.05**
16193.1	insert C	2	4.7	1	2.3	3	1.5	0.33	1	p > 0.05
16193.2	insert C	2	4.7	1	2.3	3	1.5	0.33	1	p > 0.05
16217	T -> C	3	7.0	1	2.3	4	2.0	1.00	1	p > 0.05
16223	C -> T	18	41.9	21	48.8	39	19.5	0.23	1	p > 0.05
16234	C -> T	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
16261	C -> T	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
16295	C -> T	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
16298	T -> C	18	41.9	21	48.8	39	19.5	0.23	1	p > 0.05
16311	T -> C	2	4.7	4	9.3	6	3.0	0.67	1	p > 0.05
16325	T -> C	18	41.9	21	48.8	39	19.5	0.23	1	p > 0.05
16327	C -> T	18	41.9	21	48.8	39	19.5	0.23	1	p > 0.05
16400	C -> T	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
16483	G -> A	1	2.3	1	2.3	2	1.0	0.00	1	p > 0.05
16519	T -> C	3	7.0	1	2.3	4	2.0	1.00	1	p > 0.05
72	T -> C	3	7.0	8	18.6	11	5.5	2.27	1	p > 0.05
73	A -> G	21	48.8	22	51.2	43	21.5	0.02	1	p > 0.05
143	G -> A	6	14.0	8	18.6	14	7.0	0.29	1	p > 0.05
194	C -> T	3	7.0	8	18.6	11	5.5	2.27	1	p > 0.05
195	T -> C	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
204	T -> C	1	2.3	2	4.7	3	1.5	0.33	1	p > 0.05
214	A -> G	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
249	delete A	18	41.9	21	48.8	39	19.5	0.23	1	p > 0.05
263	A -> G	21	48.8	22	51.2	43	21.5	0.02	1	p > 0.05
290	delete A	18	41.9	21	48.8	39	19.5	0.23	1	p > 0.05
291	delete A	18	41.9	21	48.8	39	19.5	0.23	1	p > 0.05
309.1	insert C	11	25.6	13	30.2	24	12.0	0.17	1	p > 0.05
309.2	insert C	5	11.6	4	9.3	9	4.5	0.11	1	p > 0.05
315.1	insert C	21	48.8	22	51.2	43	21.5	0.02	1	p > 0.05
485	T -> C	1	2.3	0	0.0	1	0.5	1.00	1	p > 0.05
489	T -> C	18	41.9	21	48.8	39	19.5	0.23	1	p > 0.05
493	A -> G	5	11.6	12	27.9	17	8.5	2.88	1	p > 0.05
499	G -> A	3	7.0	1	2.3	4	2.0	1.00	1	p > 0.05
522-523	delete CA	5	11.6	12	27.9	17	8.5	2.88	1	p > 0.05

This table shows the frequency of control region substitutions, and their association with NIDDM, in affected versus unaffected Havasupai individuals. Statistical significance was determined by the chi-square test. Percent affected/unaffected was calculated by dividing affected/unaffected sample size (n) by the total sample size (N = 43). Only one substitution (a C->T transition at bp 16192) appears to associate significantly ($p < 0.05$) with the occurance of NIDDM.

Chapter 3

Y-CHROMOSOME VARIATION IN THE HAVASUPAI

INTRODUCTION

It has been suggested that mitochondrial DNA (mtDNA) diversity is greater than Y-chromosome diversity within populations (Oota et al., 2001). Several hypotheses have been proposed to account for this discordance including higher mortality in males as a result of their role as hunters and fighters (Pena et al., 1995), polygyny (Pena et al., 1995; Fix, 1999; Hammer et al., 2001), and patrilocality (Oota et al., 2001; Hammer et al., 2001). All of these factors could lead to a smaller effective number of males within a population. With this increase of mtDNA diversity and decrease of Y-chromosome diversity within populations, comes a concomitant decrease of mtDNA diversity and increase of Y-chromosome diversity between populations (Jorde et al., 2000; Oota et al., 2001).

We recently examined mtDNA diversity in a small, homogenous population of Native Americans, the Havasupai of northern Arizona, for which genealogies, inclusive of eight generations and dating back to the middle 1800s, are available (Coon and Markow, in preparation). According to these extensive genealogies, the Havasupai underwent a population bottleneck at the turn of the century that reduced the total population from approximately 300 to 166, of which, only 42 males and 43 females reproduced. The results of this investigation revealed that the Havasupai contain only two of the four founding Native American mtDNA haplogroups, and show somewhat reduced levels of nucleotide diversity. On the other hand, they demonstrate relatively high levels of haplotype diversity and maintain a high percentage of rare haplotypes

(Coon and Markow, in preparation). These, somewhat contradictory, results suggest that the effect of the recent bottleneck on the Havasupai, if any, is difficult to discern by examining mtDNA.

Sex-specific bottlenecks, whereby the effect of a population bottleneck is more evident within the genetic history of one sex compared with the other, have been reported (Sajantila et al., 1996). The Finnish population, for example, demonstrates a significantly reduced level of Y-chromosome diversity, while maintaining a high level of mtDNA diversity. One proposed hypothesis for this observation is a 'male-specific' population bottleneck, in which only the founding population of males was significantly reduced (Sajantila et al., 1996). Although this is not the case in the Havasupai (the number of males and females reproducing after the bottleneck is essentially the same), it is possible that the potential effect(s) of the bottleneck could be more discernable in the Y-chromosome, than in the mtDNA, lineages.

It is therefore of great interest to determine what the levels of Y-chromosome diversity in the Havasupai are, and how they compare with those previously reported for mtDNA data (Coon and Markow, in preparation). As it is known that the Havasupai practice both patrilocality (Martin, 1986) and polygyny (http://www.cinprograms.org/people/coloradoriver/havasupai.html), and that a higher male mortality rate is likely, it is hypothesized, as per Oota et al. (2001), that the Havasupai will show reduced levels of Y-chromosome diversity compared to mtDNA. In addition, we hope to observe whether or not the recent population bottleneck had any noticeable effect on Y-chromosome diversity in the Havasupai. In order to test this

hypothesis, 48 Havasupai males were genotyped for the nonrecombining portion of the Y-chromosome (NRY) using three biallelic markers, consisting of single nucleotide polymorphisms (SNPs), and 12 short tandem repeat (STR) microsatellite loci.

MATERIALS AND METHODS

Population samples

Following informed consent, blood samples were obtained on Havasupai reservation between 1993 and 1995. Samples were collected in 15ml Vacutainer tubes and transported by helicopter to Arizona State University where they were either used immediately to extract DNA or to create immortalized cell lines by Epstein-Barr virus transformation (Markow et al., 1993). The genealogies were utilized to eliminate any individuals for whom admixture was known to be present.

Genotyping of Y-chromosome polymorphism

In order to determine Y-chromosome variation, 48 male Havasupai samples were screened for 3 biallelic markers DYS199 (Underhill et al., 1996; Bianchi et al., 1998), DYS257 (Hammer et al., 1998), and ARS72425 (Hammer, unpublished) that define most Native American groups on a haplotype tree (Hammer et al., 2001) as well as 12 STRs including DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS394 (http://www.ystr.org/usa), DYS388 (Kayser et al., 1997), DYS426, and TAGA13 (Redd et al., submitted).

Genotyping of the DYS199 was performed by allele-specific PCR employing methods described elsewhere (Karafet et al., 1999). DYS257 was amplified according to the conditions and primers reported by Hammer et al. (1998) and mutation detection was accomplished by overnight restriction endonuclease digestion with *Ban* I (New England Biolabs). Mutation at this locus results in the loss of the *Ban* I site and the appearance of a single band on an agarose gel. ARS72425 was amplified using a modified touchdown PCR protocol employing a positive control in the same reaction. Multiplex PCR was used to amplify 6 STRs at once. Multiplex I, containing DYS390, DYS391, DYS393, DYS394, DYS385a, and DYS385b, was amplified using the following PCR conditions: 94°/3 min. – (94°/30 sec. - 58°/30 sec. - 72°/30 sec. -> 30 cycles) - 72°/3 min. Multiplex II, containing DYS388, DYS389-1, DYS389-2, DYS426, and TAGA, was amplified using the following PCR conditions: 94°/3 min. – (94°/30 sec. - 56°/30 sec. - 72°/30 sec. -> 35 cycles) - 72°/3 min. All PCR products were visualized by agarose gel electrophoresis and then analyzed on an ABI 3100 automated DNA sequencer employing the GeneScan 3.1.2 analysis software (Applied Biosystems Inc.). Subsequent analysis was performed using the Genotyper 2.5 software (Applied Biosystems Inc.).

Statistical analysis

Statistical significance of haplogroup-specific distribution of alleles was calculated by *t*-test. Both haplotype diversity (*h*) and nucleotide diversity, as estimated by the average number of pairwise differences (*k*), are defined by Nei (1987). These

diversity indices, derived solely from STR variation on the NRY of the Havasupai, were calculated by Arlequin v. 2.0 (Schneider et al., 2000).

RESULTS

Genetic analysis of the NRY of 48 Havasupai individuals using 3 biallelic markers (SNPs) and 12 microsatellite loci (STRs) resulted in the documentation of 13 Havasupai haplotypes spanning 2 haplogroups (table 3.1). Haplogroup 1Ca is the most common with a frequency of 56.25% (n = 27) followed by 1G with a frequency of 43.75% (n = 21). Haplotypes 1-4 are associated with the 1G haplogroup, while haplotypes 5-13 are associated with the 1Ca haplogroup. The most frequently occurring haplotype is haplotype 3 (19%), followed by haplotype 8 (17%), haplotypes 1 and 6 (15% each), haplotype 13 (10%), and haplotype 2 (8%). All other haplotypes (4, 5, 7, and 9-12) are present at less than 4% and are considered rare (i.e. containing only 1 or 2 individuals). Thus, 53.8% of Havasupai haplotypes are rare and 16.7% of Havasupai display rare haplotypes and all but one (haplotype 4) of the rare haplotypes belong to the 1Ca haplogroup.

It has been suggested (Underhill et al., 1996) that the most frequently occurring haplotype is the oldest. This assumption should hold true under both an infinite allele model (Watterson and Guess, 1977) and a stepwise mutation model (Chakraborty, 1977). As the derived haplogroup 1G is one of several monophyletic subgroups of the ancestral, paraphyletic supergroup 1Ca, it is of interest to see how the modal (i.e. most frequently

occurring) haplotypes of the two differ. Table 3.2A shows the modal haplotype of the 1Ca supergroup (including 1G) in the Havasupai. It can be seen that, in the Havasupai, the modal haplotype of 1Ca+1G: DYS390/24, DYS391/11, DYS393/14, DYS394/13, DYS385a/15, DYS385b/16, DYS388/12, DYS392/14, DYS426/12, DYS389-1/13, DYS389-2/30, TAGA/16 is different than the modal haplotype of the 1G haplogroup: DYS390/24, DYS391/9, DYS393/13, DYS394/13, DYS385a/14, DYS385b/20, DYS388/12, DYS392/14, DYS426/12, DYS389-1/14, DYS389-2/30, TAGA/16 (table 2b).

Table 3.3 shows the distribution of alleles and the number of individuals who display those alleles for both haplogroups on a locus by locus basis. Of the 12 STR loci, DYS385a, DYS385b, and DYS392 were the most polymorphic in the Havasupai, displaying four alleles apiece. DYS391 and DYS389-1 both show three alleles, whereas DYS390, DYS393, DYS388, DYS389-2, and TAGA all show only two alleles. Two loci, DYS394 and DYS426, are monomorphic. Observation of these distributions of haplogroup-specific allele frequency for each polymorphic locus in a graphical form (figures 3.1-3.10) clearly demonstrates the preference of certain alleles for certain haplogroups. A two-sample *t*-test assuming unequal variance was used to test the assumption of haplogroup-specific allele distribution and the results are shown in table 3.4. The results of the *t*-tests show that allele distribution at every locus except for two (DYS388 and DYS389-1) show a statistically significant nonrandom association with either 1Ca or 1G (*p* values are shown in shaded boxes). In other words, certain alleles seem to be preferentially associated with certain haplogroups.

These haplogroup-specific allele distributions are exemplified at the DYS391, DYS393, and DYS385a loci. At the DYS391 locus (figure 3.2), the 11-repeat allele is exclusively associated with haplogroup 1Ca, whereas the 9- and 10-repeat alleles are almost, except one 1Ca individual carrying the 10-repeat allele, exclusively associated with the 1G haplogroup ($p = 4.26 \times 10^{-12}$). An even more striking illustration is seen at the DYS393 locus (figure 3.3), where the 13-repeat allele is carried solely by individuals belonging to the 1G haplogroup and the 14-repeat allele is seen almost exclusively, with the exception of one 1G individual, in 1Ca individuals ($p = 1.08 \times 10^{-14}$). Perhaps the most remarkable instance of haplogroup-specific allele distribution is seen at the DYS385a locus (figure 3.4). Although the level of statistical significance of this distribution is lower than that of the DYS393 locus due to the lack of resolution of the *t*-test for extreme versus median values, it can be seen that the 12- and 14-repeat alleles associate prohibitively with the 1G haplogroup, while the 15- and 16-repeat alleles are found exclusively in the 1Ca haplogroup ($p = 1.07 \times 10^{-9}$). Allele distributions at all other loci, except DYS388 and DYS389-1, which were not shown to be statistically significant by the *t*-test, are statistically significant to at least a $p < 0.02$ level.

A comparison of NRY and mtDNA diversity in the Havasupai can be seen in table 3.5. The sample size is roughly the same, 43 and 48 in the mtDNA and NRY studies, respectively. Several more haplotypes were observed in the analysis of the NRY (13 haplotypes) than were observed in mtDNA analysis (10 haplotypes). The Havasupai display 6 rare mtDNA lineages (60.0% of mtDNA lineages are rare) and 7 rare NRY lineages (53.8% are NRY lineages are rare), again roughly the same. The number of

polymorphic loci in the mtDNA is over twice that of the NRY, but that is to be expected when comparing DNA sequencing to STR analysis. Haplotype diversity is slightly higher in the NRY (h = 0.89) than in the mtDNA (h = 0.81) as is nucleotide diversity, as estimated by the average number of pairwise differences (k), k = 4.94 in the NRY compared to k = 3.74 in the mtDNA.

DISCUSSION

Havasupai NRY haplogroup frequencies

SNP and STR analysis of 48 male Havasupai for the NRY revealed the presence of 13 haplotypes belonging to 2 Y-chromosome haplogroups, 1Ca and 1G, as most recently defined by Hammer et al. (2001). 56.25% (n =27) of Havasupai belong to haplogroup 1Ca and 43.75% (n = 21) belong to haplogroup 1G (table 1). Haplogroup 1G is defined by a single polymorphic C->T transition at locus DYS199 on the human Y-chromosome. The frequency of the DYS199T mutation (i.e. haplogroup 1G) ranges from 0.35-0.95 in the Americas and averages 0.376 in North American Indians (Karafet et al., 1998). The Havasupai frequency of 0.438, thus, represents an above average frequency of this allele among North American Indians. This figure correlates closely with previous reports showing that half of Havasupai sampled, albeit with a much smaller sample size (n = 10), were shown to carry the DYS199T variant (Karafet et al, 1997; Bianchi et al., 1998).

Modal haplotypes

Haplogroup 1G is one of several monophyletic subgroups of the haplogroup 1Ca. Therefore, it has been proposed that all individuals presently belonging to haplogroup 1G are presumed to be descendants of a single individual who first displayed the DYS199C->T mutation (Bianchi et al., 1998). This suggests that the first 1G haplotype was probably very similar to the 1Ca haplotypes from which it derived with the exception of DYS199T allele (Bosch et al., 1999). Thus, under the assumption of neutrality, 1G individuals should display a modal haplotype that is very similar to that found in the older 1Ca+1G supergroup, with the exception of DYS199T allele. A "founding" 1G haplotype, OA, has been proposed by Bianchi et al. (1998). The modal haplotype for OA, assuming DYS199T and omitting markers that were not typed in this investigation (αh II and DYS19) is: DYS389-1/10, DYS389-2/27, DYS390/24, DYS391/10, DYS392/14, and DYS393/13 (table 3.2C).

Modal haplotypes derived from Havasupai STR data are presented in table 3.2. It can be seen that the modal haplotype of 1Ca+1G (table 3.2A) differs from that of 1G (table 3.2B) at 5 of the 12 loci (DYS391, DYS393, DYS38a, DYS385b, and DYS389-1). Two of these loci (DYS391 and DYS385b) would most likely require multiple mutations in order to coalesce the two modal haplotypes. It can also be seen that the Havasupai modal 1G (Hav 1G) haplotype differs from OA at 3 of the 6 common loci (DYS391, DYS389-1, and DYS389-2). In fact, a four-step mutation and a three-step mutation separate Hav 1G and OA at the DYS389-1 and DYS389-2 loci, respectively. Thus, eight allelic shifts separate Hav 1G and OA, whereas, only seven allelic shifts separate the

most disparate Native American haplotypes from OA in a much larger sample (n = 89) where two more potentially polymorphic loci (αh II and DYS19) were included (Bianchi et al., 1998). None of the 13 Havasupai haplotypes match the OA exactly.

It is especially interesting to note that the DYS392 locus, while hypervariable in Native Americans, is generally represented by alleles >13 repeats and has not been observed to transgress the 13-11 repeat leap (Forster et al., 2000). Thus, although the DYS392 locus is purportedly prone to double repeat mutations, a jump from 11 to 13 repeats is observed most frequently and the DYS392 allele of 12 repeats, found in one Havasupai individual belonging to haplotype 9, is rarely found anywhere in the world (Kayser et al., 1997). The amount of difference in the modal haplotypes of 1Ca+1G and 1G could suggest that there was some type of selective pressure, perhaps related to mating practices, that might have acted on individuals within their respective haplogroups. Separation of Hav 1G from OA could also reflect the reproductive and geographic isolation of the Havasupai and the subsequent evolution of rare haplotypes. Alternatively, it could merely be a representation of the propensity for an increase of repeats as opposed to the loss (Kayser et al., 2000).

Haplogroup-specific allele distribution

According to de Knijff (2000), haplogroup-specific distribution of alleles is expected in a non-recombining, haploid element such as the Y-chromosome, as fixation of alleles after episodes of genetic drift would be more rapid. In order to determine whether or not this haplogroup-specific distribution of alleles is occurring in the

Havasupai Y-chromosome, we analyzed the allele distribution at each of the twelve STR loci (table 3.3). A graphical representation of allele distributions at each of the ten polymorphic loci clearly demonstrated that allele distribution is indeed haplogroup-specific (figures 3.1-3.10). *t*-tests, performed in order to see if these haplogroup-specific allele distributions were statistically significant, revealed that 8 (DYS390, DYS391, DYS393, DYS385a, DYS385b, DYS392, DYS389-2, and TAGA) of the 10 polymorphic loci showed statistically significant levels of haplogroup-specific distribution of STR alleles to at least a $p = 0.02$ level (table 3.4). The most striking examples, DYS391, DYS393, and DYS385a, show statistical significance to at least the $p = 1.07 \times 10^{-9}$ level.

However, in some instances, the *t*-test statistic, which makes comparisons *t*-test based on averages, tends to underestimate the actual level of haplogroup-specific allele distribution. For instance, the DYS385a locus (figure 3.4) reveals that the 12- and 14-repeat alleles were exclusively associated with haplogroup 1G, while the 15- and 16-repeat alleles are exclusively associated with the 1Ca haplogroup ($p = 1.07 \times 10^{-9}$). Contrarily, the DYS393 locus (figure 3.3), where the 13-repeat allele is carried solely by individuals belonging to the 1G haplogroup and the 14-repeat allele is shared by both haplogroups, albeit not evenly (26:1), shows a much greater level of statistical significance ($p = 1.08 \times 10^{-14}$).

Another example is the DYS385b locus (figure 3.5), which was shown to be statistically significant at only a, comparatively, very low level ($p = 0.02$). At this locus, the 15- and 20-repeat alleles are found only in individuals of the 1G haplotype while the 17-repeat allele is found solely in the 1Ca haplogroup, and the 16-repeat allele is shared

by both haplotypes (20:4). This distribution is very similar to that of DYS393 which shows an extremely high level of statistical significance ($p = 1.08 \times 10^{-14}$). However, due to fact that one group (1G) consists of mostly extreme alleles (15- and 20-repeats) and the other (1Ca) consists of median alleles (16- and 17-repeats), the difference between the two haplogroups gets diluted out as the scores are averaged.

With this in mind, a re-examination of the DYS389-1 locus (figure 3.8), which was shown to be statistically non-significant ($p > 0.05$, $p = .385$), could prove valuable. At this locus also, one haplogroup (1G) shows extreme alleles (12- and 14- repeats) while the other group (1Ca) shows mainly median alleles (13-repeats). Thus, the difference between the two haplogroups at this locus is higher than it appears based on the results of the, though perhaps still not statistically significant, due to dilution of scores through averaging. Therefore, it would appear that there is a strong haplogroup-specific bias in the distribution of alleles at all but one of the ten polymorphic loci in the Havasupai.

NRY vs. mtDNA diversity

Examination of the NRY has revealed that, in the Havasupai, the number of haplotypes, 13 for the NRY and 10 for mtDNA, as well as diversity of haplotypes, 0.89 for NRY compared with 0.81 for mtDNA, is greater for the Y-chromosome (table 3.5). Nucleotide diversity, as estimated by the average number of pairwise differences, is also greater for the NRY than mtDNA, 4.94 and 3.74 respectively. Thus, it would appear that, contrary to Oota et al. (2001), the NRY displays slightly higher levels of genetic diversity than does mtDNA, at least within this population and we must, therefore, reject our initial

hypothesis that the NRY will contain reduced genetic diversity compared with mtDNA in the Havasupai. This finding, although contrary to Oota et al. (2001) and the five European populations investigated by Sajantila et al. (1996), is consistent with that of Jorde et al. (2000) who found that NRY diversity was higher than mtDNA diversity in 15 geographically diverse populations worldwide. Additionally, 4 of 5 populations surveyed by Mesa et al. (2000) demonstrated greater genetic diversity in the NRY than in mtDNA.

Due, however, to the lack of a between-group comparison, these results have to be interpreted carefully. Specifically, as there are many different factors that contribute to total diversity, we cannot know whether the difference in diversity seen between the Y-chromosome and mtDNA of the Havasupai is the result of differences in mutation rate or cultural differences (e.g. polygyny) or a true difference in relative diversity. Examination of another (preferentially closely related) population such as the Hualapai or the Yavapai would allow the desired between-group comparisons and make estimates of relative diversity more reliable.

In summary, it would appear that the proposed hypothesis, that mtDNA diversity is greater than Y-chromosome diversity within populations, proposed by Oota et al. (2001), is not true for the Havasupai Indians of northern Arizona. Analysis of Havasupai paternal lineages showed no significant reduction in NRY diversity compared to mtDNA. Indeed, the NRY displayed higher levels of diversity than did mtDNA, though again, it should be firmly established that, it is not possible to distinguish between the forces shaping the observed disparity in the levels of diversity between mtDNA and the Y-chromosome. Additionally, the haplotypes found among the Havasupai did not associate

significantly with the OA founder haplotype reported by Bianchi et al. (1998), perhaps suggesting some sort of selective pressure, or the effects of geographic/reproductive isolation, acting on the Havasupai. As expected, a pronounced haplogroup-specific distribution of alleles at most of the STR loci examined was demonstrated. Lastly, the relatively high levels of Y-chromosome diversity seen in the Havasupai would appear to suggest that the population bottleneck undergone by the Havasupai at the turn of the century had little effect on the population.

Table 3.1: Y-chromosome Haplotypes

			Single Nucleotide Polymorphisms (SNPs)				Microsatellite Short Tandem Repeats (STRs)											
Haplo-type	n	Freq.	SY103/ DYS199	SY57/ DYS257	ARS 72425	Haplo-group	DYS390	DYS391	DYS393	DYS394	DYS385a	DYS385b	DYS388	DYS392	DYS426	DYS389-1	DYS389-2	TAGA
1	7	0.15	+	+	+	1G	24	9	13	13	14	15	12	14	12	14	30	16
2	4	0.08	+	+	+	1G	24	9	13	13	14	16	12	14	12	14	30	16
3	9	0.19	+	+	+	1G	24	10	13	13	12	20	12	13	12	12	29	16
4	1	0.02	+	+	+	1G	24	10	14	13	12	20	12	13	12	12	29	16
5	2	0.04	-	+	+	1Ca	23	11	14	13	15	16	12	14	12	13	30	15
6	7	0.15	-	+	+	1Ca	23	11	14	13	15	16	12	14	12	13	30	16
7	1	0.02	-	+	+	1Ca	24	10	14	13	16	16	12	15	12	13	30	15
8	8	0.17	-	+	+	1Ca	24	11	14	13	15	16	12	14	12	13	30	15
9	1	0.02	-	+	+	1Ca	24	11	14	13	15	16	12	12	12	13	30	15
10	1	0.02	-	+	+	1Ca	24	11	14	13	15	16	12	13	12	13	30	15
11	1	0.02	-	+	+	1Ca	24	11	14	13	15	17	12	15	12	14	30	15
12	1	0.02	-	+	+	1Ca	24	11	14	13	16	17	11	15	12	14	30	16
13	5	0.10	-	+	+	1Ca	24	11	14	13	16	17	12	15	12	14	30	15

This table shows the 13 observed Havasupai Y-chromosome haplotypes, representing two Native American haplogroups (1G and 1Ca), and their respective frequencies within this population.

Table 3.2: Modal STR-based Y-chromosome Haplotypes

	Haplogroup		STR Loci											
			DYS390	DYS391	DYS393	DYS394	DYS385a	DYS385b	DYS388	DYS392	DYS426	DYS389-1	DYS389-2	TAGA
A.	1Ca + 1G	Low	23	9	13	13	12	15	11	12	12	12	29	15
		High	24	11	14	13	16	20	12	15	12	14	30	16
		Median	23.8	10.3	13.6	13.0	14.3	16.8	12.0	13.9	12.0	13.2	29.8	15.6
		Mode	24	11	14	13	15	16	12	14	12	13	30	16

			STR Loci											
			DYS390	DYS391	DYS393	DYS394	DYS385a	DYS385b	DYS388	DYS392	DYS426	DYS389-1	DYS389-2	TAGA
B.	1G only	Low	24	9	13	13	12	15	12	13	12	12	29	16
		High	24	10	14	13	14	20	12	14	12	14	30	16
		Median	24.0	9.5	13.0	13.0	13.0	17.6	12.0	13.5	12.0	13.0	29.5	16.0
		Mode	24	9	13	13	14	20	12	14	12	14	30	16

		STR Loci											
		DYS390	DYS391	DYS393	DYS394	DYS385a	DYS385b	DYS388	DYS392	DYS426	DYS389-1	DYS389-2	TAGA
C.	OA (Bianchi et al., 1998)	24	10	13					14		10	27	

This table shows the median and modal number of repeats at each loci for the STR-based haplotypes found in the Havasupai. The OA modal haplotype has been described as the ancestral founder haplotype by Bianchi et al., 1998.

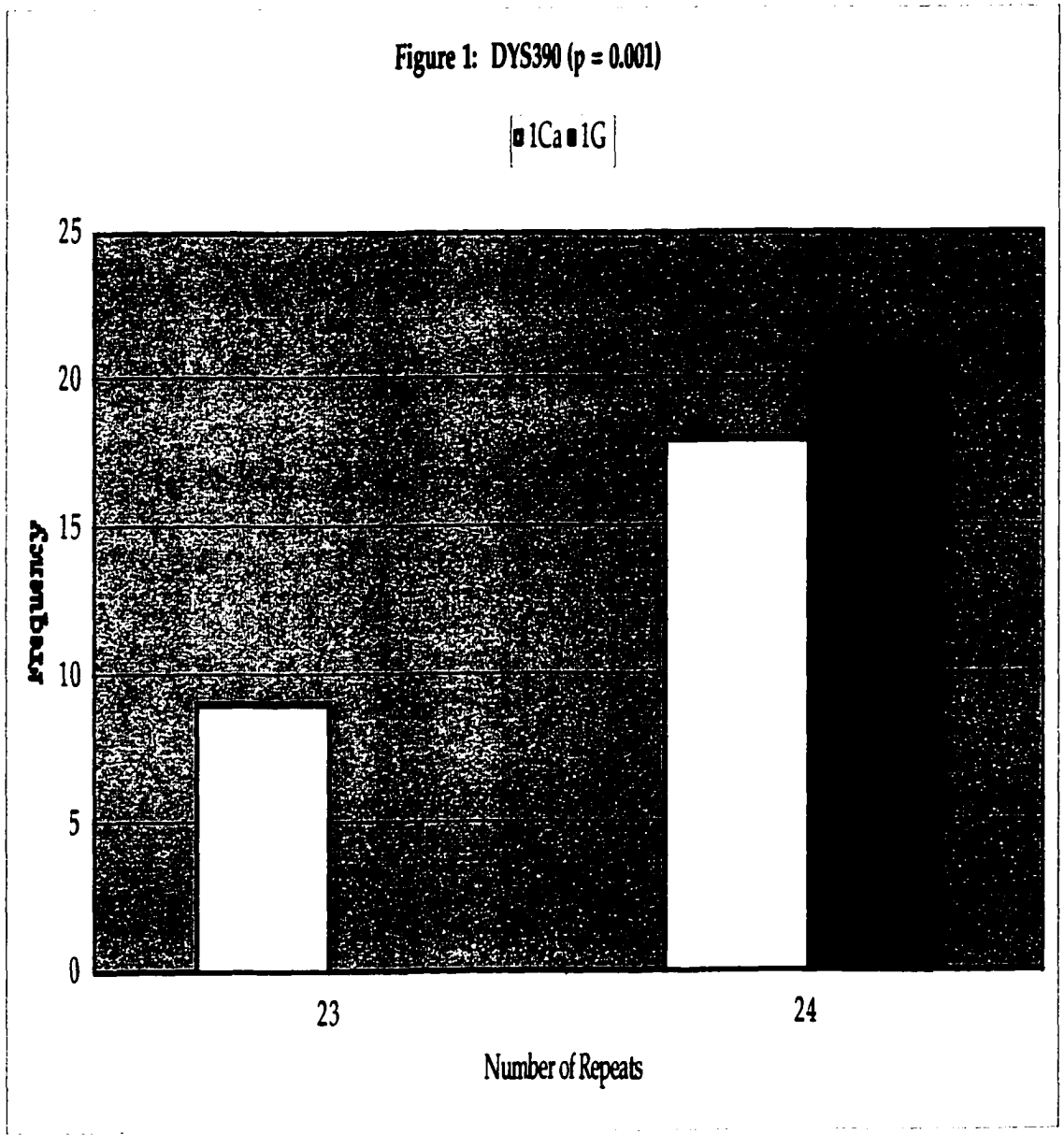
Table 3.3: Comparison of STR Repeat Number at 10 Polymorphic Loci

Tabular representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY

	DYS390				DYS391				DYS393				DYS385a					DYS385b				
		# of repeats				# of repeats				# of repeats				# of repeats					# of repeats			
		23	24	n		9	10	11	n		13	14	n		12	14	15	16	n		15	16	17	20	n
Haplotype	**1Ca**	9	18	27	**1Ca**	0	1	26	27	**1Ca**	0	27	27	**1Ca**	0	0	20	7	27	**1Ca**	0	20	7	0	27
Haplotype	**1G**	0	21	21	**1G**	11	10	0	21	**1G**	20	1	21	**1G**	10	11	0	0	21	**1G**	7	4	0	10	21

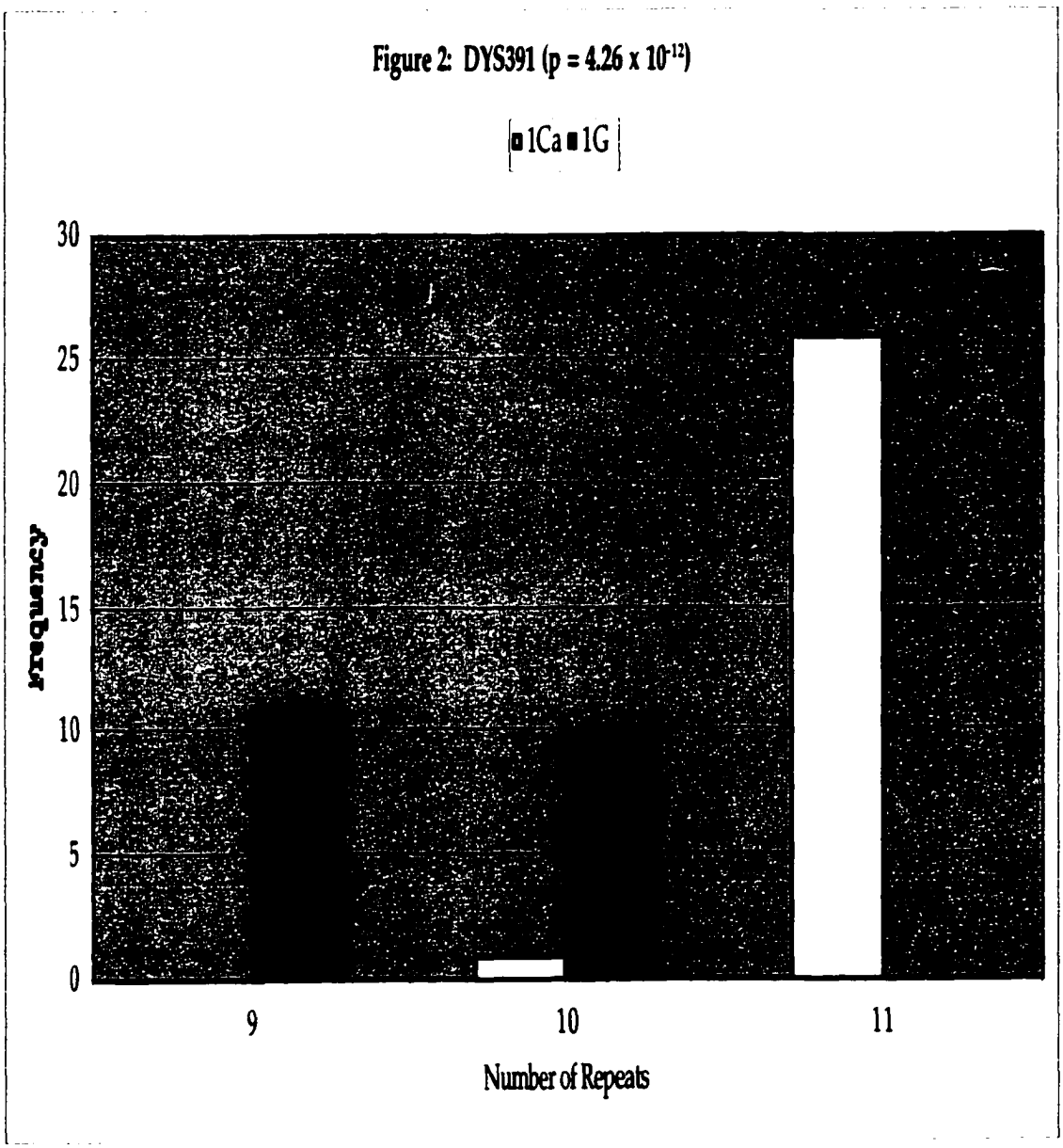
	DYS388				DYS389-1				DYS389-2				DYS392					TAGA			
		# of repeats				# of repeats				# of repeats				# of repeats					# of repeats		
		11	12	n		12	13	14	n		29	30	n		12	13	14	15	n		15	16	n
Haplotype	**1Ca**	1	26	27	**1Ca**	0	20	7	27	**1Ca**	0	27	27	**1Ca**	1	1	17	8	27	**1Ca**	19	8	27
Haplotype	**1G**	0	21	21	**1G**	10	0	11	21	**1G**	10	11	21	**1G**	0	10	11	0	21	**1G**	0	21	21

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



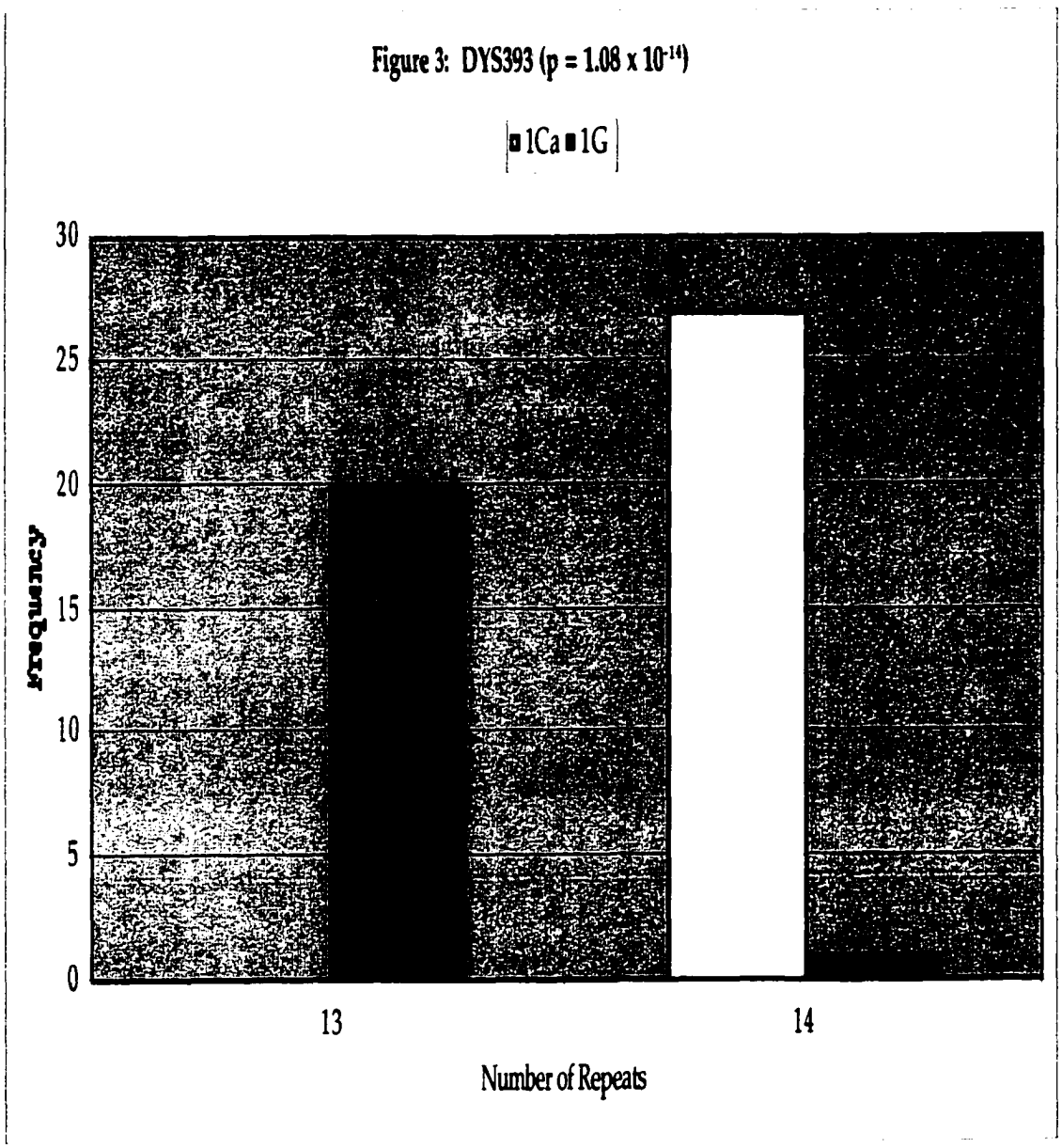
Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



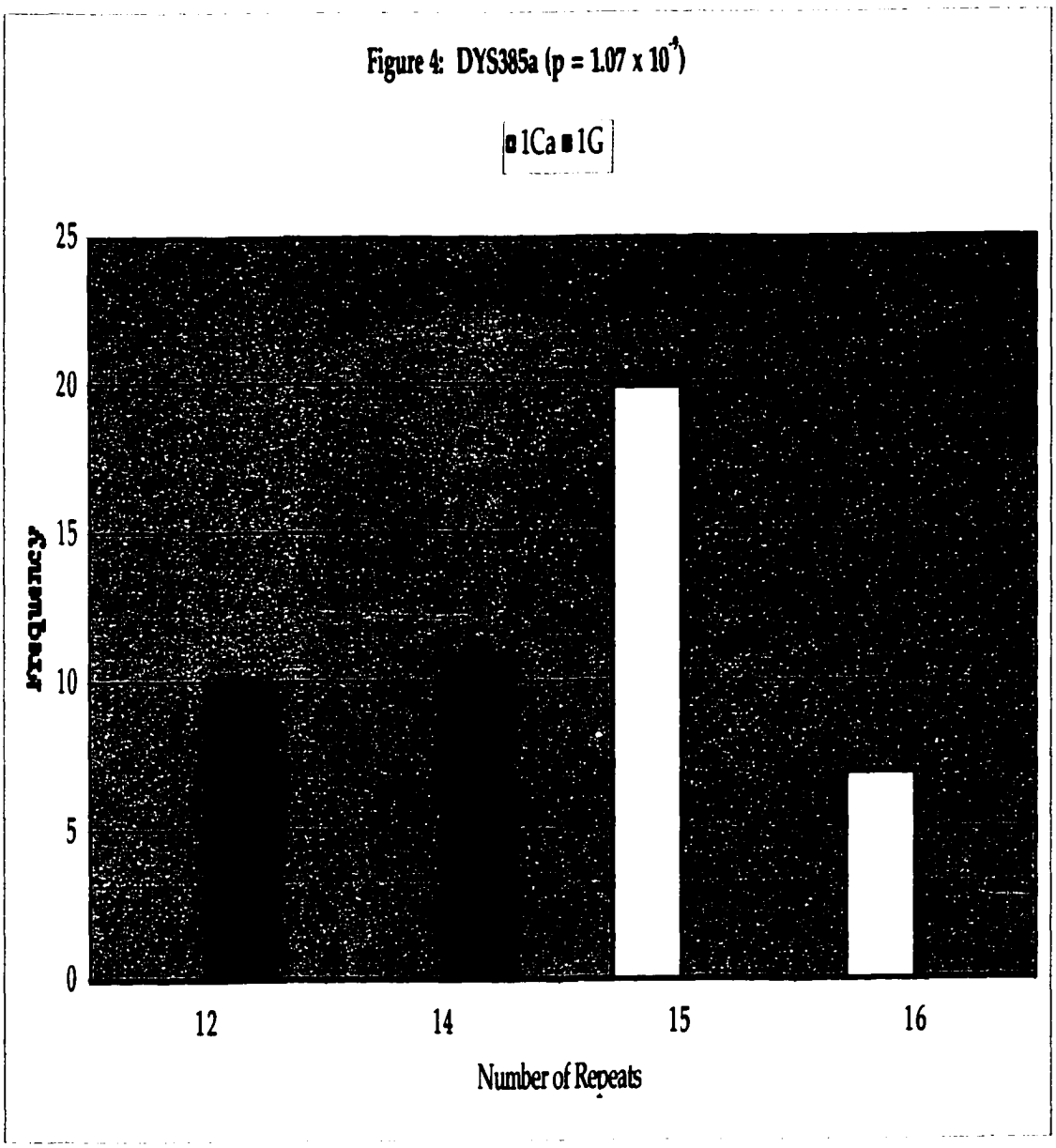
Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



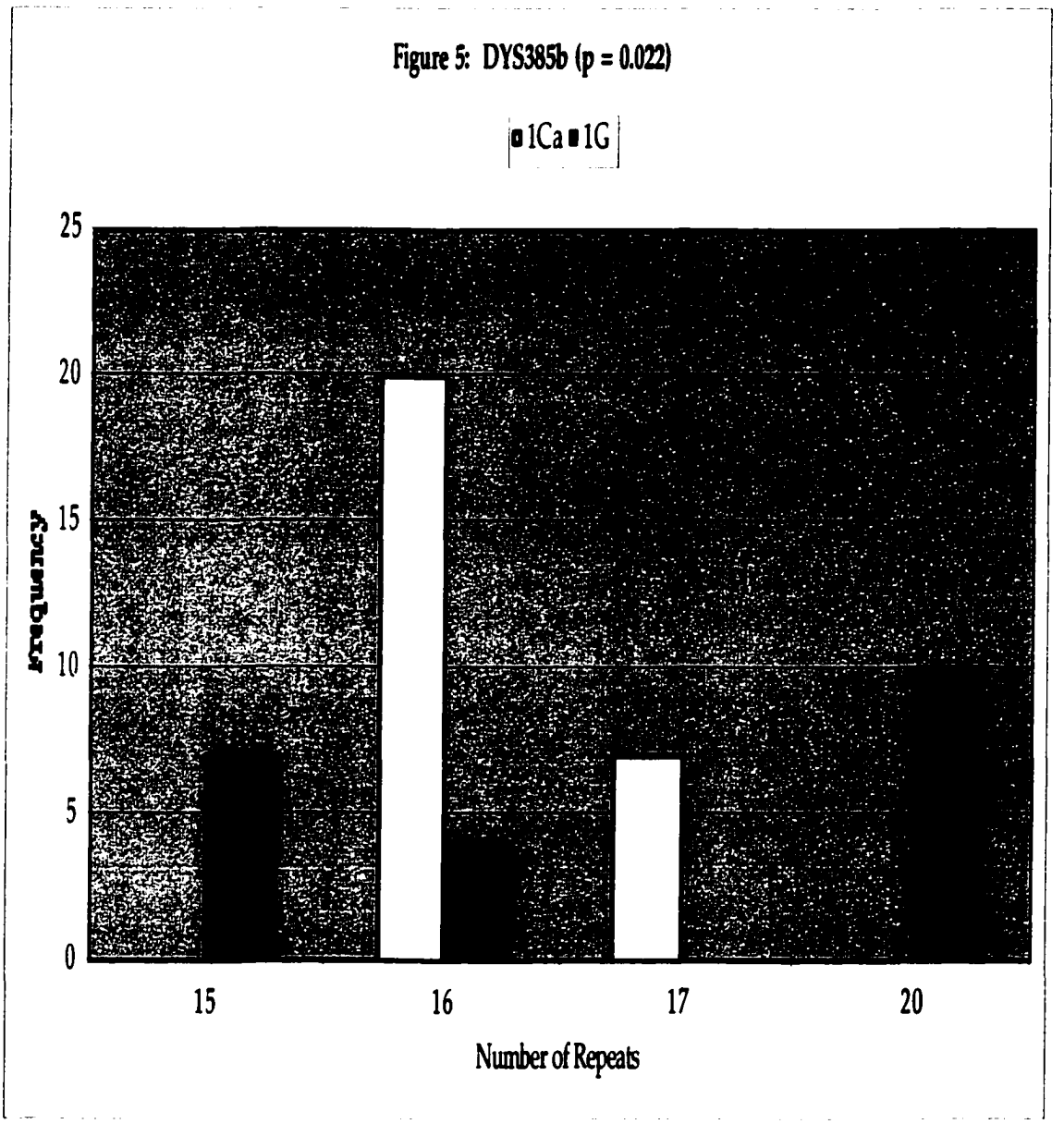
Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



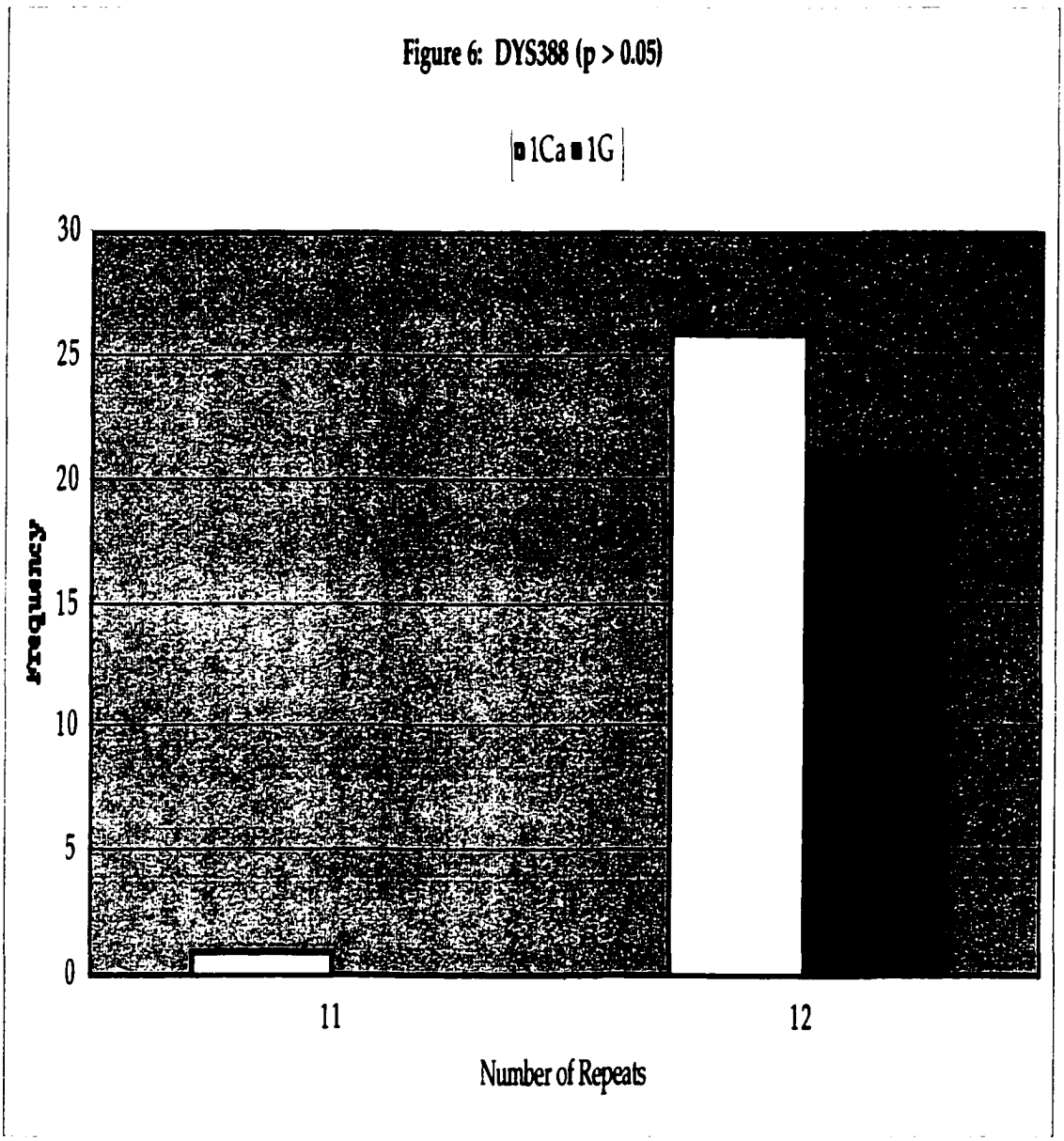
Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



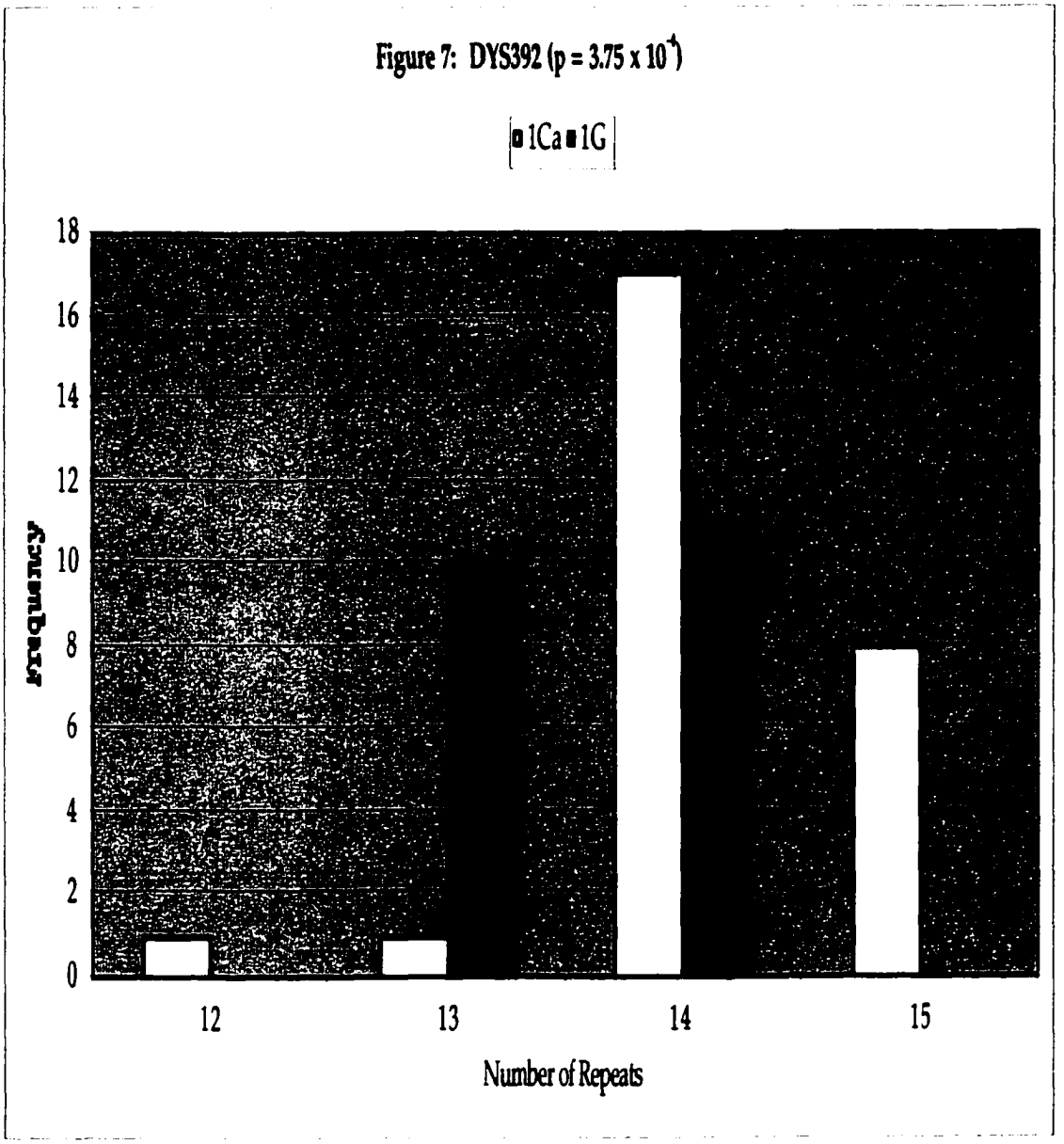
Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



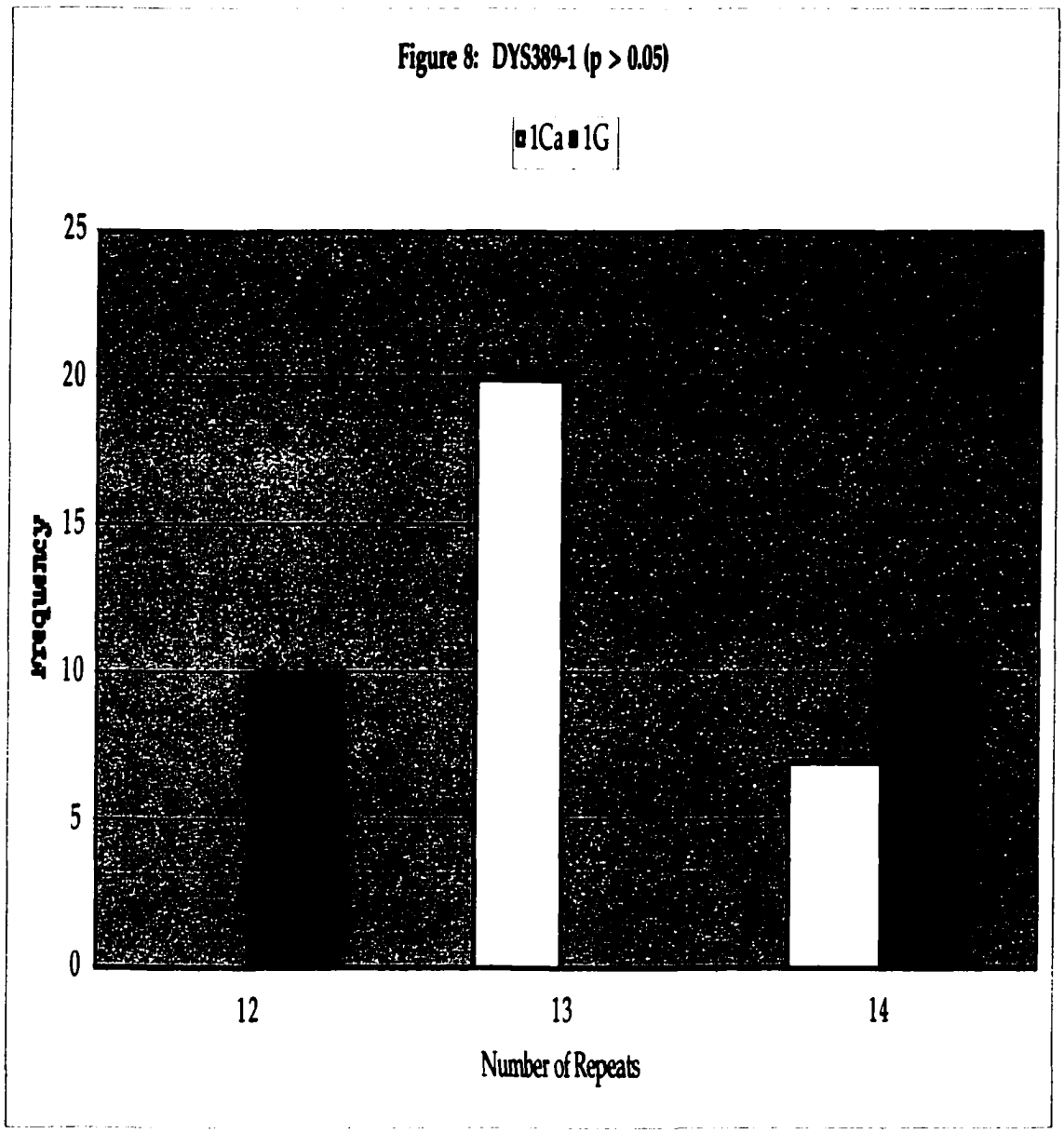
Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



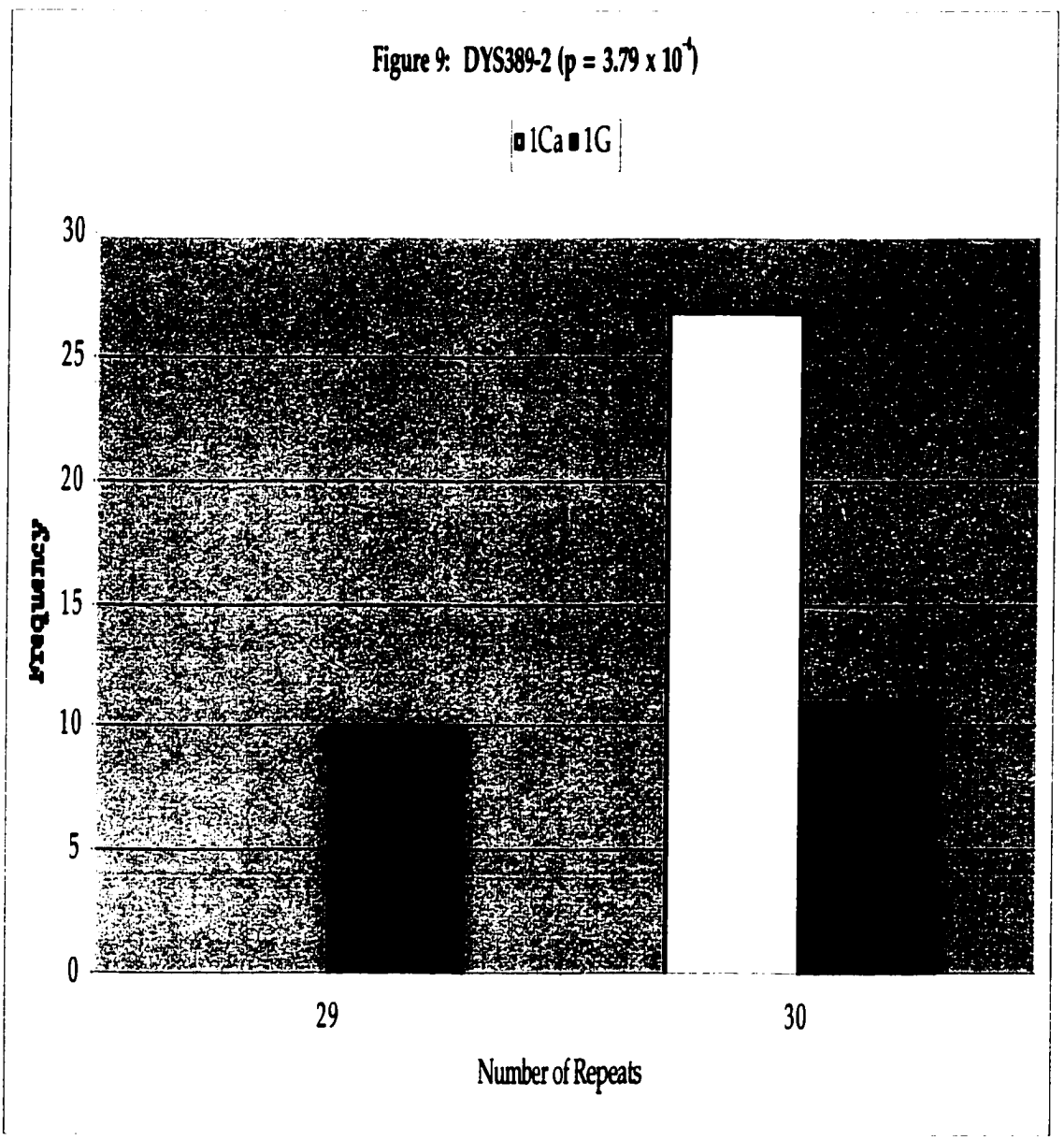
Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



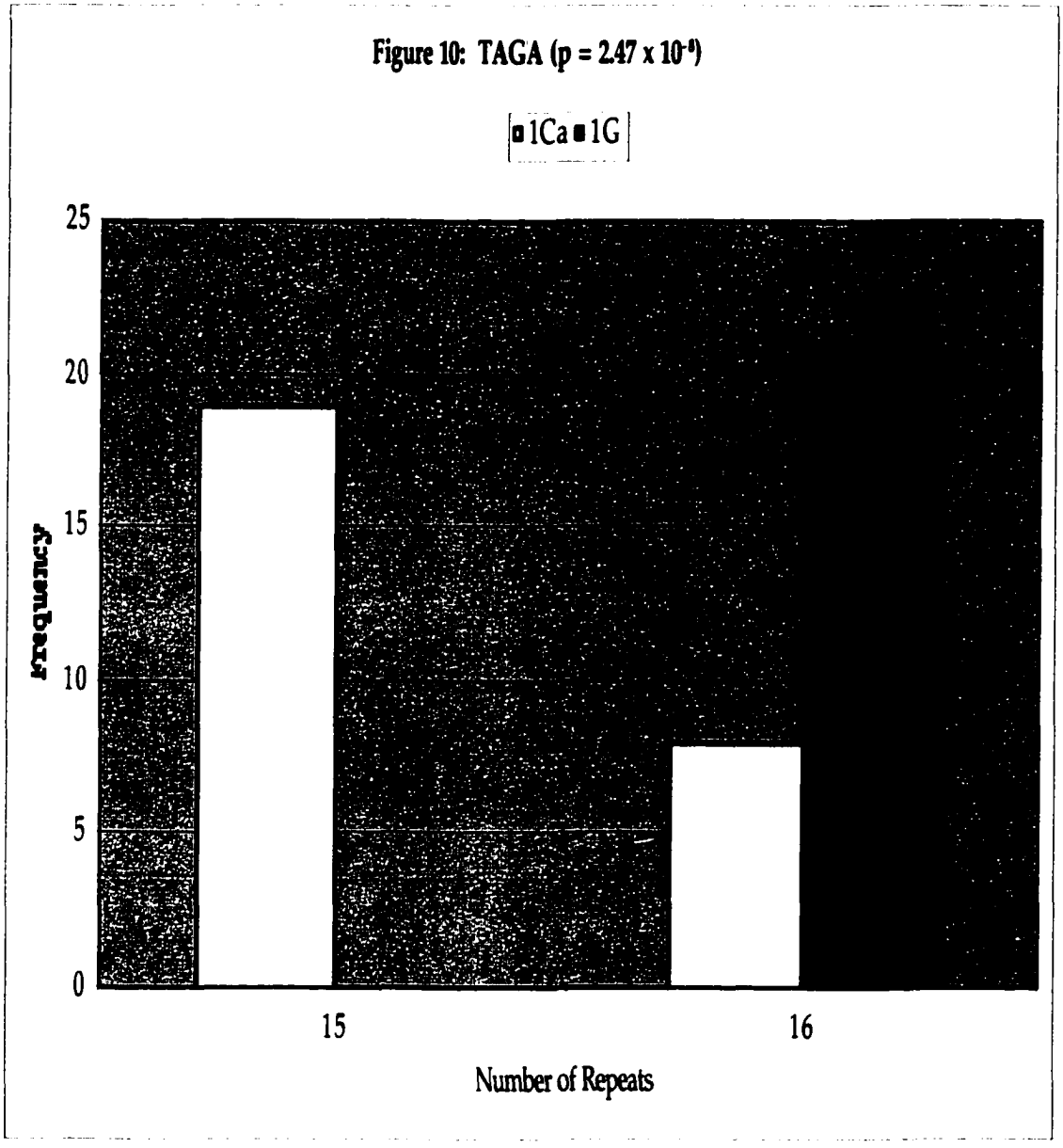
Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Figures 3.1-3.10: Comparison of STR Repeat Number at 10 Polymorphic Loci
Graphical representation of a comparison of the frequency, within each observed Havasupai haplogroup, of the number of repeats observed at 10 polymorphic STR loci on the NRY



Note: Values in parentheses represent t-test results demonstrating the probability that the observed differences in allele frequency between the two NRY haplogroups (1Ca and 1G) occurred by chance.

Table 3.4: t-Tests

Two-sample t-tests assuming unequal variances were used to measure haplogroup-specific distribution of alleles at 10 polymorphic STR loci

DYS390

	ICa	IG	DYS 390	ICa	IG
	23	24	Mean	23.67	24.00
	23	24	Var	0.23	0.00
	23	24	Obs.	27	21
	23	24	U1-U2=	0	
	23	24	df	26	
	23	24	t Stat	-3.61	
	23	24	P(T<=t) one-tail	6.48E-04	
	23	24	t Critical one-tail	1.71	
	23	24	P(T<=t) two-tail	0.001	
	24	24	t Critical two-tail	2.06	
	24	24			
	24	24			
	24	24			
	24	24			
	24	24			
	24	24			
	24	24			
	24	24			
	24	24			
	24	24			
	24	24			
	24				
	24				
	24				
	24				
	24				
	24				
n	27	21			
avg	23.67	24.00			
var	0.23	0.00			

DYS391

	ICa	IG	DYS 391	ICa	IG
	10	9	Mean	10.96	9.48
	11	9	Var	0.04	0.26
	11	9	Obs.	27	21
	11	9	U1-U2=	0	
	11	9	df	24	
	11	9	t Stat	12.64	
	11	9	P(T<=t) one-tail	2.13E-12	
	11	9	t Critical one-tail	1.71	
	11	9	P(T<=t) two-tail	4.26E-12	
	11	9	t Critical two-tail	2.06	
	11	9			
	11	10			
	11	10			
	11	10			
	11	10			
	11	10			
	11	10			
	11	10			
	11	10			
	11	10			
	11	10			
	11				
	11				
	11				
	11				
	11				
	11				
n	27	21			
avg	10.96	9.48			
var	0.04	0.26			

DYS393

	ICa	IG	DYS393	ICa	IG
	14	13	Mean	14.00	13.05
	14	13	Var	0.00	0.05
	14	13	Obs.	27	21
	14	13	U1-U2=	0	
	14	13	df	20	
	14	13	t Stat	20.00	
	14	13	P(T<=t) one-tail	5.40E-15	
	14	13	t Critical one-tail	1.72	
	14	13	P(T<=t) two-tail	1.08E-14	
	14	13	t Critical two-tail	2.09	
	14	13			
	14	13			
	14	13			
	14	13			
	14	13			
	14	13			
	14	13			
	14	13			
	14	13			
	14	13			
	14	14			
	14				
	14				
	14				
	14				
	14				
	14				
n	27	21			
avg	14.00	13.05			
var	0.00	0.05			

DYS385a

	ICa	IG	DYS385a	ICa	IG
	15	12	Mean	15.26	13.05
	15	12	Var	0.20	1.05
	15	12	Obs.	27	21
	15	12	U1-U2=	0	
	15	12	df	26	
	15	12	t Stat	9.24	
	15	12	P(T<=t) one-tail	5.34E-10	
	15	12	t Critical one-tail	1.71	
	15	12	P(T<=t) two-tail	1.07E-09	
	15	12	t Critical two-tail	2.06	
	15	14			
	15	14			
	15	14			
	15	14			
	15	14			
	15	14			
	15	14			
	15	14			
	15	14			
	15	14			
	16	14			
	16				
	16				
	16				
	16				
	16				
	16				
n	27	21			
avg	15.26	13.05			
var	0.20	1.05			

DYS385b

	ICa	IG	DYS385b	ICa	IG
	16	15	Mean	16.26	17.57
	16	15	Var	0.20	5.76
	16	15	Obs.	27	21
	16	15	U1-U2=	0	
	16	15	df	21	
	16	15	t Stat	-2.47	
	16	15	P(T<=t) one-tail	0.011	
	16	16	t Critical one-tail	1.72	
	16	16	P(T<=t) two-tail	0.022	
	16	16	t Critical two-tail	2.08	
	16	16			
	16	20			
	16	20			
	16	20			
	16	20			
	16	20			
	16	20			
	16	20			
	16	20			
	16	20			
	17	20			
	17				
	17				
	17				
	17				
	17				
	17				
n	27	21			
avg	16.26	17.57			
var	0.20	5.76			

DYS388

	ICa	IG	DYS388	ICa	IG
	11	12	Mean	11.96	12.00
	12	12	Var	0.04	0.00
	12	12	Obs.	27	21
	12	12	U1-U2=	0	
	12	12	df	26	
	12	12	t Stat	-1.00	
	12	12	P(T<=t) one-tail	0.163	
	12	12	t Critical one-tail	1.71	
	12	12	P(T<=t) two-tail	0.327	
	12	12	t Critical two-tail	2.06	
	12	12			
	12	12			
	12	12			
	12	12			
	12	12			
	12	12			
	12	12			
	12	12			
	12	12			
	12	12			
	12	12			
	12				
	12				
	12				
	12				
	12				
	12				
n	27	21			
avg	11.96	12.00			
var	0.04	0.00			

DYS392

	ICa	IG	DYS392	ICa	IG
	12	13	Mean	14.19	13.52
	13	13	Var	0.46	0.26
	14	13	Obs.	27	21
	14	13	U1-U2=	0	
	14	13	df	46	
	14	13	t Stat	3.84	
	14	13	P(T<=t) one-tail	1.87E-04	
	14	13	t Critical one-tail	1.68	
	14	13	P(T<=t) two-tail	3.75E-04	
	14	13	t Critical two-tail	2.01	
	14	14			
	14	14			
	14	14			
	14	14			
	14	14			
	14	14			
	14	14			
	14	14			
	14	14			
	15	14			
	15	14			
	15				
	15				
	15				
	15				
	15				
	15				
n	27	21			
avg	14.19	13.52			
var	0.46	0.26			

DYS389-I

	ICa	IG	DYS 389-I	ICa	IG
	13	12	Mean	13.26	13.05
	13	12	Var	0.20	1.05
	13	12	Obs.	27	21
	13	12	U1-U2=	0	
	13	12	df	26	
	13	12	t Stat	0.88	
	13	12	P(T<=t) one-tail	0.192	
	13	12	t Critical one-tail	1.71	
	13	12	P(T<=t) two-tail	0.385	
	13	12	t Critical two-tail	2.06	
	13	14			
	13	14			
	13	14			
	13	14			
	13	14			
	13	14			
	13	14			
	13	14			
	13	14			
	13	14			
	14	14			
	14				
	14				
	14				
	14				
	14				
	14				
n	27	21			
avg	13.26	13.05			
var	0.20	1.05			

DYS389-II

	ICa	IG	DYS 389-I	ICa	IG
	30	29	Mean	30.00	29.52
	30	29	Var	0.00	0.26
	30	29	Obs.	27	21
	30	29	U1-U2=	0	
	30	29	df	20	
	30	29	t Stat	4.26	
	30	29	P(T<=t) one-tail	1.90E-04	
	30	29	t Critical one-tail	1.72	
	30	29	P(T<=t) two-tail	3.79E-04	
	30	29	t Critical two-tail	2.09	
	30	30			
	30	30			
	30	30			
	30	30			
	30	30			
	30	30			
	30	30			
	30	30			
	30	30			
	30	30			
	30	30			
	30				
	30				
	30				
	30				
	30				
	30				
n	27	21			
avg	30.00	29.52			
var	0.00	0.26			

TAGA

	ICa	IG		ICa	IG
	15	16	Mean	15.30	16.00
	15	16	Var	0.22	0.00
	15	16	Obs.	27	21
	15	16	U1-U2=	0	
	15	16	df	26	
	15	16	t Stat	-7.86	
	15	16	P(T<=t) one-tail	1.24E-08	
	15	16	t Critical one-tail	1.71	
	15	16	P(T<=t) two-tail	2.47E-08	
	15	16	t Critical two-tail	2.06	
	15	16			
	15	16			
	15	16			
	15	16			
	15	16			
	15	16			
	15	16			
	15	16			
	15	16			
	16	16			
	16	16			
	16				
	16				
	16				
	16				
	16				
	16				
n	27	21			
avg	15.30	16.00			
var	0.22	0.00			

Table 3.5: mtDNA and Y-chromosome Diversity

This table shows a comparison of mtDNA and Y-chromosome diversity In the Havasupai. Nucleotide diversity is measured by, k , the average number of pairwise differences.

	Number of				Haplotype Diversity (Nei, 1987)	Nucleotide Diversity (Nei, 1987)
Group	Individuals	Haplotypes	Rare Alleles	Polymorphic Sites	h	k
mtDNA (HVS-I+II)	43	10	6	21	0.81 ± 0.0350	3.74 ± 1.9230
Y-chromosome	48	13	7	10	0.89 ± 0.0185	4.94 ± 2.4466

Chapter 4

ESTIMATION OF mtDNA AND Y-CHROMOSOME MUTATION RATES FROM HAVASUPAI PEDIGREES

INTRODUCTION

A vast number of estimates of mitochondrial mutation rates, using a variety of methods, mutational models, and generation times, have been derived from phylogenetic studies. They range from 0.025-0.26 per site per million years (Vigilant et al., 1991; Hasegawa et al., 1993; Horai et al., 1995) to 1.5-4.95 per site per million generations for HVS-I (Ward et al., 1991; Tamura and Nei, 1993). However, it has been observed that mutation rates derived from phylogenetic studies are generally underestimated and that a more accurate estimate of mutation rate comes from the application of coalescent models (Lundstrom et al., 1992) or the study of known pedigrees (Howell et al., 1996). There is, in fact, a proposed 200-fold disparity between the "evolutionary" and "pedigree" mutation rates for mtDNA (Macaulay et al., 1997 vs. Howell and Mackey). Although reports of mutation rates are much sparser for the Y-chromosome, "evolutionary" mutation rates as low as 0.00026 per site per generation have been reported (Forster et al., 2000) which is, on average, an order of magnitude less than that of 0.0028 per site per generation reported for established pedigrees (Kayser et al., 2000).

Variation in mutation rates can have many causes including different origins and size of data sets, different models of mutation, different generation times and/or divergence times between lineages, recurrent mutations, and rate heterogeneity (Wakeley, 1993; Yang, 1996; Excoffier and Yang, 1999; Heyer et al., 2001). Indeed, it has been proposed that phylogenetic mutation rates, which rely on assumptions of effective

population size and/or population separation time, more accurately represent the fixation rate of mutations rather than the mutation rate itself (Howell et al., 1996).

Thus, accurate determination of mutation rates can be ascertained much more easily through the use of established pedigrees. Rate heterogeneity is irrelevant to estimation of average mutation rate in a population pedigree because counting mutations is unbiased except in the remote chance that a particular locus mutates so rapidly, within a single generation, that it fails to be observed (Sigurdardottir et al., 2000). However, there are some problems with pedigree estimates. Due to the limited number of meioses that are being examined, rapidly mutating sites may be preferentially identified (Heyer et al., 2001). In addition, the small number of meioses being examined results in a significantly larger standard error. Nonetheless, it is generally agreed that the most accurate mutation rate estimates are obtained through the use of pedigrees.

The Havasupai Indians of northern Arizona are unique among populations that have been examined for mutation rates in that they are a historically small population (~600 members at present). In addition, they are both reproductively (80% of Havasupai marriages are endogamous) and geographically (they are located in a reservation at the bottom of the Grand Canyon) isolated. The availability of pedigrees, inclusive of eight generations, for the Havasupai provides a unique opportunity to investigate the rate of mutation, for both mtDNA and the Y-chromosome, in this population.

MATERIALS AND METHODS

Pedigrees

Pedigrees, both maternal and paternal, inclusive of eight generations, and dating back to approximately 1850, were constructed from tribal roles and census data (Markow and Martin, 1993) as part of a long-term investigation involving NIDDM. These pedigrees were then used in determining relatedness among individuals in order to calculate mutation rates.

Molecular analyses of mtDNA polymorphism

Automated DNA sequence analysis of 1127 bp of the mitochondrial control region (CR), as described previously in chapter 1 "mtDNA variation in the Havasupai", was used to genotype the 43 Havasupai individuals in the mtDNA pedigrees.

Molecular analyses of Y-chromosome polymorphism

The 48 Havasupai males in the Y-chromosome pedigree were screened for 3 biallelic markers DYS199, DYS257, and ARS72425 as well as 12 short tandem repeats (STRs) DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS394, DYS388, DYS426, and TAGA13, as detailed previously in chapter 3 "Y-chromosome variation in the Havasupai". Only the STR data were used in the derivation of Y-chromosome mutation rates.

Calculation of mutation rate

Average mutation rates for both mtDNA and Y-chromosome pedigrees were estimated by dividing the number of observed mutations by the product of the number of meioses analyzed and the number of sites analyzed (e.g. 1127 nucleotide loci for mtDNA and 12 STR loci for the Y-chromosome).

RESULTS

mtDNA pedigrees

Maternal pedigrees consisted of 11 maternal lineages spanning 10 mtDNA haplotypes. Analysis of 87 meioses events among the 11 maternal pedigrees revealed one potential mutation in the mtDNA pedigree in figure 4.8. On the assumption that the pedigrees are accurate, the average mutation rate of mtDNA in the Havasupai was calculated to be 1.0204×10^{-5} per site per generation (table 4.2A). In order to facilitate comparison with other investigations, this estimate may also be expressed as 10.2 per site per million generations.

Y-chromosome pedigrees

Paternal pedigrees included 12 paternal lineages and incorporated 13 Y-chromosome haplotypes. Five mutations were identified by examination of 90 independent meioses events among the 12 paternal lineages (table 4.1). Four one-step mutations were observed among three STR loci: DYS385a (figure 4.25), DYS385b (which harbored two mutations, see figure 4.14), and DYS388 (figure 4.25). A two-step

mutation was detected at the DYS392 locus (figure 4.23). Under a stepwise mutation model (Ohta and Kimura, 1973; Chakraborty, 1977), this single two-step mutation could be counted as two separate mutations. However, since we know the genotype of the father and the son and can, thus, prove that the mutation occurred in one generation, the two-step mutation is counted only once. Calculation of mutation rate was performed under both assumptions (table 4.2B and 4.2C) so as to obtain a range of mutation rate that accounts for both hypotheses, however, the majority of the analyses (and text) will refer to this two-step mutation as one event. The average Y-chromosome mutation rate in the Havasupai was found to be 0.0046 per site per generation (table 4.2B). However, if we assume a stepwise mutation model (Ohta and Kimura, 1973; Chakraborty, 1977) and count the two-step mutation at the DYS392 locus as two mutations, then the average mutation rate for the Havasupai Y-chromosome increases to 0.0056 per site per generation (table 4.2C). Lastly, 4 of the 5 identified mutations resulted from a loss of repeat number. One mutation, however, one of the two at the DYS385b locus, within pedigree 4.14, showed an increase in the number of repeats (table 4.1).

DISCUSSION

mtDNA pedigree analysis

The maternal lineages determined by DNA sequencing of the mitochondrial CR match precisely, with a few exceptions, the pedigrees constructed from tribal roles and census data (Markow and Martin, 1993). First, it can be observed that mtDNA haplotype

#1 (Hav 1) contains three maternal lineages (figures 4.3-4.5) that can be traced back to three maternal ancestors. Two were born at approximately the same time (1862 and 1867 respectively), thus it is likely they were either sisters or cousins. It is possible that the third, whose date of birth predates the pedigree record, is the mother or aunt of one or both of the other two. Regardless, the fact that they all have the same haplotype would appear to suggest that all three share a common maternal ancestor.

Haplotypes Hav 2 (figure 4.6) and Hav 3 (figure 4.7) present a slightly more difficult issue. There is one individual in the Hav 2 haplotype that falls out in the pedigree represented by the Hav 3 haplotype (figure 4.7). Since there are six nucleotide changes between Hav 2 and Hav 3, mutation can be safely ruled out. Thus, the discrepancy observed in this haplotype must represent an error in the pedigree record, an error in sampling (mis-labeling a tube, etc.), or a complex living arrangement sometime in the past whereby one of the children/grandchildren of one ancestor was raised in the home of another. Unfortunately, there is no way to be certain of the exact cause of the observed incongruity, thus, it was not counted as a mutation and was omitted from further analysis. Haplotypes Hav 2 and Hav 3 are consistent with the pedigree record, with the exception of the one individual discussed above.

The pedigree in figure 4.8 represents the occurrence of two different haplotypes (Hav 4 and Hav 5) that trace back to a common ancestor. Comparison of CR sequences between the two haplotypes reveals only one nucleotide change (a T -> C transition at bp 204 that is present in Hav 5, but not in Hav 4; see figure 4.1) separating the two, suggesting that mutation could be responsible for the divergence of the two haplotypes.

However, since we do not know the genotype of the maternal ancestor of either haplotype, or their common ancestor, it is impossible to tell which haplotype, if either, is the one carried by the rest of the lineage or where the potential mutation might have occurred. However, we can infer, with relative safety, that the rest of the lineage carries one of the two haplotypes and that the other is the result of mutation. Thus, this assumption was made and the mtDNA mutation rate was calculated based on this one mutation. The remaining Havasupai mtDNA haplotypes Hav 6-10 are consistent with their respective pedigrees and, thus, need no further interpretation.

mtDNA mutation rate

A total of 89 female germline transmission events were examined among 11 maternal pedigrees consisting of 10 mtDNA haplotypes. When the lineage most likely to contain some type of error in the mtDNA pedigree (as discussed previously; and see figure 4.7) was omitted, one lineage consisting of two meiotic events was excluded from the analysis, leaving a total of 87 maternal meioses examined. Among these 87 meioses, one potential mutation was identified (figure 4.8). On the assumption that the pedigrees are accurate, the average mutation rate of mtDNA in the Havasupai was estimated to be 1.0204×10^{-5} per site per generation (table 4.2A). This translates to 10.2 per site per million generations. Our estimated mutation rate is consistent with recent investigations, which have established mutation rates by direct examination of pedigrees.

Most recently, Heyer et al. (2001), using deep-rooted pedigrees from northeastern Quebec, demonstrated an average mutation rate of 11.6 per site per million generations.

Assuming a generation time of 30 years, an Icelandic pedigree was shown to have a mutation rate of 6.4 per site per million generations (Sigurdardottir et al., 2000). However, Howell et al. (1996) and Parsons et al. (1997), investigating an Australian pedigree and a combination of pedigrees, respectively, have demonstrated mutation rates ranging from 22.0 to 45.1 per site per million generations. Standardizing the methodology and assuming a 30 year generation time, an average mutation rate for all known mitochondrial pedigrees (18 mutations in 1,729 meioses) was estimated by Heyer et al. (2001) to be 15.5 per site per million generations (Howell et al., 1996; Parsons et al., 1997; Soodyall et al., 1997; Jazin et al., 1998; Sigurdardottir et al., 2000; Heyer et al., 2001). The estimated mitochondrial mutation rate in the Havasupai is, therefore, consistent with this newly established and standardized average mutation rate for mtDNA.

Y-chromosome pedigrees

The 13 haplotypes determined by Y-chromosome genotyping, as might be expected with the potential for nonpaternity, show a much higher level of discordance with the paternal pedigrees than do the mtDNA haplotypes with the maternal pedigrees. The pedigree in figure 4.14 contains three haplotypes (1, 2, and 13) all tracing back to the same common paternal ancestor. Haplotype 13 can be easily characterized as a clear case of nonpaternity. Not only does it fall into another haplogroup (1Ca for Hap 13 compared with 1G for Hap 1 and 2), at least six mutations, half of them two-step mutations, would be required for Hap 1 or 2 to diverge into Hap 13. Discerning the origins of Hap 1 and

Hap 2 is more complicated, however, as there appears to be haplotype switching within lineages. Haplotype 1 and 2 differ by a single one-step mutation at the DYS385b locus, thus, mutation (i.e. addition/deletion of a single tetranucleotide repeat) is a possible explanation for a single origin for the two haplotypes. If we assume mutation, then a one-step mutation would have had to occurred in both sons of one individual and a reversal of that mutation would have had to occur in only one (of three) of that individual's grandsons. Though this seems unlikely initially, it should be noted that the father of that individual was 43 years old at the time of his birth compared to 29 and 36 years old at the birth of his brothers, and the transmission of germline mutations is known to increase with paternal age (Crow, 1997; Crow, 1999; Crow, 2000; Malispina et al., 2001). Alternatively, the observed switching of haplotypes could suggest unusual practices, whereby one individual could be the son or cousin of his grandfather. Again, however, there is no way to know for certain how this discordance between the pedigrees and the genetic data is rectified.

There are four paternal lineages characterized as haplotype 3 and all but one are consistent with a single common paternal ancestor that dates back prior to our genealogical evidence. Genetic evidence demonstrates that an individual within the pedigree in 4.18 belongs to haplotype 4, whereas his son belongs to haplotype 3. There are several plausible interpretations of this observation including the potential for mutation. However, this individual revealed marked errors in the mtDNA pedigree record and it is, thus, likely that, as mentioned previously, there was some potential mishandling of this sample. As this individual (and their respective lineage) was omitted

from the analysis of mtDNA mutation rates, it was likewise ignored and omitted from further analysis of Y-chromosome mutation rates.

Haplotype 6 is comprised of two paternal lineages, one of which, contains a single individual belonging to another haplotype (haplotype 7). As five mutations would be required to consolidate haplotype 6 and 7, nonpaternity is the obvious rationale for the exclusion of this individual from haplotype 6. Haplotype 8 presents a similar problem. The genetic data suggests that one individual in this pedigree falls out in haplotype 9. In this case, however, only one mutation, albeit a two-step mutation, at the DYS392 locus, separates haplotype 8 from haplotype 9. Thus, although haplotype divergence resulting from a two-step mutation is slightly more rare, a mutation in could still account for this discrepancy in the pedigree record. Alternatively, nonpaternity could be responsible for the incongruence.

The pedigree in figure 4.25 encompasses three haplotypes (11, 12, and 13), of which haplotype 13 is the most numerous. There is only one individual in haplotype 12, which is separated from haplotype 13 by one one-step mutation at the DYS388 locus. This being the case, mutation is the likely rationale for the genetic separation, although nonpaternity is not inconsistent with this observation either. The same is true for haplotype 11. It is separated from haplotype 13 by a single one-step mutation at the DYS385a locus. As before, this incongruence could be the result of mutation or nonpaternity, although the pattern of polymorphism (i.e. single one-step mutations separating haplotypes) leans toward mutation as the probable explanation.

In regards to nonpaternity, there are two confirmed cases (figures 4.14 and 4.21) and one potential case (figure 4.18) identified among the 48 males examined in the Havasupai. Assuming that all of the potential mutations are indeed that, we can, thus, estimate the rate of nonpaternity in the Havasupai at 4-6%. If all observed polymorphism is assumed to be nonpaternity, rather than mutation, then the exclusion rate rises to 17%.

Y-chromosome mutation rate

STR mutation rate is a function of both the mean number of repeats and the motif length (Forster et al., 2000). Thus, pentanucleotide repeats mutate less frequently than tetranucleotide repeats which mutate less frequently than trinucleotide repeats. A total of 99 male germline meioses were examined at 12 microsatellite loci. Two lineages, one consisting of three meioses events and one consisting of two, were omitted due to suspected nonpaternity (i.e. multiple one or two-step mutations at several different loci separate the haplotypes of father and son). Additionally, one lineage of four meiotic events was excluded due to the likelihood of errors in the pedigree record (figure 4.18), leaving a total of 90 germline transmissions analyzed. Among the 90 meioses, 5 potential mutations were observed and a mutation rate of 0.0046 per site per generation was obtained (table 4.2B). In addition, if we accept the stepwise mutation model (Ohta and Kimura, 1973; Chakraborty, 1977) and count the two-step mutation at DYS392 as two mutations, then the average mutation rate increases to 0.0056 per site per generation (table 4.2C).

These rates are two- to four-times higher than those reported for the Y-chromosome in previous pedigree studies, which range from an average of ~0.0012 (Bianchi et al., 1998 averaged their own data with that of Heyer et al., 1997 and Kayser et al., 1997) to 0.0028 per site per generation (Kayser et al., 2000). Additionally, they are an order of magnitude greater than those reported for "evolutionary" mutation rates, estimated by phylogenetics, which average ~0.00026 per site per generation (Forster et al., 2000), as well as those from autosomal microsatellites, which range from 0.0006-0.0027 per site per generation (Brinkmann et al., 1998; Henke and Henke, 1999; Sajantila et al., 1999). It should be noted, however, that some of the phylogenetic estimates including those of Forster et al. (2000) excluded rapidly mutating STRs including DYS389 and DYS392, which would result in lower mutation rates.

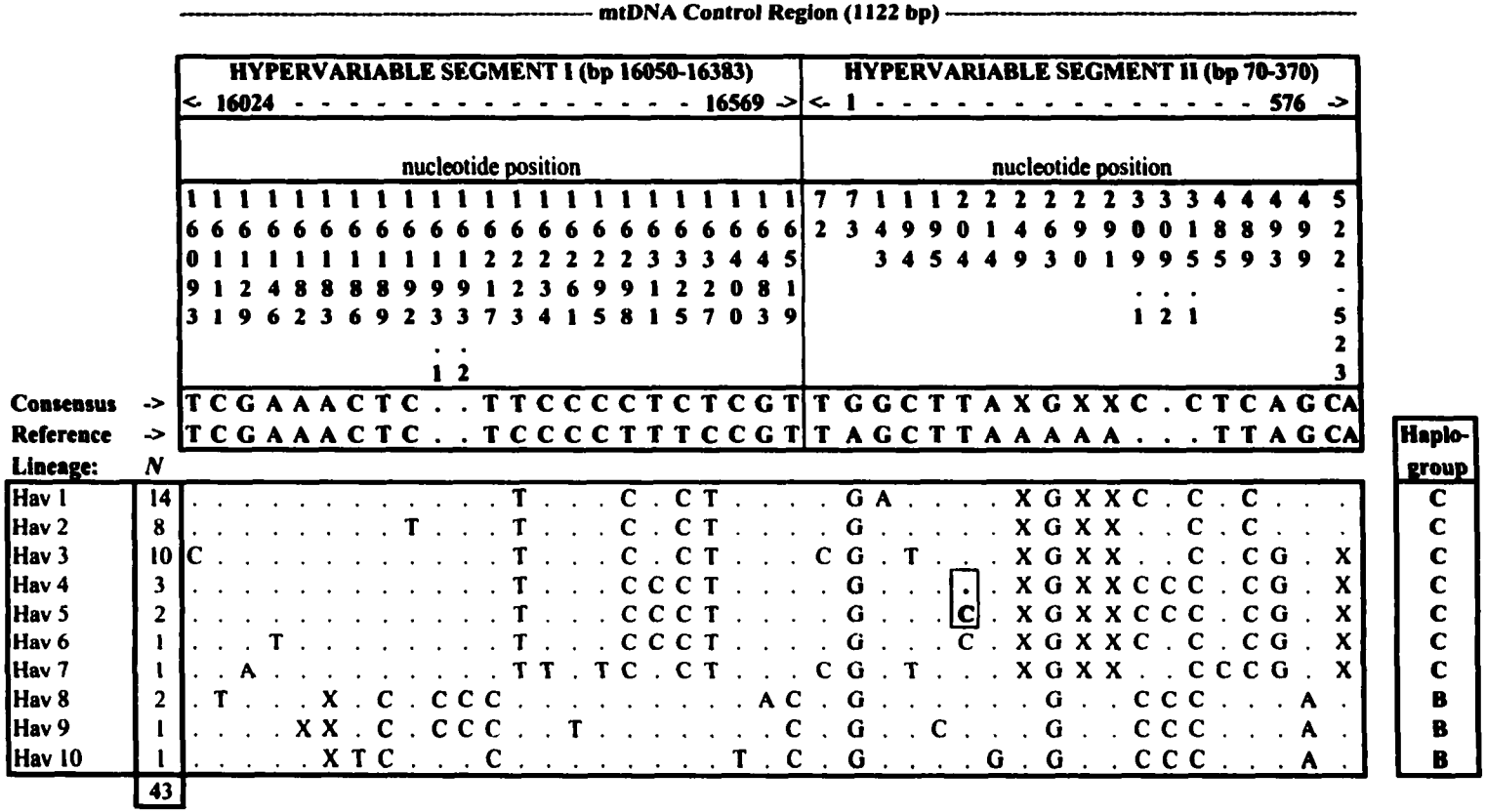
Despite the accuracy of mutation rates ascertained from pedigree studies such as this one, the question of nonpaternity versus mutation cannot be resolved completely here, as the only way to calculate a "true" mutation rate is through observation of father/son pairs with confirmed paternity. Three of the five mutations identified in this investigation occur in father/son pairs, where both father and son have been sampled (table 4.1). On the contrary, in two of the potential mutations (DYS385a and DYS388), the genotype of the father was inferred from the pedigree. Elimination of these two mutations yields a mutation rate of 0.0028 per site per generation, which is identical to the estimate reported by Kayser et al. (2000). However, it remains imperative to test paternity in those father/son pairs for which we have adequate samples in order to verify the accuracy of our derived Y-chromosome mutation rates. Although it is possible that

the mutation rate in the Havasupai has been artificially increased due to a reported increase in mutation rates in lymphoblastoid cell lines (Banchs et al., 1994), it is not known whether or not the individual samples used in this study were immortalized as cells and then extracted as DNA or extracted as DNA directly from the blood samples obtained from volunteers. Additionally, another investigation reports no increase in mutation among samples derived from cell lines (Bianchi et al., 1998), thus, there is no clear indication that cell lines actually accumulate mutations more rapidly.

Although a mutational bias in favor of repeat expansion has been previously demonstrated in German and Polish populations (Cooper et al., 1999; Kayser et al., 2000), the Havasupai show a distinct preference for loss of repeats (table 4.1). Of the five observed mutations, four were found to involve repeat loss. However, this observation was not supported statistically by the chi-square test ($p = 0.1$). Our data support the findings of Kayser et al. (2000) in that the premutation allele size of all mutations observed in the Havasupai was greater than 11 repeats. In fact, our data reveal a minimum premutation repeat size of 12 and a postmutation repeat size of 11 repeats despite a propensity for repeat reduction.

In summary, the Havasupai population shows a pattern of Y-chromosome mutation that is slightly different from that observed previously. The Havasupai display a mutation rate that is two- to four-times higher than those reported in previous pedigree studies. In addition, and contrary to preceding investigations, they display a clear bias in favor of repeat loss at STR loci. Future investigations into Y-chromosome mutation in

the Havasupai should include paternity testing, when possible, to verify the accuracy of the proposed mutation rates.



Reference sequence is that of Anderson et al., 1981.
Consensus sequence is that of the majority of Havasupai without comparison to reference sequence.
An X in any sequence denotes a deletion.
☐ indicates a potential mutational change within a pedigree

Figure 4.1: Polymorphic Nucleotide Positions in mtDNA
Polymorphic nucleotide positions within the mtDNA control region of Havasupai Indians

Figure 4.2: Polymorphic Y-chromosome Loci

			Single Nucleotide Polymorphisms (SNPs)				Microsatellite Short Tandem Repeats (STRs)											
Haplo-type	n	Freq.	SY183/ DYS199	SY57/ DYS257	ARS 72425	Haplo-group	DYS390	DYS391	DYS393	DYS394	DYS385a	DYS385b	DYS388	DYS392	DYS426	DYS389-1	DYS389-2	TAGA
1	7	0.15	+	+	+	1G	24	9	13	13	14	20	12	14	12	14	30	16
2	4	0.08	+	+	+	1G	24	9	13	13	14	21	12	14	12	14	30	16
3	9	0.19	+	+	+	1G	24	10	13	13	12	20	12	13	12	12	29	16
4	1	0.02	+	+	+	1G	24	10	14	13	12	20	12	13	12	12	29	16
5	2	0.04	-	+	+	1Ca	23	11	14	13	15	16	12	14	12	13	30	15
6	7	0.15	-	+	+	1Ca	23	11	14	13	15	16	12	14	12	13	30	16
7	1	0.02	-	+	+	1Ca	24	10	14	13	16	16	12	15	12	13	30	15
8	8	0.17	-	+	+	1Ca	24	11	14	13	15	16	12	14	12	13	30	15
9	1	0.02	-	+	+	1Ca	24	11	14	13	15	16	12	12	12	13	30	15
10	1	0.02	-	+	+	1Ca	24	11	14	13	15	16	12	13	12	13	30	15
11	1	0.02	-	+	+	1Ca	24	11	14	13	15	17	12	15	12	14	30	15
12	1	0.02	-	+	+	1Ca	24	11	14	13	16	17	11	15	12	14	30	16
13	5	0.10	-	+	+	1Ca	24	11	14	13	16	17	12	15	12	14	30	15

This table shows the 13 observed Havasupai Y-chromosome haplotypes, representing two Native American haplogroups (1G and 1Ca), and their respective frequencies within this population.

indicates a potential mutational change within a pedigree

indicates a potential mutational change as well as a reversal of this same change within a pedigree, such "haplotype switching" represents two potential mutations

Figure 4.3: mtDNA haplotype 1a (Hav 1)

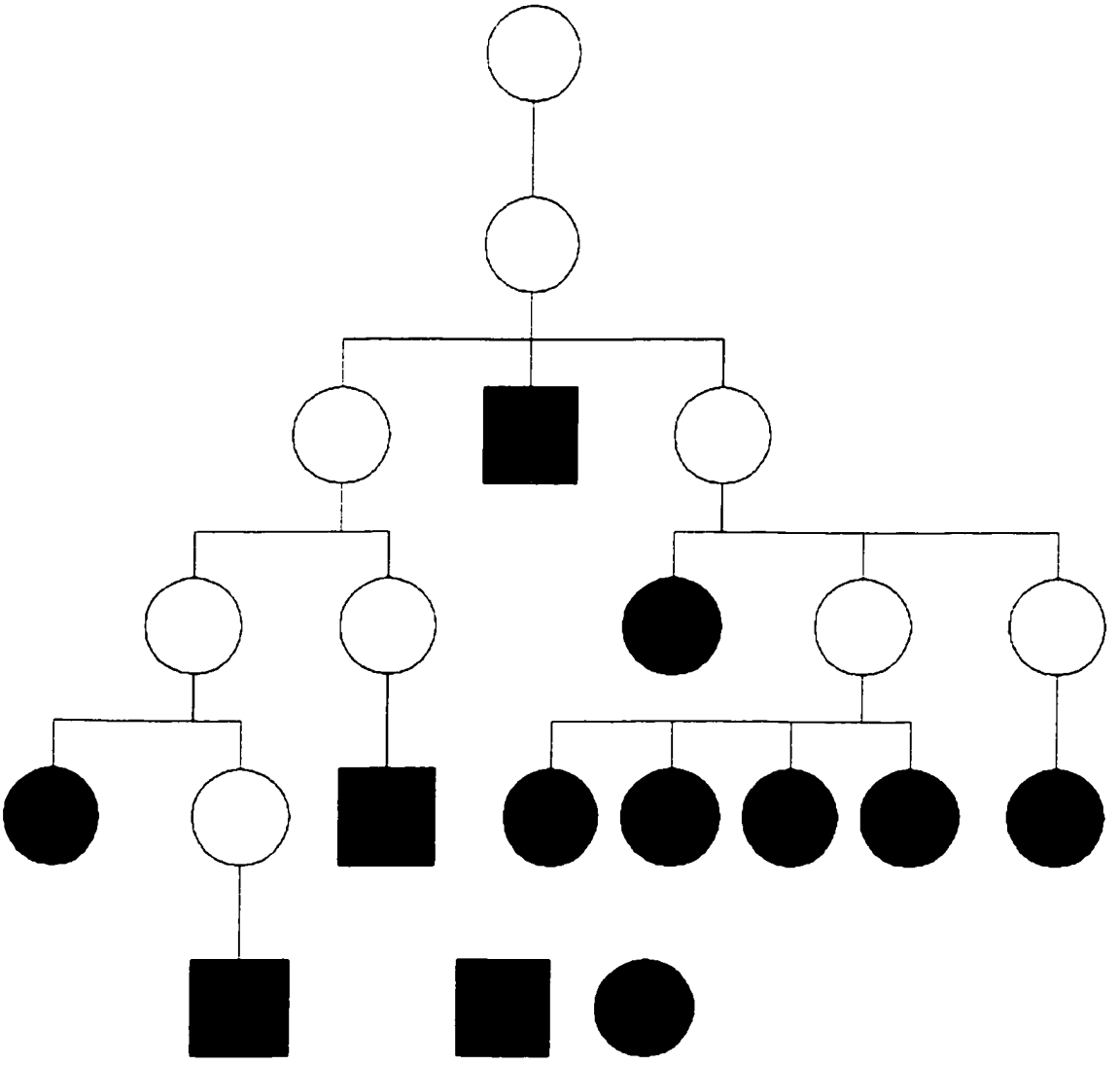


Figure 4.4: mtDNA haplotype 1b (Hav 1)

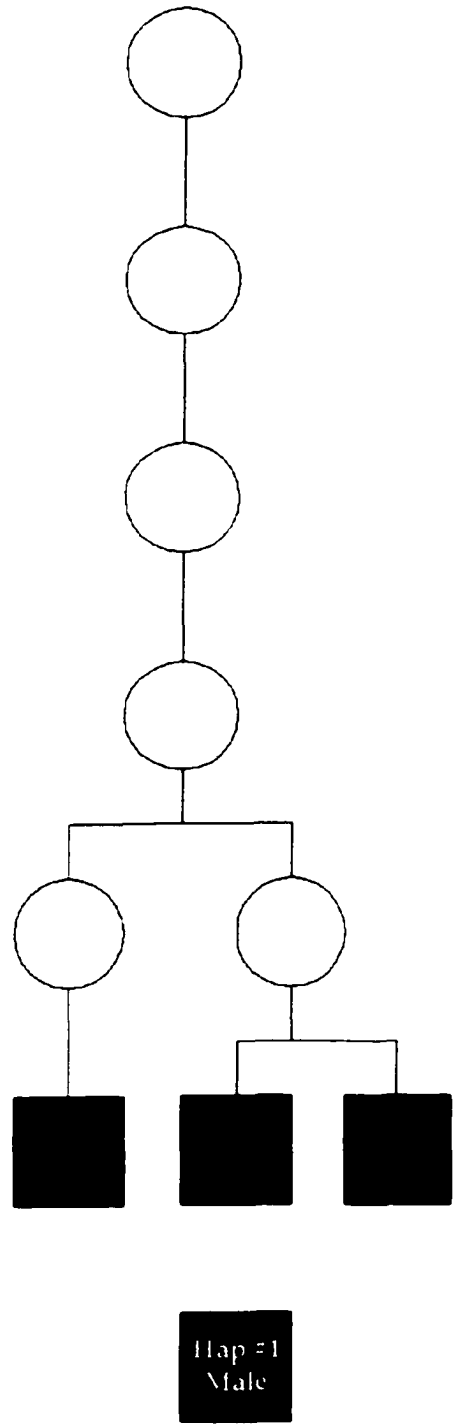


Figure 4.5: mtDNA haplotype 1c (Hav 1)

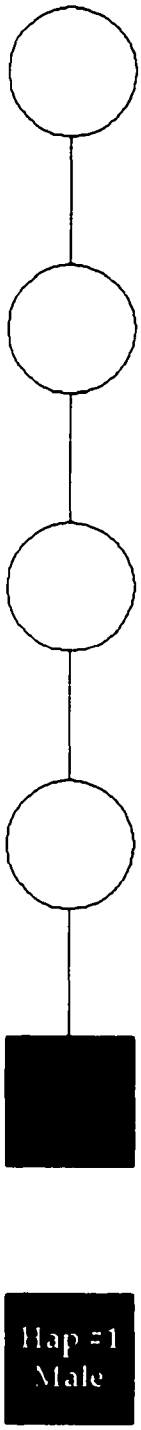


Figure 4.6: mtDNA haplotype 2 (Hav 2)

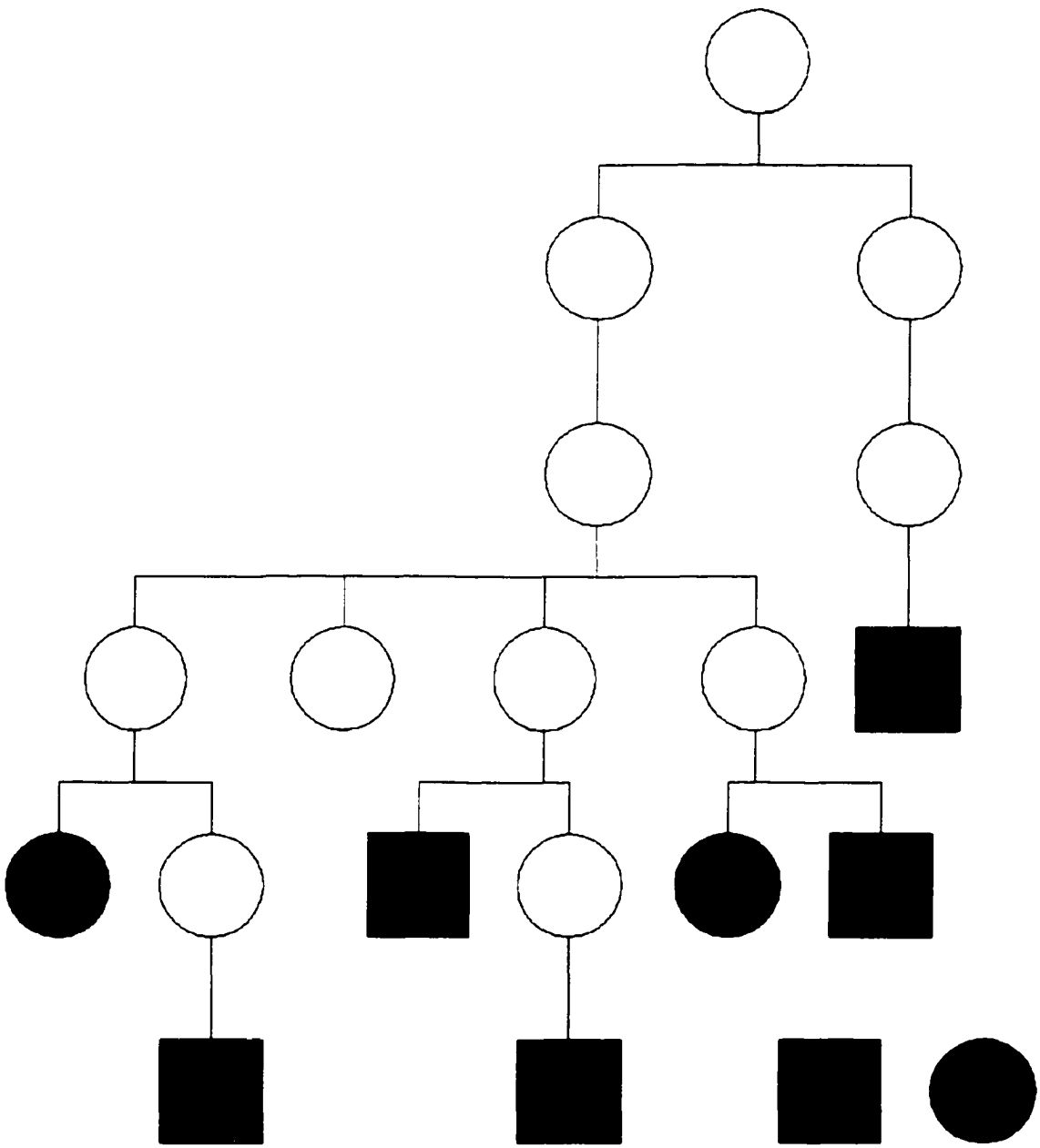


Figure 4.7: mtDNA haplotype 3 (Hav 3)

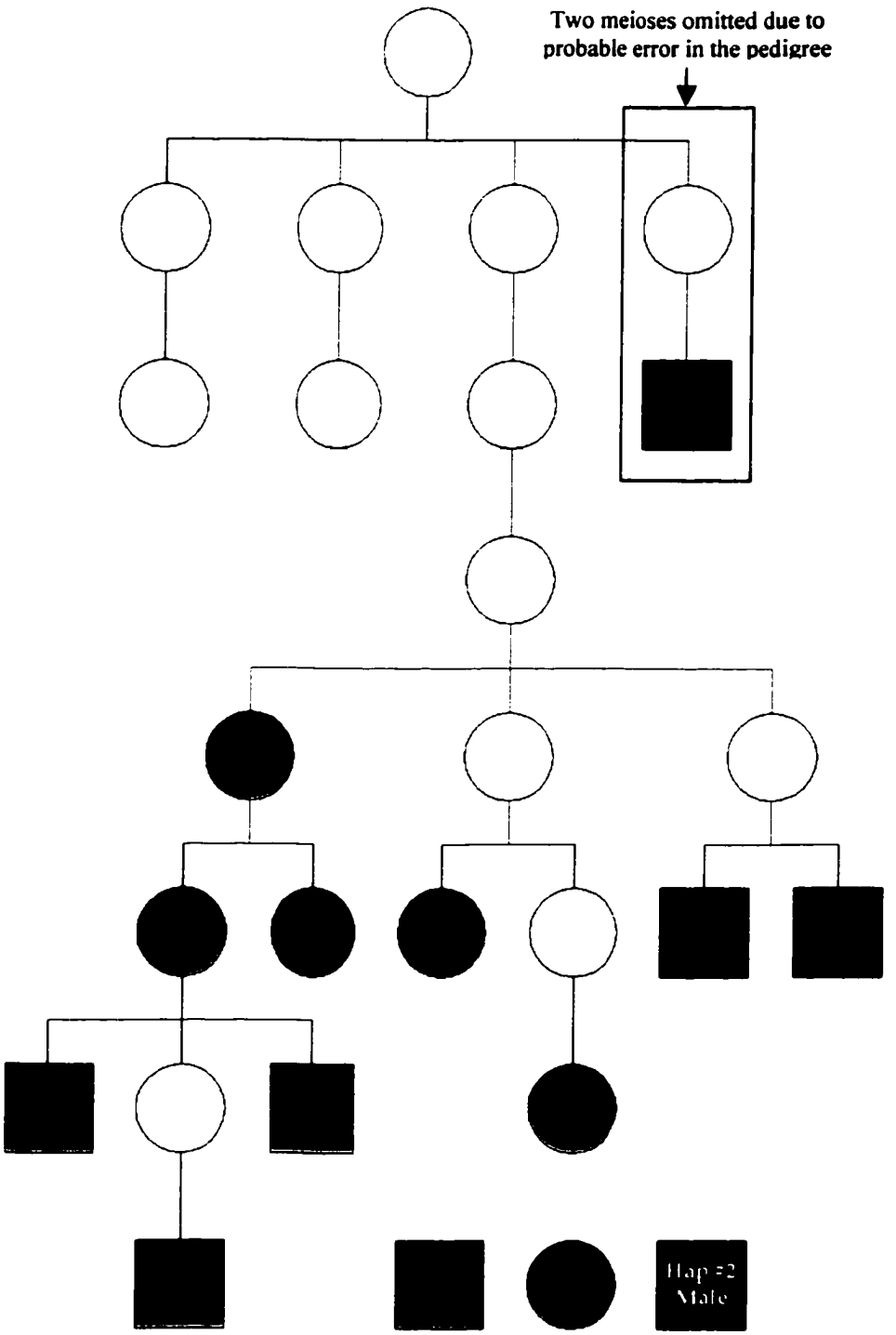


Figure 4.8: mtDNA haplotypes 4 and 5 (Hav 4 and Hav 5)

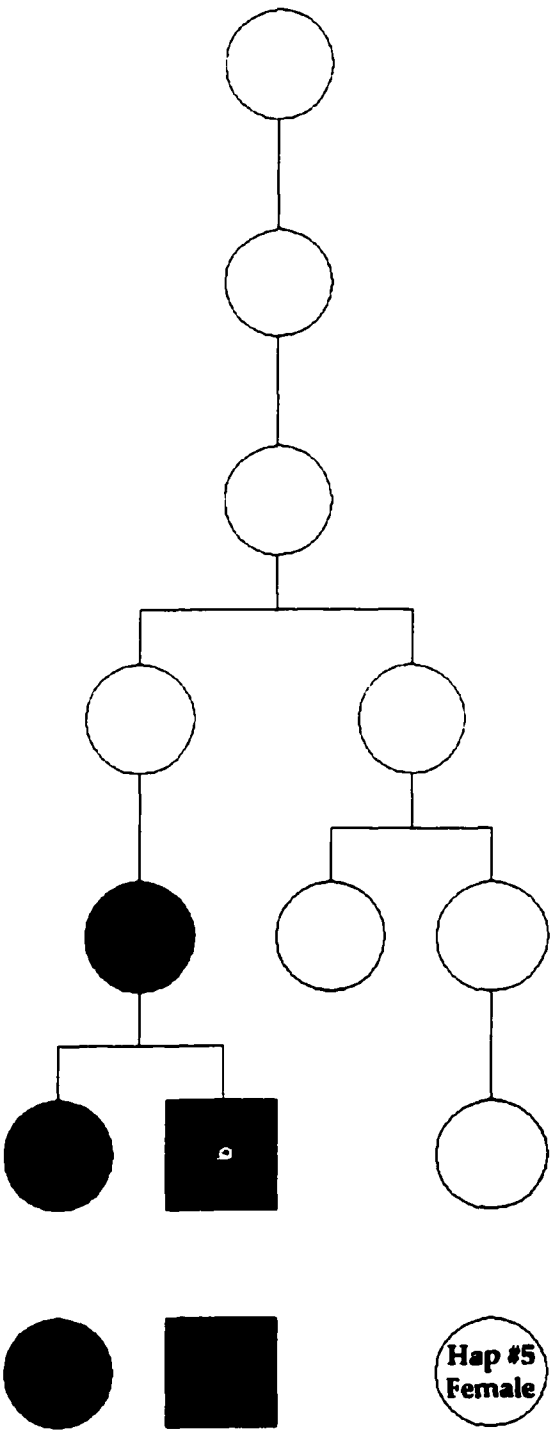


Figure 4.9: mtDNA haplotype 6 (Hav 6)

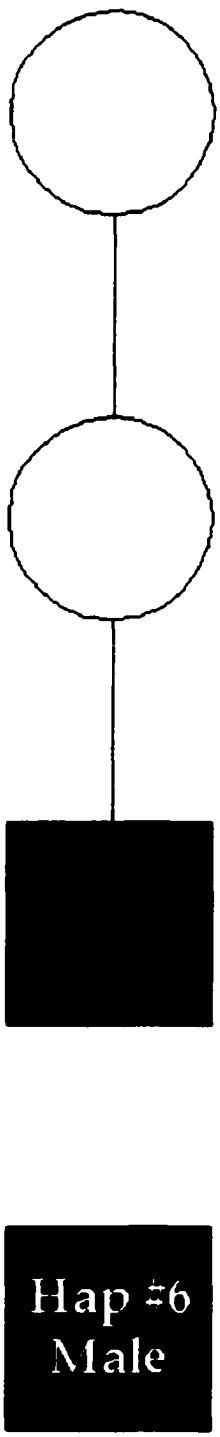


Figure 4.10: mtDNA haplotype 7 (Hav 7)

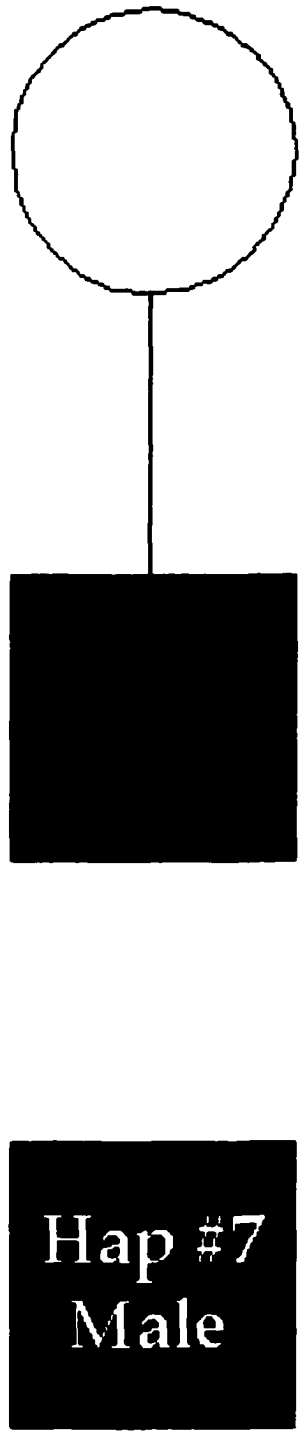


Figure 4.11: mtDNA haplotype 8 (Hav 8)

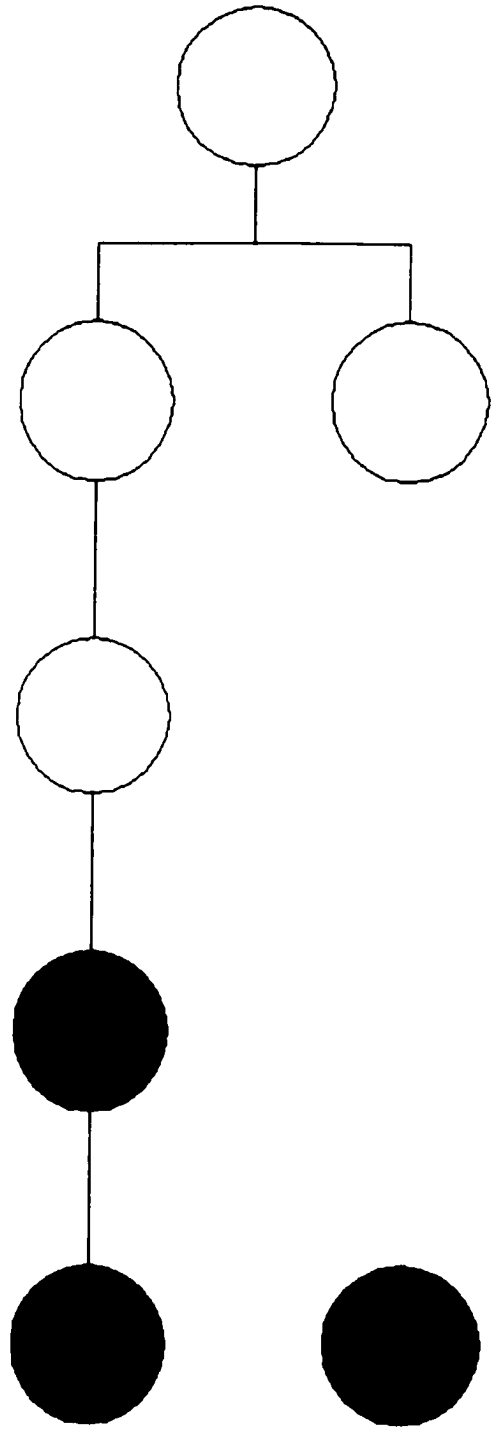


Figure 4.12: mtDNA haplotype 9 (Hav 9)

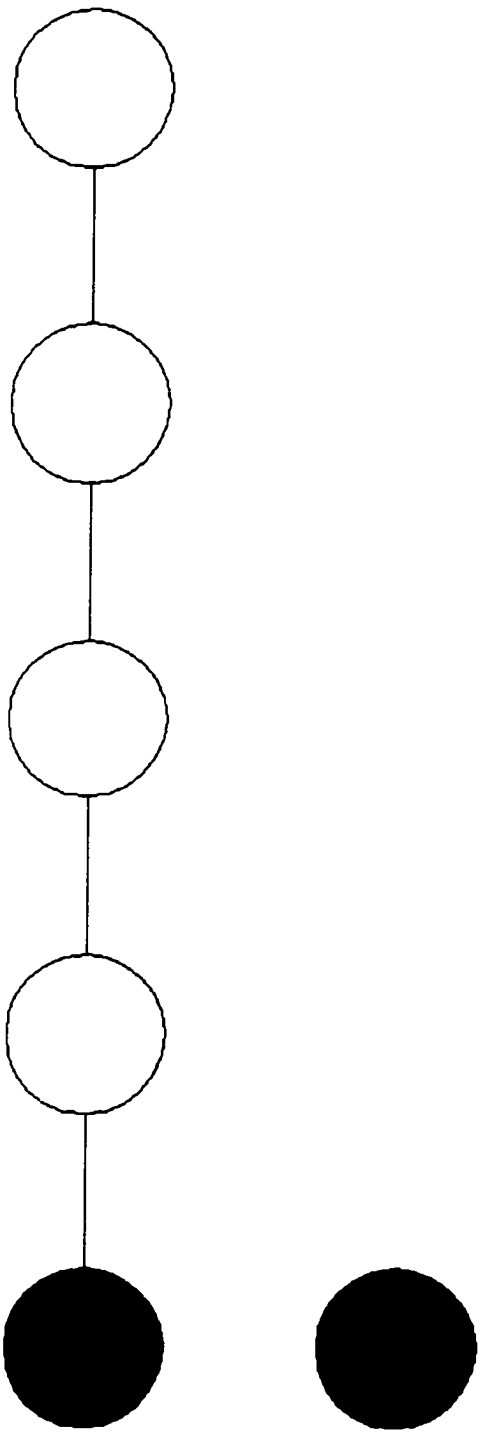


Figure 4.13: mtDNA haplotype 10 (Hav 10)

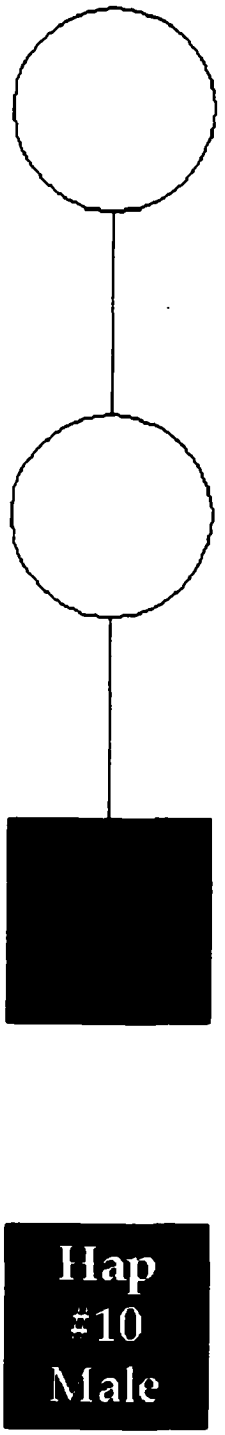


Figure 4.14: Y-chromosome haplotypes 1, 2, and 13a

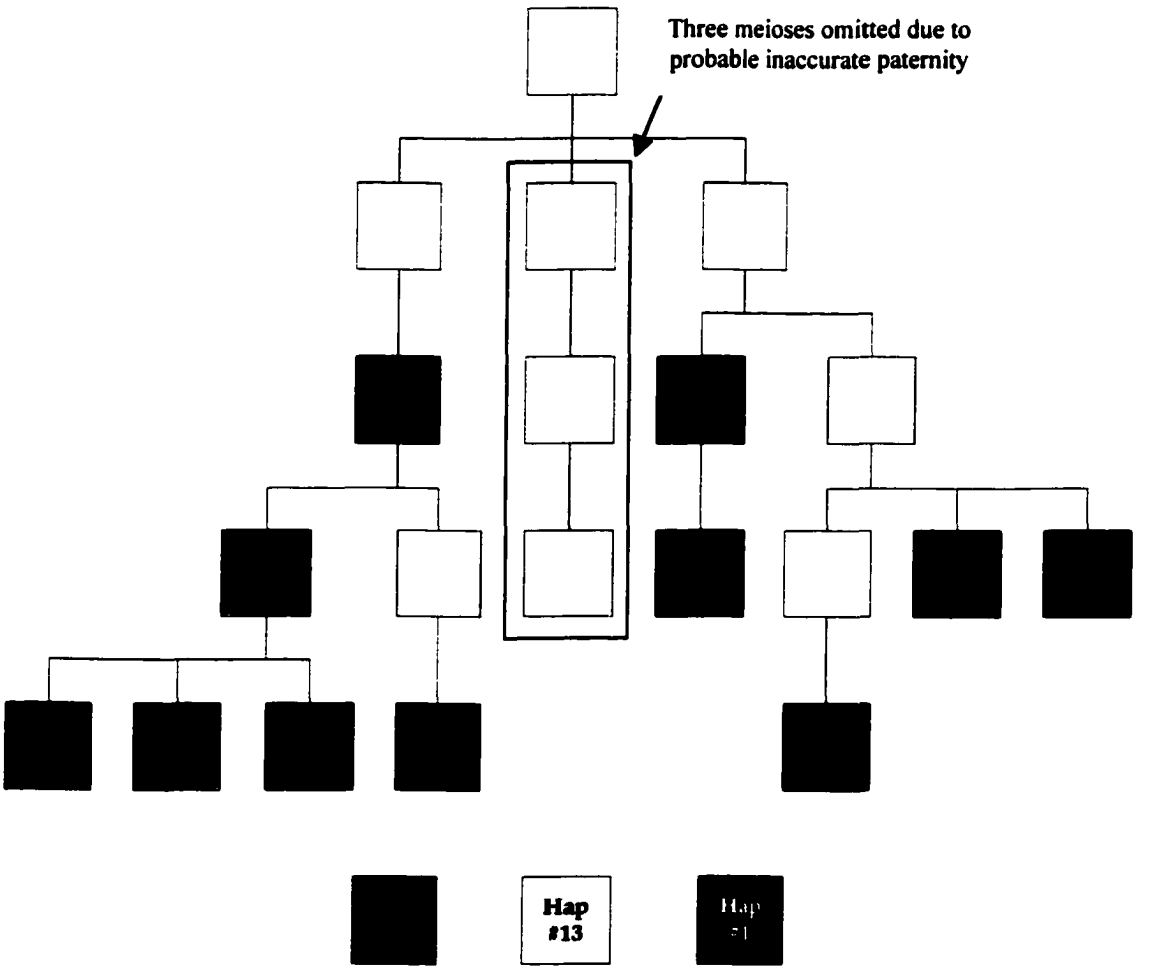


Figure 4.15: Y-chromosome haplotype 3a

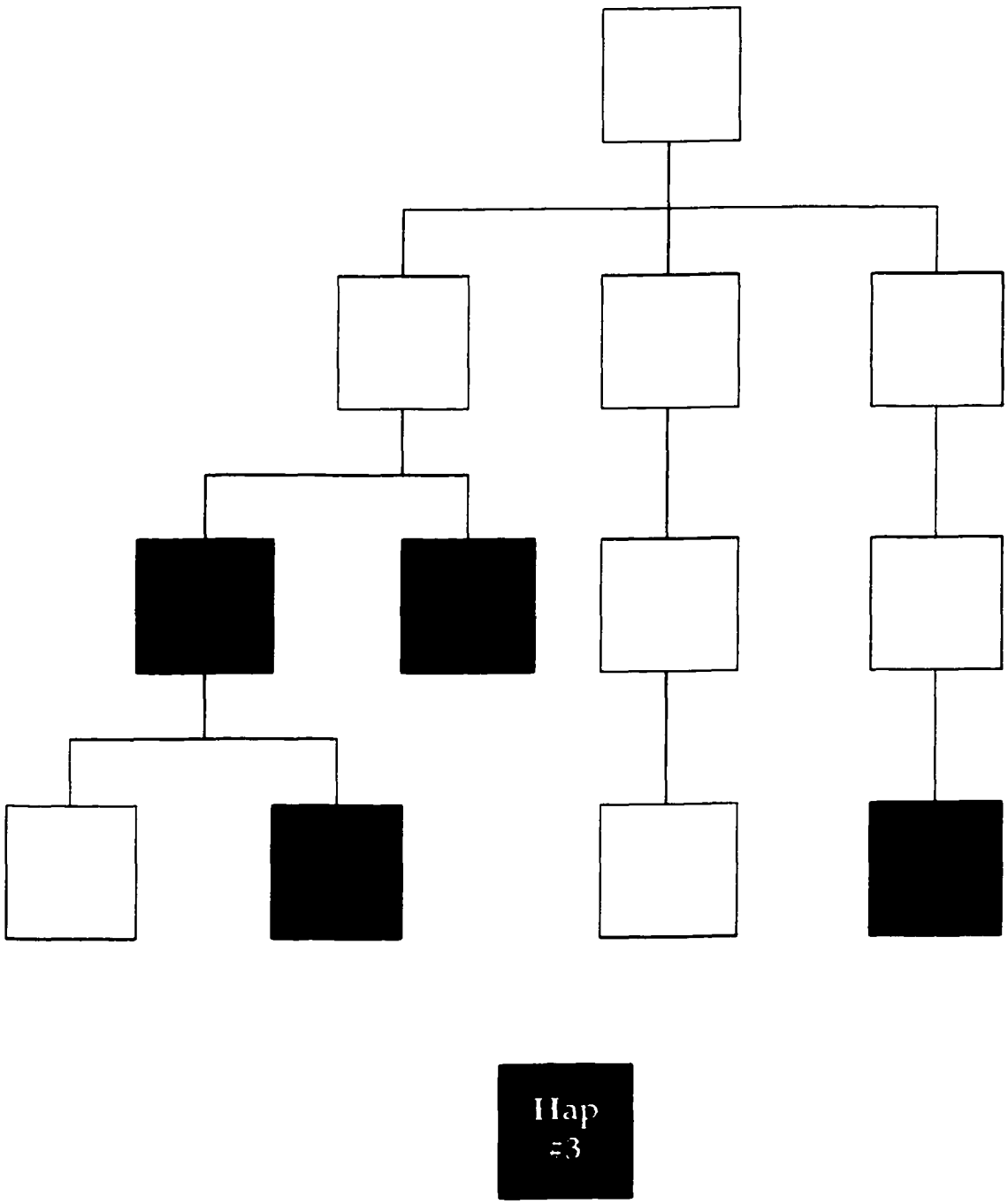


Figure 4.16: Y-chromosome haplotype 3b

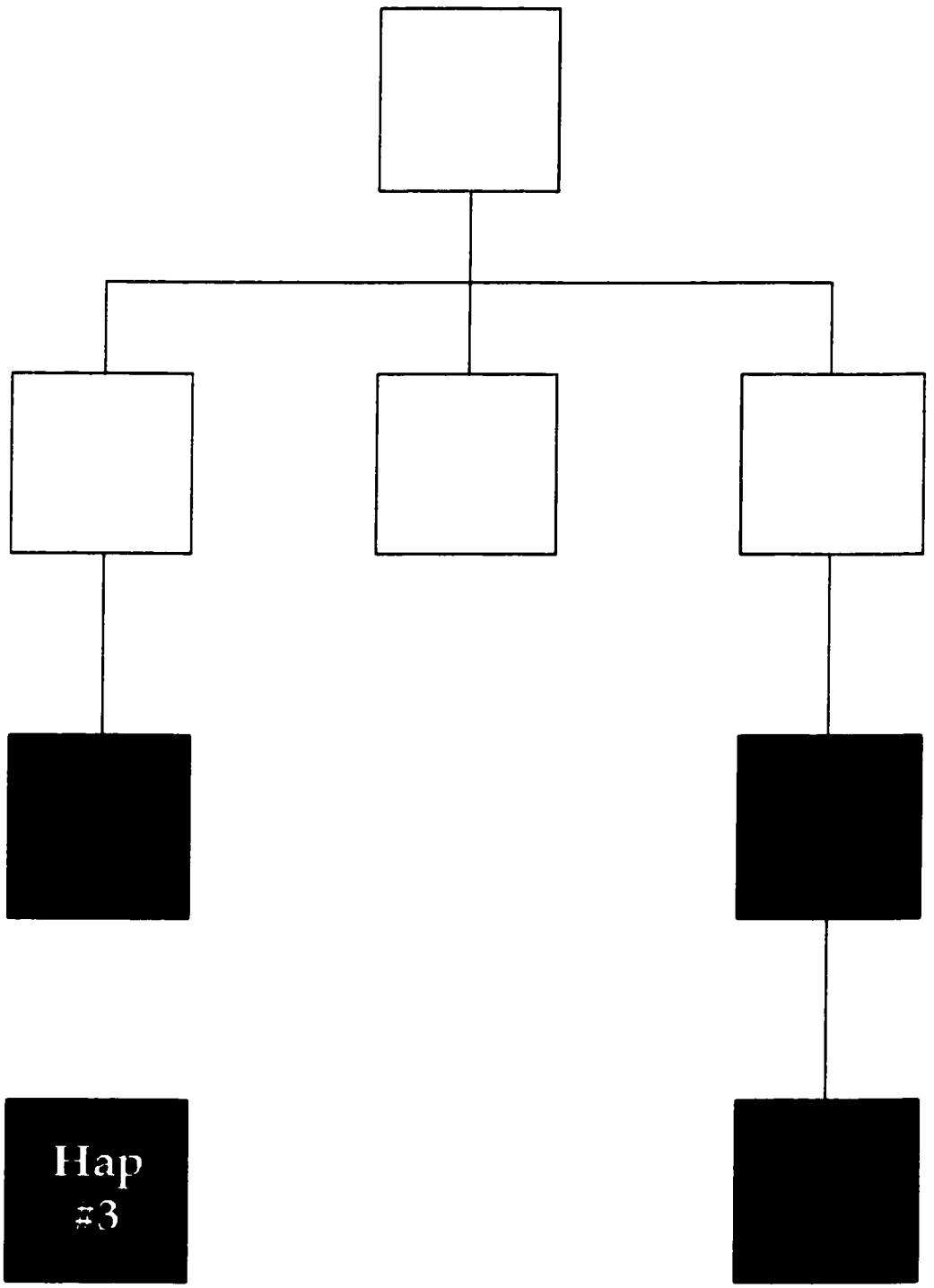


Figure 4.17: Y-chromosome haplotype 3c

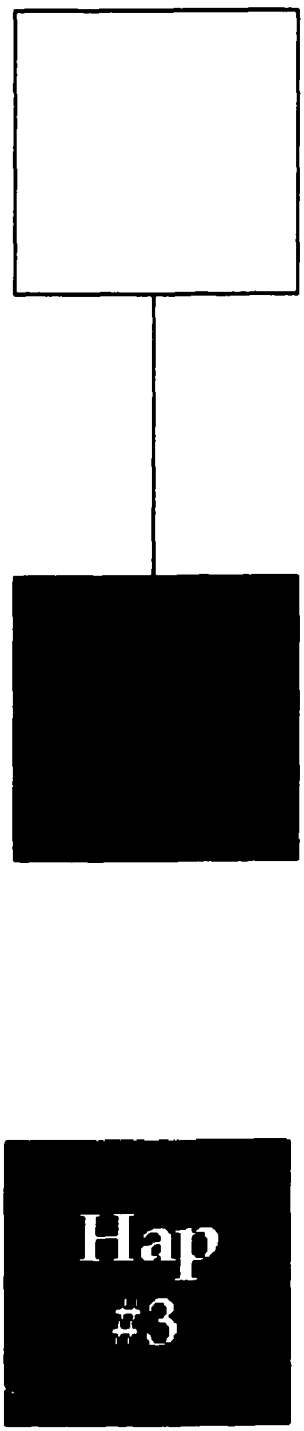


Figure 4.18: Y-chromosome haplotypes 3d and 4

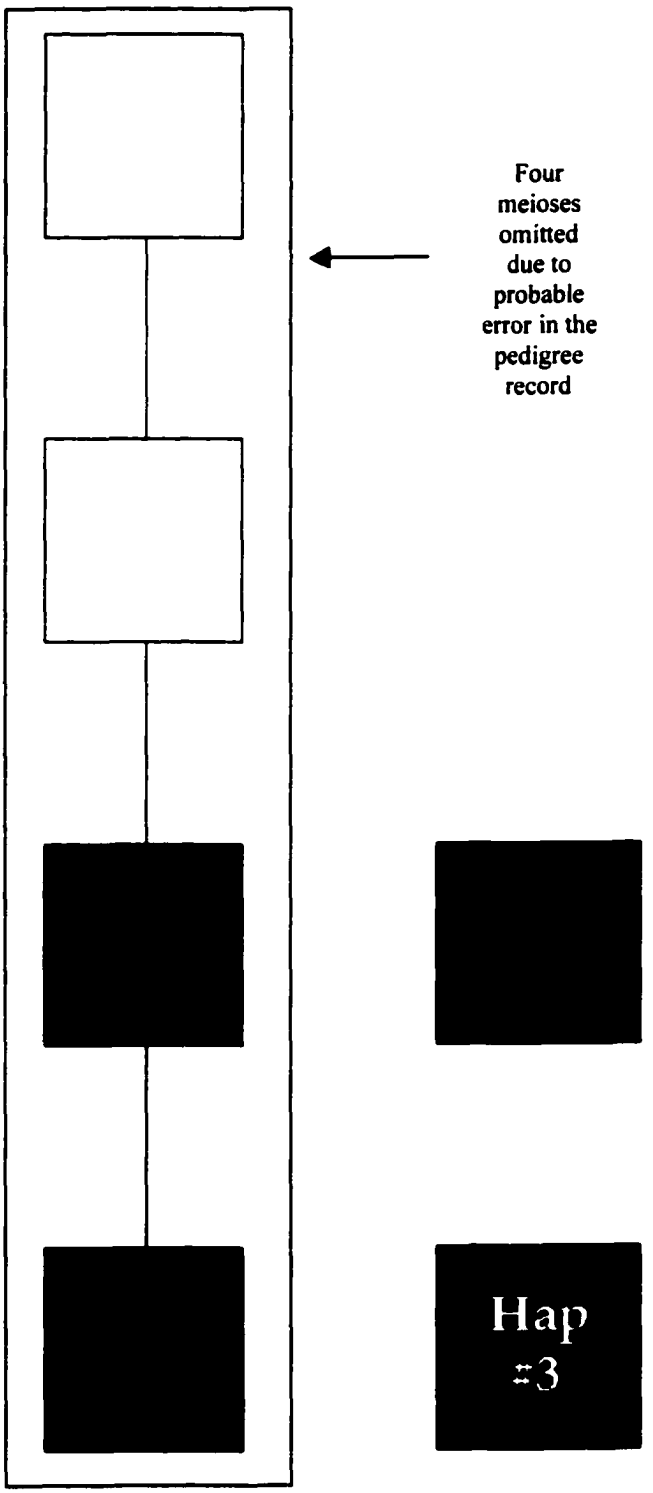


Figure 4.19: Y-chromosome haplotype 5

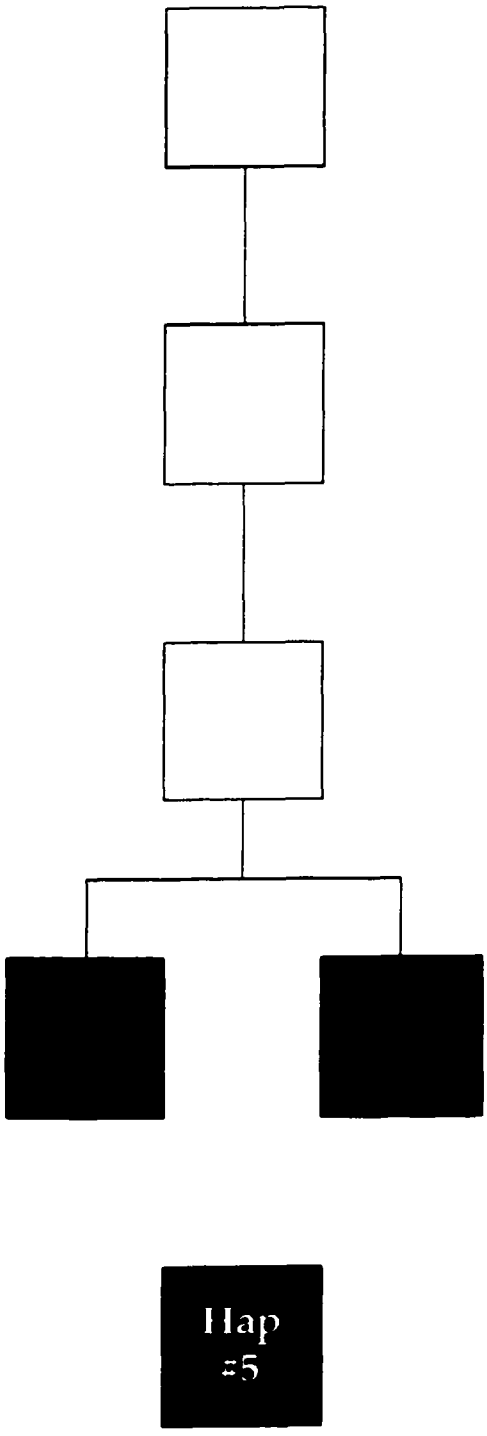


Figure 4.20: Y-chromosome haplotype 6a

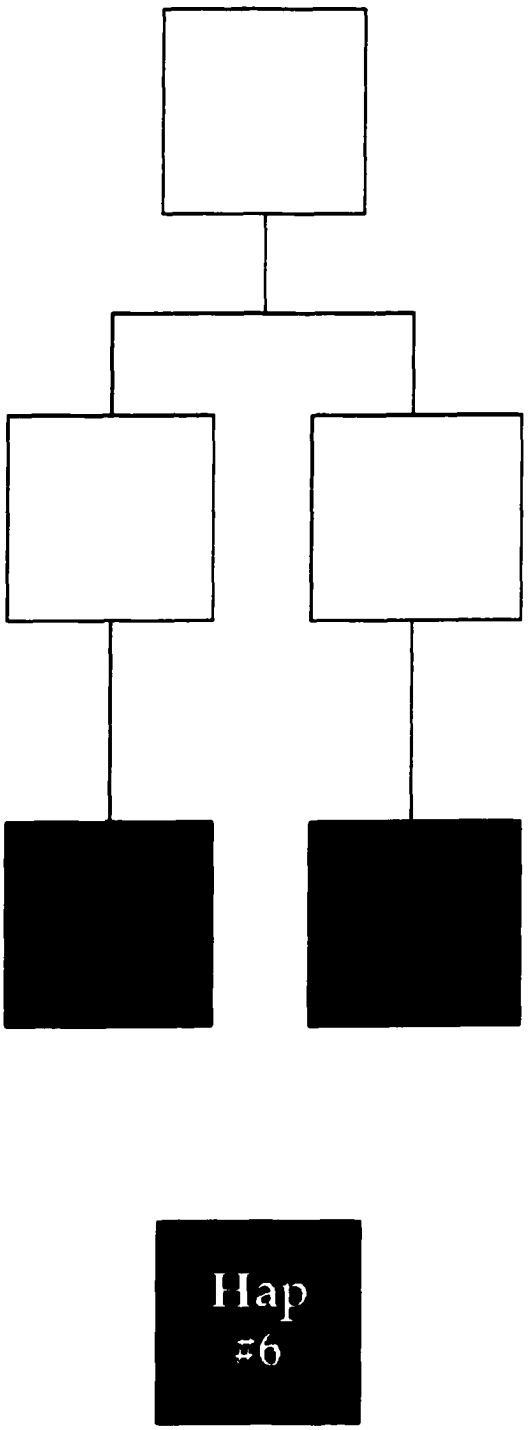


Figure 4.21: Y-chromosome haplotypes 6b and 7

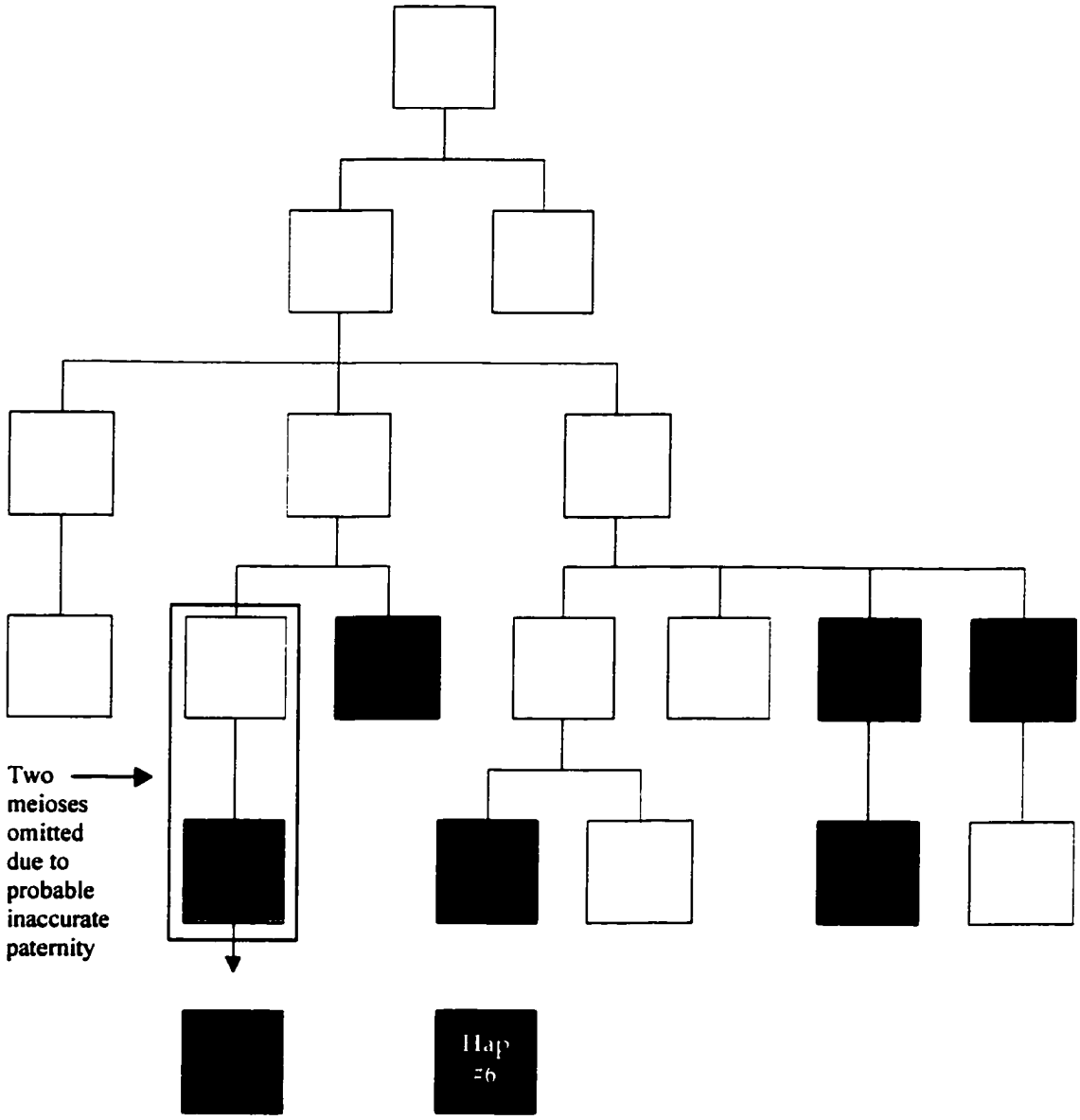


Figure 4.22: Y-chromosome haplotype 8a

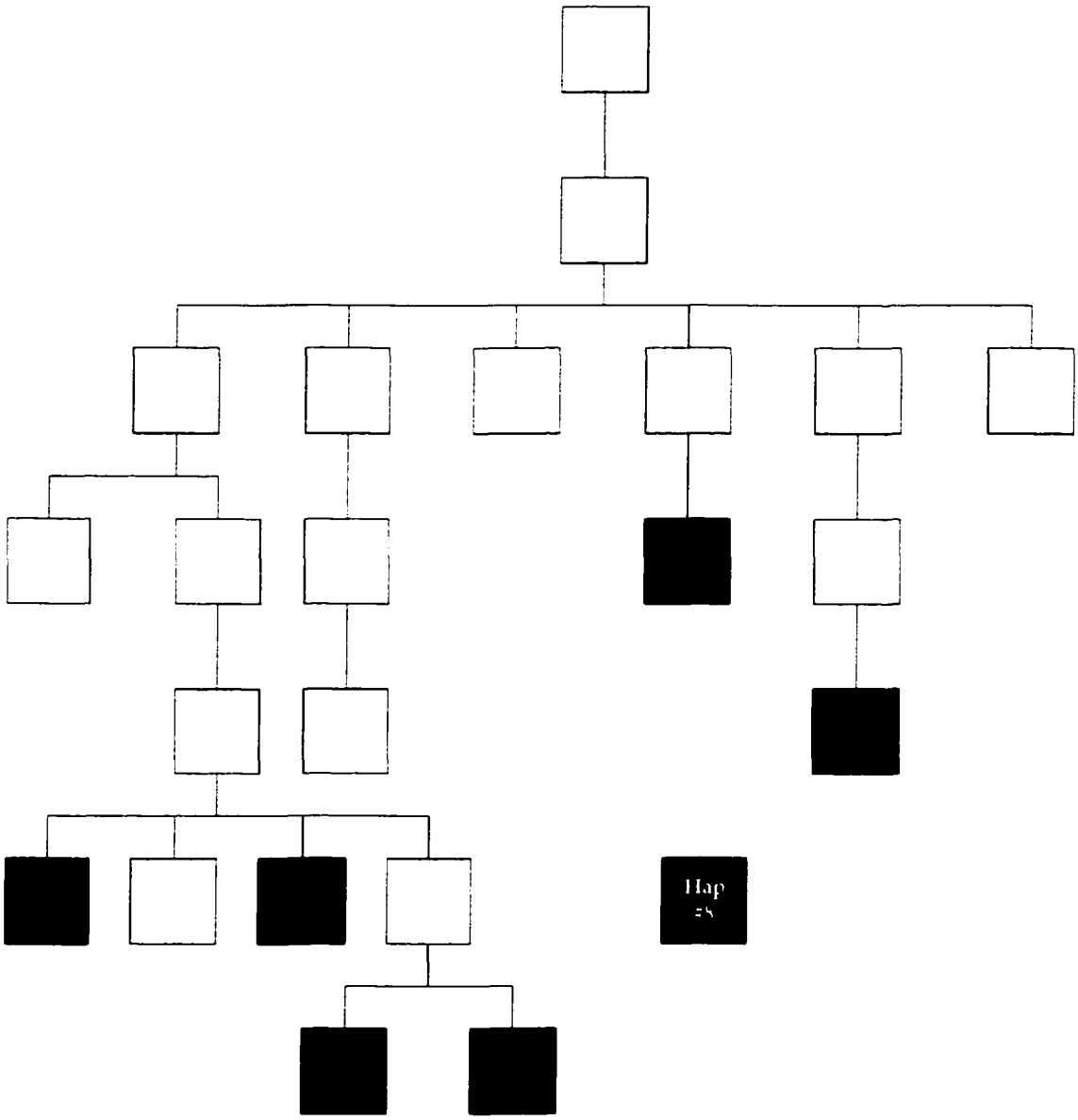


Figure 4.23: Y-chromosome haplotypes 8b and 9

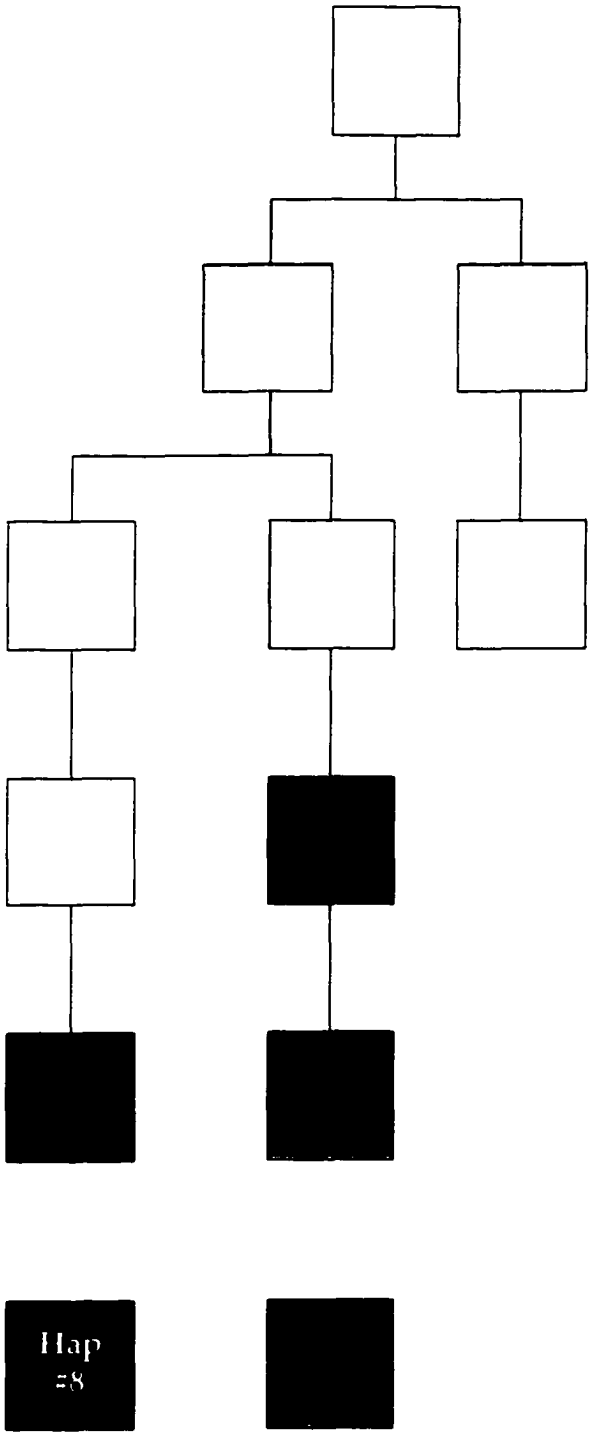


Figure 4.24: Y-chromosome haplotype 10

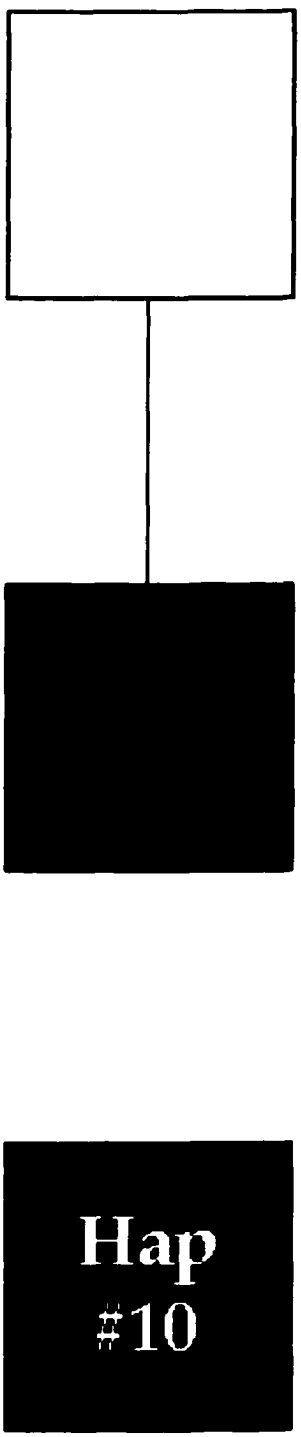


Figure 4.25: Y-chromosome haplotypes 11, 12, and 13a

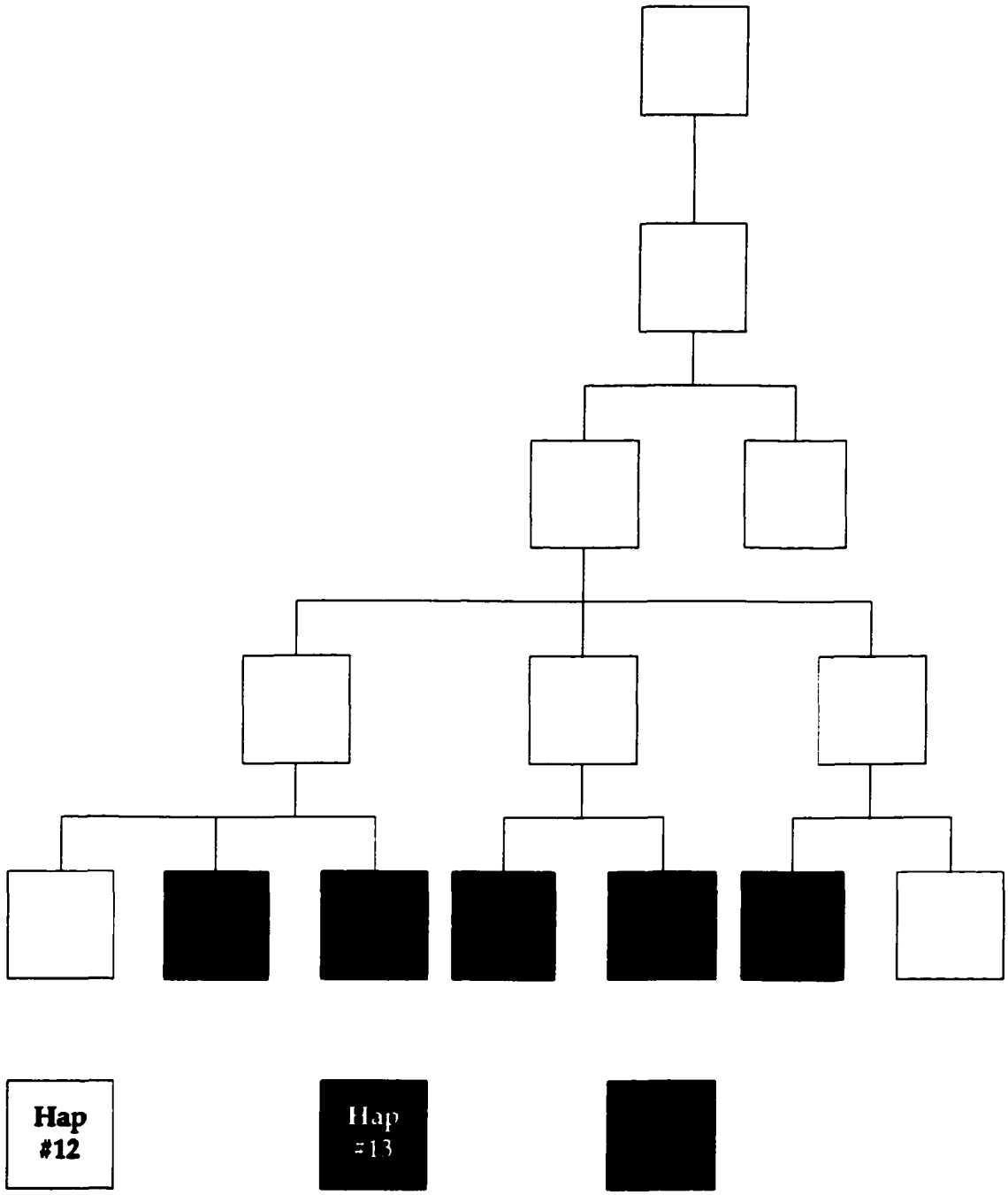


Table 4.1: Y-chromosome Mutation Within Havasupai Pedigrees

Summary of 5 mutations, at 4 STR loci, observed within Havasupai Y-Chromosome pedigrees. Mutation type refers to a gain or loss of repeat number at the loci, number of steps refers to the number of repeat changes between father and son at a particular loci, and pedigree location represents the pedigree (by figure number) in which the mutation was found.

Mutated STR Loci	Repeat Type	Father's Allele	Son's Allele	Mutation Type	No. of Steps	Pedigree Location
DYS385a	tetra-	16 (?)	15	loss	one	4.25
DYS385b	tetra-	15	16	gain	one	4.14
DYS385b	tetra-	16	15	loss	one	4.14
DYS388	tri-	12 (?)	11	loss	one	4.25
DYS392	tri-	14	12	loss	two	4.23

(?) - father not sampled (i.e. father's haplotype inferred from pedigree)

Table 4.2: mtDNA and Y-chromosome Mutation Rates

This table shows the average mutation rate (per generation and per site/generation) of Havasupai mtDNA, as well as Y-chromosome, pedigrees. Mutation rates were estimated by dividing the number of observed mutations by the number of meioses examined. Per site/generation mutation rates were estimated by dividing the mutation rate by the number of sites examined (1127 nucleotide loci for mtDNA and 12 STR loci for the NRY).

A.

mtDNA mutation rate			
No. of mutations	meioses	Mutation rate (μ/generation)	Mutation rate/site (μ/site/generation)
1	87	0.011	1.0204×10^{-5}

B.

Y-chromosome mutation rate			
No. of mutations	meioses	Mutation rate (μ/generation)	Mutation rate/site (μ/site/generation)
5	90	0.056	0.0046

C.

Y-chromosome mutation rate (stepwise model)			
No. of mutations	meioses	Mutation rate (μ/generation)	Mutation rate/site (μ/site/generation)
6	90	0.067	0.0056

A. Average rate of mtDNA mutation

B. Average rate of Y-chromosome STR mutation

C. Average rate of Y-chromosome STR mutation assuming stepwise mutation model

SUMMARY

Extensive analysis of genetic variation within the Havasupai Indians of northern Arizona has revealed several interesting observations. First, the Havasupai show a significantly higher level of mtDNA diversity than was initially predicted. An examination of mtDNA sequence diversity among 43 Havasupai revealed 10 unique haplotypes, of which 60% are considered rare. The unique haplotypes observed in the Havasupai may potentially exist in the populations surveyed herein, but missed due to inadequately sampling. Alternatively, they could be present, even frequent, in populations from the southwest United States that have yet to be characterized for mtDNA CR sequence. Regardless, without extensive sampling of New World populations, especially those that are geographically and/or culturally related to the Havasupai, no indication of genetic diversity can be attained by the detection of these unique haplotypes.

On the contrary, the presence of rare haplotypes suggests that the level of diversity in the Havasupai prior to the bottleneck might have been substantial despite their small population size, probably a remnant of a greater genetic landscape associated with the larger Pai population from which the they are derived (Martin, 1986). The presence (i.e. maintenance) of rare haplotypes after a population bottleneck is unusual, as bottlenecks are known to reduce sequence diversity. The Havasupai, which contain 60% rare haplotypes, display as much diversity, as characterized by percent of rare haplotypes, as populations that are 100 times (e.g. the Kuna) to 200 times (e.g. the Ngöbé) their size.

Indeed, the average percentage of rare haplotypes in all Native American populations is only 45.7% (Stone and Stoneking, 1998).

Haplotype diversity in the Havasupai was also considerably higher than expected, as was nucleotide diversity, as estimated by E(ν). Increased haplotype diversity was likely the result of the high incidence of rare haplotypes. E(ν), though affected by rare polymorphisms, is independent of haplotype frequency and measures a long-term average diversity rather than that of a current generation (Watterson, 1975). It is, therefore, more reflective of changes in demography (e.g. expansion, reduction, and/or selection) and likely represents the higher levels of diversity that were apparently present in the Havasupai in the past.

These data indicate that the Havasupai, historically, maintained a relatively high level of genetic diversity for a population of such small size. The small population size of the Havasupai is supported both empirically and historically. Empirically, nucleotide diversity, as estimated by π, is, unlike E(ν), not affected by rare polymorphisms, and serves principally as an indicator of population size. Hence, the low estimates of π observed in the Havasupai probably represent low long-term effective population size as opposed to loss of diversity. This, of course, is supported by historical evidence, which purports that the entire Pai population numbered less than 1500 in the 1850's and the Havasupai, once separated from the rest of the Pai, had an initial population of only 250-300 members (Martin, 1986). Another indication of the small initial population size ostensibly displayed by the Havasupai is the observation that only two (B and C) of the four (A, B, C, and D) major founding lineages are represented, with haplogroup C far

more common than haplogroup B. Although a limited number of founding haplotypes could also be suggestive of populations that have undergone bottlenecks in the past (e.g. the Kuna and Ngöbé of Panama), the maintenance of rare haplotypes since the bottleneck appears to indicate that the limited number of founding haplotypes observed in the Havasupai is a reflection of the small initial population size rather than the effect of the recent bottleneck.

The data from the mtDNA CR, taken together, seems to suggest that despite high levels of inbreeding (Markow and Martin, 1993) and a population bottleneck at the turn of the century that reduced the number of reproducing females and males to 43 and 42, respectively, the tribe maintains a relatively high level of genetic variation. Thus, our original hypothesis, that the Havasupai would exhibit less mtDNA variation than other New World groups, must be rejected. Examination of mitochondrial diversity in the Havasupai seems to suggest a picture whereby the Havasupai, initially a small population (as indicated by the limited representation of founder haplogroups and the low estimate of π), underwent a population expansion at some point that generated a relatively large amount of diversity (as evidenced by the number of rare haplotypes, the high haplotype diversity, and the high estimates of [E(ν)]). As the level of diversity displayed by the Havasupai seems to have been maintained within the population since the bottleneck at the turn of the century, it must have been too small and/or short to have any detectable effect on the mtDNA variability of the population.

Our second hypothesis, that the Havasupai Y-chromosome would display less genetic variation than Havasupai mtDNA, was also not supported. Examination of the

NRY of 48 male Havasupai using 3 SNPs and 12 STRs revealed the presence of 13 haplotypes belonging to 2 Y-chromosome haplogroups, 1Ca (56.25%) and 1G (43.75%). Thus, the number of haplotypes, 13 for the NRY and 10 for mtDNA, as well as diversity of haplotypes, 0.89 for NRY compared with 0.81 for mtDNA, was greater for the Y-chromosome. Nucleotide diversity, as estimated by the average number of pairwise differences, was also greater for the NRY than for mtDNA, 4.94 and 3.74 respectively. Hence, it would appear that the NRY displays slightly higher levels of genetic diversity than does mtDNA, at least within this population and we must, therefore, reject our initial hypothesis that the NRY will contain reduced genetic diversity compared with mtDNA in the Havasupai.

Due to the lack of a between-group comparison, however, these results must be interpreted carefully. Specifically, as there are many different factors that contribute to total diversity, we cannot know whether the difference in diversity seen between the Y-chromosome and mtDNA of the Havasupai is the result of differences in mutation rate, cultural differences, (e.g. polygyny), or a true difference in relative diversity. Examination of another, preferentially closely related, population such as the Hualapai or the Yavapai would allow the desired between-group comparisons and make estimates of relative diversity more reliable.

Another interesting observation was made in the analyses of Havasupai Y-chromosomes. The DYS392 allele of 12 repeats, which is rarely found anywhere in the world (Forster et al., 2000), was found in the Havasupai. While the locus is known to be hypervariable in Native Americans, it is generally represented by alleles >13 repeats and

has not been observed to transgress the 13-11 repeat leap (Forster et al., 2000). Lastly, the high levels of Y-chromosome diversity detected in the Havasupai appear to confirm the previously established conclusion, that the bottleneck at the turn of the century had little observable effect on genetic diversity, either mtDNA or NRY, of the tribe.

The availability of Havasupai pedigrees warranted an estimation of mutation rates for both mtDNA and the NRY in the Havasupai. Examination of 87 female germline transmission events among 11 maternal pedigrees consisting of 10 mtDNA haplotypes revealed one potential mutation. Assuming that the pedigrees were accurate, the average mutation rate of mtDNA in the Havasupai was estimated to be 1.0204×10^{-5} per site per generation or 10.2 per site per million generations. This estimate is consistent with recent investigations that have also established mtDNA mutation rates by direct examination of pedigrees, estimated by Heyer et al. (2001) to be 15.5 per site per million generations.

Among the 90 paternal meioses examined in the Y-chromosome pedigrees, 5 potential mutations were observed and a mutation rate of 0.0046 per site per generation was obtained. This rate estimation is two- to four-times higher than those reported for the Y-chromosome in previous pedigree studies, which range from an average of ~0.0012 (Bianchi et al., 1998 averaged their own data with that of Heyer et al., 1997 and Kayser et al., 1997) to 0.0028 per site per generation (Kayser et al., 2000). Additionally, they are an order of magnitude greater than those reported for "evolutionary" mutation rates, estimated by phylogenetics, which average ~0.00026 per site per generation (Forster et al., 2000), as well as those from autosomal microsatellites, which range from 0.0006-0.0027 per site per generation (Brinkmann et al., 1998; Henke and Henke, 1999; Sajantila

et al., 1999). Analysis of NRY mutation in the pedigrees also led to the observation that the Havasupai show a distinct, though not statistically significant ($p = 0.1$), preference for loss of repeats despite a previously demonstrated mutational bias in favor of repeat expansion in other populations (Cooper et al., 1999; Kayser et al., 2000). Furthermore, our data support the findings of Kayser et al. (2000) in that the premutation allele size of all mutations observed in the Havasupai, despite a propensity for repeat reduction, was greater than 11 repeats.

Finally, the observed maternal transmission of NIDDM in the Havasupai (Fenger, 1992) and the existence of NIDDM-related retinopathies in some pedigrees, lead to the hypothesis that members of these affected lineages should exhibit mtDNA mutations associated with NIDDM and NIDDM-related retinopathy in other studies. DNA sequence analysis of several potentially etiological loci including the tRNA Leu(UUR) gene, the tRNA Lys gene, and the intergenic region between COII and tRNA Lys in 52 Havasupai individuals revealed no variation whatsoever in either of the tRNA genes. The only identified mutation, a 9 bp deletion in the COII-tRNA Lys intergenic region, has not been associated with disease in any known populations. The 9 bp deletion, characteristic of individuals belonging to founder Native American haplogroup B, was found in 26.9% of individuals surveyed. The prevalence of haplogroup B based on the 9 bp deletion is almost three times the percentage of haplotype B individuals found in our earlier study based on CR sequencing (Chapter 1), which indicated that the Havasupai contain only 9.3% haplogroup B. Pooling the data from this investigation with that from the previous investigation of 43 mtDNA CR sequences (Chapter 1), yielded a total of 55 individuals,

of which 25.5% belong to haplogroup B. Although there is no evidence to suggest that the 9 bp deletion characterizing haplogroup B has any association with NIDDM, it is interesting to note that there is an increase, though not statistically significant, in the occurrence of NIDDM in haplogroup B individuals (64.3%) compared with those belonging to haplogroup C (43.9%). Examination of the ten mtDNA haplotypes characterized by CR sequencing revealed that 87.5% of individuals displaying the Hav 2 haplotype are affected by NIDDM, whereas only 12.5% are unaffected ($p < 0.05$). Analysis of the CR polymorphisms on a individual site basis demonstrated that a C -> T transition at bp 16192 was found in 7/43 individuals (16.3%) affected by NIDDM compared to only 1/43 unaffected individual (2.3%) ($p < 0.05$). It is noteworthy that the presence of this substitution is what predominantly characterizes haplotype Hav 2.

In summary, the Havasupai are an interesting and unique population both historically and genetically. Despite their small size, isolation, and a recent population bottleneck, the Havasupai maintain relatively high levels of genetic diversity in both their maternal and paternal lineages. As mentioned previously, the loss of diversity seen after a population bottleneck is dependent on the size and duration of that bottleneck. Although the Havasupai bottleneck lasted approximately ten years and reduced the population by less than half (from ~300 to 166 individuals), a bottleneck of this size and duration is apparently not sufficient to cause a substantial reduction in genetic diversity. If not, the number of rare haplotypes contained within the population before the bottleneck, would be largely maintained during and after the bottleneck and this appears to be the case with the Havasupai. Thus, whatever loss of genetic diversity affecting the

Havasupai through genetic drift, founder effect, and consanguineous mating, was present before and apparently, unaffected by, the population bottleneck that occurred at the turn of the century. They display a rate of mtDNA mutation that is consistent with those of other populations surveyed, whereas, the Y-chromosome mutation rate is twice that of any population yet observed. Although the Havasupai demonstrate the third largest incidence of NIDDM of any population worldwide, it would appear that the etiology of NIDDM in this population does not lie within the examined regions of the mtDNA. Future investigations should focus on examination of closely related populations such as the Hualapai, Yavapai, and/or Pai Pai to determine whether or not the interesting trends observed in the Havasupai are unique to them, or common among tribes of the Colorado river or the southwest United States in general.

REFERENCES

Alcolado JC, Majid A, Brockington M, Sweeney MG, Morgan R, Rees A, Harding AE, Barnett AH (1994) Mitochondrial gene defects in patients with NIDDM. Diabetologia 37(4): 372-6.

Allen JF (1996) Separate sexes and the mitochondrial theory of ageing. J Theor Biol 180(2): 135-40.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290(5806): 457-65.

Ankel-Simons F, Cummins JM (1996) Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. Proc Natl Acad Sci U S A. 93(24): 13859-63.

Awadalla P, Eyre-Walker A, Smith JM (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286(5449): 2524-5.

Bailliet G, Rothhammer F, Carnese FR, Bravi CM, Bianchi NO (1994) Founder mitochondrial haplotypes in Amerindian populations. Am J Hum Genet 55(1): 27-33.

Banchs I, Bosch A, Guimera J, Lazaro C, Puig A, Estivill X (1994) New alleles at microsatellite loci in CEPH families mainly arise from somatic mutations in the lymphoblastoid cell lines. Hum Mutat. 3(4): 365-72.

Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci U S A 87(5): 1816-9.

Barrantes R, Smouse PE, Mohrenweiser HW, Gershowitz H, Azofeifa J, Arias TD, Neel JV (1990) Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. Am J Hum Genet 46(1): 63-84.

Batista O, Kolman CJ, Bermingham E (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panama. Hum Mol Genet 4(5): 921-9.

Bendall KE, Sykes BC (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 57(2): 248-56.

Bergstrom CT, Pritchard J (1998) Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics 149(4): 2135-46.

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Lopez-Camelo JS (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 63(6): 1862-71.

Black FL, Salzano FM, Berman LL, Gabbay Y, Weimer TA, Franco MH, Pandey JP (1983) Failure of linguistic relationships to predict genetic distances between the Waiapi and other tribes of lower Amazonia. Am J Phys Anthropol 60(3): 327-35.

Bosch E, Calafell F, Santos FR, Perez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet. 65(6): 1623-38.

Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62(6): 1408-15.

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A 76(4): 1967-71.

Brown MD, Wallace DC (1994) Molecular basis of mitochondrial DNA disease. J Bioenerg Biomembr 26(3): 273-89.

Brown DT, Samuels DC, Michael EM, Turnbull DM, Chinnery PF (2001) Random genetic drift determines the level of mutant mtDNA in human primary oocytes. Am J Hum Genet 68(2): 533-6.

Byrne E (1991) Mitochondrial DNA abnormalities in human disease. Med J Aust 154(10): 646-7.

Calabresi PA, Silvestri G, DiMauro S, Griggs RC (1994) Ekbom's syndrome: lipomas, ataxia, and neuropathy with MERRF. Muscle Nerve 17(8): 943-945.

Cardaioli E, Fabrizi GM, Grieco GS, Dotti MT, Federico A (2000) Heteroplasmy of the A3243G transition of mitochondrial tRNA(Leu(UUR)) in a MELAS case and in a 25-week-old miscarried fetus. J Neurol 247(11): 885-7.

Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci U S A 85(16): 6002-6.

Cavalli-Sforza LL, Minch E, Mountain JL (1992) Coevolution of genes and languages revisited. Proc Natl Acad Sci U S A 89(12): 5620-4.

Chakraborty R (1976) Cultural, language and geographical correlates of genetic variability in Andean highland Indians. Nature 264(5584): 350-2.

Chakraborty R (1977) Simulation results with stepwise mutation model and their interpretations. J Mol Evol. 9(4): 313-22.

Chinnery PF, Turnbull DM (1999) Mitochondrial DNA and disease. Lancet 354 Suppl 1:SI17-21.

Chinnery PF, Thorburn DR, Samuels DC, White SL, Dahl HM, Turnbull DM, Lightowlers RN, Howell N (2000) The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? Trends Genet 16(11): 500-5.

Chomyn A, Martinuzzi A, Yoneda M, Daga A, Hurko O, Johns D, Lai ST, Nonaka I, Angelini C, Attardi G (1992) MELAS mutation in mtDNA binding site for transcription termination factor causes defects in protein synthesis and in respiration but no change in levels of upstream and downstream mature transcripts. Proc Natl Acad Sci U S A 89(10): 4221-5.

Chomyn A, Lai ST, Shakeley R, Bresolin N, Scarlato G, Attardi G (1994) Platelet-mediated transformation of mtDNA-less human cells: analysis of phenotypic variability among clones from normal individuals and complementation behavior of the tRNALys mutation causing myoclonic epilepsy and ragged red fibers. Am J Hum Genet 54(6): 966-974.

Christianson TW, Clayton DA (1986) *In vitro* transcription of human mitochondrial DNA: accurate termination requires a region of DNA sequence that can function bidirectionally. Proc Natl Acad Sci U S A 83(17): 6277-81.

Coon and Markow, in preparation.

Cooper G, Burroughs NJ, Rand DA, Rubinsztein DC, Amos W (1999) Markov chain Monte Carlo analysis of human Y-chromosome microsatellites provides evidence of biased mutation. Proc Natl Acad Sci U S A 96(21): 11916-21.

Crow JF (1997) Molecular evolution--who is in the driver's seat? Nat Genet. 17(2): 129-30.

Crow JF (1999) Spontaneous mutation in man. Mutat Res. 437(1): 5-9.

Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet. 1(1): 40-7.

de Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet. 67(5): 1055-61.

de Vries DD, De Wijs IJ, Wolff G, Ketelsen UP, Ropers HH, Van Oost BA (1993) X-linked myoclonus epilepsy explained as a maternally inherited mitochondrial disorder. Hum Genet 91: 51-54.

DeMartino CA, Floride A, Marcante ML, Malorni W, Scorza-Barcellona P, Bellocci M, and Silvestrini B (1979) Morphological, histochemical, and biochemical studies on germ cell mitochondria of normal rats. Cell and Tissue Research 196: 1-22.

DiMauro S, Bonilla E, Davidson M, Hirano M, Schon EA (1998) Mitochondria in neuromuscular disorders. Biochim Biophys Acta 1366(1-2): 199-210.

Doda JN, Wright CT, Clayton DA (1981) Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc Natl Acad Sci U S A 78(10): 6116-20.

Easton RD, Merriwether DA, Crews DE, Ferrell RE (1996) mtDNA variation in the Yanomami: evidence for additional New World founding lineages. Am J Hum Genet 59(1): 213-25.

Excoffier L, Yang Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16(10): 1357-68.

Eyre-Walker A, Smith NH, Smith JM (1999) How clonal are human mitochondria? Proc R Soc Lond B Biol Sci 266(1418): 477-83.

Eyre-Walker A, Awadalla P (2001) Does human mtDNA recombine? J Mol Evol 53(4-5): 430-5.

Fenger DP (1992) A Genetic epidemiological study of non-insulin-dependent diabetes mellitus in the Havasupai. MS Thesis. Arizona State University.

Fix A (1999) Migration and colonization in human microevolution. Cambridge University Press, New York.

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67(1): 182-96.

Froguel P, Vaxillaire M, Sun F, Velho G, Zouali H, Butel MO, Lesage S, Vionnet N, Clement K, Fougerousse F, et al. (1992) Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. Nature 356: 162-164.

Froguel P, Velho G (1993) Non-sense mutation of glucokinase gene. Lancet 341: 385-386.Fenger DP (1992) A Genetic epidemiological study of non-insulin-dependent diabetes mellitus in the Havasupai. MS Thesis. Arizona State University.

Gelfand R, Attardi G (1981) Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable. Mol Cell Biol 1(6): 497-511.

Gerbitz KD, van den Ouweland JMW, Maassen JA, Jaksch M (1995) Mitochondrial diabetes mellitus: a review. Biochim Biophys Acta 1271: 253-260.

Gerbitz KD, Gempel K, Brdiczka D (1996) Mitochondria and diabetes. Genetic, biochemical, and clinical implications of the cellular energy circuit. Diabetes 45(2): 113-26.

Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, Hutz MH, Anderson A, Just J, Salzano FM, King MC (1993) Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. EXS 67: 211-9.

Graf WD, Sumi SM, Copass MK, Ojemann LM, Longstreth WT, Shanske S, Lombes A, DiMauro S (1993) Phenotypic heterogeneity in families with the myoclonic epilepsy and ragged-red fiber disease point mutation in mitochondrial DNA. Ann Neurol 33: 640-645.

Greenberg JH, Turner CG II, Zegura SL (1986) The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence. Curr Anthropol 27(5): 477-497.

Greenberg JH (1987) Language in the Americas. Stanford University Press. Stanford, CA.

Gyllensten U, Wharton D, Wilson AC (1985) Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. J Hered. 1985 76(5): 321-4.

Gyllensten U, Wharton D, Josefsson A, Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. Nature 352(6332): 255-7.

Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc Lond B Biol Sci 266(1418):485-92.

Hammans SR, Sweeney MG, Wicks DA, Morgan-Hughes JA, Harding AE (1992) A molecular genetic study of focal histochemical defects in mitochondrial encephalomyopathies. Brain 115: 343-65.

Hammans SR, Sweeney MG, Brockington M, Lennox GG, Lawton NF, Kennedy CR, Morgan-Hughes JA, Harding AE (1993) The mitochondrial DNA transfer RNALys A-G(8344) mutation and the syndrome of myoclonic epilepsy with ragged red fibres (MERRF): Relationship of clinical phenotype to proportion of mutant mitochondrial DNA. Brain 116: 617-632.

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15(4): 427-41.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol. 18(7): 1189-203.

Hao H, Bonilla E, Manfredi G, DiMauro S, Moraes CT (1995) Segregation patterns of a novel mutation in the mitochondrial tRNA glutamic acid gene associated with myopathy and diabetes mellitus. Am J Hum Genet 56(5): 1017-25.

Harpending HC, Sherry ST, Rogers AR, and Stoneking M (1993) The genetic structure of ancient human populations. Curr. Anthropol 34: 483-496.

Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum Biol 66(4): 591-600.

Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol. 37(4): 347-54.

Hauswirth WW, Laipis PJ (1982) Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. Proc Natl Acad Sci U S A 79(15): 4686-90.

Hauswirth WW, Laipis PJ (1985) Achievements and Perspectives of Mitochondrial Research Volume II: Biogenesis. Ed. Quagliariello E. Elsevier Science Publishers. New York, NY.

Hayashi J, Tagashira Y, Yoshida MC, Ajiro K, Sekiguchi T (1983) Two distinct types of mitochondrial DNA segregation in mouse-rat hybrid cells. Stochastic segregation and chromosome-dependent segregation. Exp Cell Res 147(1): 51-61.

Hecht NB, Liem H, Kleene KC, Distel RJ, Ho SM (1984) Maternal inheritance of the mouse mitochondrial genome is not mediated by a loss or gross alteration of the paternal mitochondrial DNA or by methylation of the oocyte mitochondrial DNA. Dev Biol 102(2): 452-61.

Henke J, Henke L (1999) Mutation rate in human microsatellites. Am J Hum Genet. 64(5): 1473-4.

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6(5): 799-803.

Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D (2001) Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. Am J Hum Genet. 69(5): 1113-26.

Holt IJ, Harding AE, Petty RK, Morgan-Hughes JA (1990) A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. Am J Hum Genet 46(3): 428-33.

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Mol Biol Evol 10(1): 23-47.

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci U S A 92(2): 532-6.

Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet. 59(3): 501-9.

Howell N, Smejkal CB (2000) Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction. Am J Hum Genet 66(5): 1589-98.

http://www.cinprograms.org/people/coloradoriver/havasupai.html

Ikebe S, Tanaka M, Ohno K, Sato W, Hattori K, Kondo T, Mizuno Y, Ozawa T (1990) Increase of deleted mitochondrial DNA in the striatum in Parkinson's disease and senescence. Biochem Biophys Res Commun 170(3): 1044-8.

Jacobi FK, Leo-Kottler B, Mittelviefhaus K, Zrenner E, Meyer J, Pusch CM, Wissinger B (2001) Segregation patterns and heteroplasmy prevalence in Leber's hereditary optic neuropathy. Invest Ophthalmol Vis Sci 42(6): 1208-14.

Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllensten U (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. Nat Genet. 18(2): 109-10.

Jenuth JP, Peterson AC, Fu K, Shoubridge EA (1996) Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet 14(2): 146-51.

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11(11): 449-56.

Johns DR, Drachman DB, Hurko O (1989) Identical mitochondrial DNA deletion in blood and muscle. Lancet 1(8634): 393-4.

Jones JS, Rouhani S (1986) Human evolution. How small was the bottleneck? Nature 319(6053): 449-50.

Jorde LB, Bamshad M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288(5473): 1931.

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet. 66(3): 979-88.

Kadowaki T, Kadowaki H, Mori Y, Tobe K, Sakuta R, Suzuki Y, Tanabe Y, Sakura H, Awata T, Goto Y, et al. (1994) A subtype of diabetes mellitus associated with a mutation of mitochondrial DNA. N Engl J Med 330(14): 962-8.

Kameoka K, Isotani H, Tanaka K, Azukari K, Fujimura Y, Shiota Y, Sasaki E, Majima M, Furukawa K, Haginomori S, Kitaoka H, Ohsawa N (1998) Novel mitochondrial DNA mutation in tRNA(Lys) (8296A-->G) associated with diabetes. Biochem Biophys Res Commun 245(2): 523-7.

Kaneda H, Hayashi J, Takahama S, Taya C, Lindahl KF, Yonekawa H (1995) Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. Proc Natl Acad Sci U S A 92(10): 4542-6.

Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, Long JC, Harihara S, Jin F, Dashnyam B, Gerelsaikhan T, Omoto K, Hammer MF (1997) Y

chromosome markers and Trans-Bering Strait dispersals. Am J Phys Anthropol. 102(3): 301-14.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am J Hum Genet. 64(3): 817-31.

Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Roewer L, et al. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110(3):125-33, 141-9.

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66(5): 1580-8.

King MP, Koga Y, Davidson M, Schon EA (1992) Defects in mitochondrial protein synthesis and respiratory chain activity segregate with the tRNA(Leu(UUR)) mutation associated with mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes. Mol Cell Biol 12(2): 480-90.

Kivisild T, Villems R (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288(5473): 1931.

Koehler CM, Lindberg GL, Brown DR, Beitz DC, Freeman AE, Mayfield JE, Myers AM (1991) Replacement of bovine mitochondrial DNA by a sequence variant within one generation. Genetics 129(1): 247-55.

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995) Reduced mtDNA diversity in the Ngobe Amerinds of Panama. Genetics 140(1): 275-83.

Kolman CJ, Bermingham E (1997) Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. Genetics 147(3): 1289-302.

Kruse B, Narasimhan N, Attardi G (1989) Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. Cell 58(2): 391-7.

Kumar S, Hedrick P, Dowling T, Stoneking M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288(5473): 1931.

Kumar S, Tamura K, Jakobsen IB, and Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17(12): 1244-5.

Kunkel TA, Loeb LA (1981) Fidelity of mammalian DNA polymerases. Science 213(4509): 765-7.

Lauber J, Marsac C, Kadenbach B, Seibel P (1991) Mutations in mitochondrial tRNA genes: a frequent cause of neuromuscular diseases. Nucleic Acids Res 19: 1393-1397.

Lertrit P, Noer AS, Byrne E, Marzuki S (1992) Tissue segregation of a heteroplasmic mtDNA mutation in MERRF (myoclonic epilepsy with ragged red fibers) encephalomyopathy. Hum Genet 90: 251-254.

Liang P, Hughes V, Fukagawa NK (1997) Increased prevalence of mitochondrial DNA deletions in skeletal muscle of older individuals with impaired glucose tolerance. Possible markers of glycemic stress. Diabetes 46: 920-923.

Lorenz JG, Smith DG (1996) Distribution of four founding mtDNA haplogroups among Native North Americans. Am J Phys Anthropol 101(3): 307-23.

Lundstrom R, Tavare S, Ward RH (1992) Estimating substitution rates from molecular data using the coalescent. Proc Natl Acad Sci U S A 89(13): 5961-5.

Macaulay VA, Richards MB, Forster P, Bendall KE, Watson E, Sykes B, Bandelt HJ (1997) mtDNA mutation rates--no need to panic. Am J Hum Genet. 61(4): 983-90.

Malaspina D, Harlap S, Fennig S, Heiman D, Nahon D, Feldman D, Susser ES (2001) Advancing paternal age and the risk of schizophrenia. Arch Gen Psychiatry 58(4): 361-7.

Marchington DR, Poulton J, Sellar A, Holt IJ (1996) Do sequence variants in the major non-coding region of the mitochondrial genome influence mitochondrial mutations associated with disease? Hum Mol Genet 5(4): 473-9.

Marchington DR, Macaulay V, Hartshorne GM, Barlow D, Poulton J (1998) Evidence from human oocytes for a genetic bottleneck in an mtDNA disease. Am J Hum Genet 63(3): 769-75.

Markow TA, Martin JF (1993) Inbreeding and developmental stability in a small human population. Ann Hum Biol 20(4): 389-94.

Markow T, Hedrick PW, Zuerlein K, Danilovs J, Martin J, Vyvial T, Armstrong C (1993) HLA polymorphism in the Havasupai: evidence for balancing selection. Am J Hum Genet 53(4): 943-52.

Martin JF (1986) The Havasupai. Plateau 56(4).

Massin P, Guillausseau PJ, Vialettes B, Paquis V, Orsini F, Grimaldi AD, Gaudric A (1995) Macular pattern dystrophy associated with a mutation of mitochondrial DNA. Am J Ophthalmol 120(2): 247-8.

Merriwether DA (1994) mtDNA and the peopling of the New World. Mother Tongue 23: 22-28.

Merriwether DA, Rothhammer F, Ferrell RE (1995) Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am J Phys Anthropol 98(4): 411-30.

Merriwether DA, Ferrell RE (1996) The four founding lineage hypothesis for the New World: a critical reevaluation. Mol Phylogenet Evol 5(1): 241-6.

Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC, Ruiz-Linares A (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. Am J Hum Genet. 67(5): 1277-86.

Michaels GS, Hauswirth WW, Laipis PJ (1982) Mitochondrial DNA copy number in bovine oocytes and somatic cells. Dev Biol 94(1): 246-51.

Mitchell RJ, Hammer MF (1996) Human evolution and the Y chromosome. Curr Opin Genet Dev 6(6): 737-42.

MITOMAP: A Human Mitochondrial Genome Database. Center for Molecular Medicine, Emory University, Atlanta, GA, USA. http://www.gen.emory.edu/mitomap.html, 2002.

Montoya J, Gaines GL, Attardi G (1983) The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell 34(1): 151-9.

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene tress vs. nuclear-gene trees. Evolution 49: 718-726.

Moraes CT, DiMauro S, Zeviani M, Lombes A, Shanske S, Miranda AF, Nakase H, Bonilla E, Werneck LC, Servidei S, et al. (1989) Mitochondrial DNA deletions in

progressive external ophthalmoplegia and Kearns-Sayre syndrome. N Engl J Med 320(20): 1293-9.

Moraes CT, Ricci E, Bonilla E, DiMauro S, Schon EA (1992a) The mitochondrial tRNA(Leu(UUR)) mutation in mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes (MELAS): genetic, biochemical, and morphological correlations in skeletal muscle. Am J Hum Genet 50(5): 934-49.

Moraes CT, Ricci E, Petruzzella V, Shanske S, DiMauro S, Schon EA, Bonilla E (1992b) Molecular analysis of the muscle pathology associated with mitochondrial DNA deletions. Nat Genet 1(5): 359-67.

Moraes CT, Ciacci F, Bonilla E, Jansen C, Hirano M, Rao N, Lovelace RE, Rowland LP, Schon EA, DiMauro S (1993) Two novel pathogenic mitochondrial DNA mutations affecting organelle number and protein synthesis. Is the tRNA(Leu(UUR)) gene an etiologic hot spot? J Clin Invest 92(6): 2906-15.

Morten KJ, Cooper JM, Brown GK, Lake BD, Pike D, Poulton J (1993) A new point mutation associated with mitochondrial encephalomyopathy. Hum Mol Genet 2(12): 2081-7.

Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 1: 2-9.

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York.

Noer AS, Sudoya H, Lertrit P, Thyagarajan D, Utthanaphol P, Kapsa R, Byrne E, Marzuki S (1991) A tRNALys mutation in the mtDNA is the causal genetic lesion underlying myoclonic epilepsy and ragged-red fiber (MERRF) syndrome. Am J Hum Genet 49: 715-722.

Normanly J, Abelson J (1989) tRNA identity. Annu Rev Biochem 58: 1029-49.

Ohta T, Kimura M (1973) The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. Genet. Res. 22: 201-204.

Olivo PD, Van de Walle MJ, Laipis PJ, Hauswirth WW (1983) Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. Nature 306(5941): 400-2.

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet. 29(1): 20-1.

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet. 15(4): 363-8.

Parsons TJ, Irwin JA (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288(5473): 1931.

Pena SD, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B, Oyunsuren T (1995) A major founder Y-chromosome haplotype in Amerindians. Nat Genet. 11(1): 15-6.

Permutt MA, Chiu KC, Tanizawa Y (1992) Glucokinase and NIDDM. A candidate gene that paid off. Diabetes 41: 1367-1372.

Poulton J, Brown MS, Cooper A, Marchington DR, Phillips DI (1998) A common mitochondrial DNA variant is associated with insulin resistance in adult life. Diabetologia 41(1): 54-8.

Poulton J, Marchington DR (2002) Segregation of mitochondrial DNA (mtDNA) in human oocytes and in animal models of mtDNA disease: clinical implications. Reproduction 123(6): 751-5.

Reardon W, Ross RJ, Sweeney MG, Luxon LM, Pembrey ME, Harding AE, Trembath RC (1992) Diabetes mellitus associated with a pathogenic point mutation in mitochondrial DNA. Lancet 340(8832): 1376-9.

Redd AJ, Agellon AB, Kearney VA, Karafet T, de Knijff P, Park H, Butler JM, Hammer MF. Forensic value of fourteen novel STRs on the human Y chromosome. Submitted to Forensic Science International.

Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, Cann RL (1999) mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. Am J Hum Genet 65(2): 519-30.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9(3): 552-69.

Rozas J and Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15: 174-175.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-91.

Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Paabo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci U S A 93(21): 12035-9.

Sajantila A, Lukka M, Syvanen AC (1999) Experimentally observed germline mutations at human micro- and minisatellite loci. Eur J Hum Genet. 7(2): 263-6.

Saker PJ, Hattersley AT, Barrow B, Hammersley MS, McLellan JA, Lo YM, Olds RJ, Gillmer MD, Holman RR, Turner RC (1996) High prevalence of a missense mutation of the glucokinase gene in gestational diabetic patients due to a founder-effect in a local population. Diabetologia 39(11): 1325-8.

Santos M, Ward RH, Barrantes R (1994) mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. Hum Biol 66(6): 963-77.

Sathananthan AH, Ratnam SS, Ng SC, Tarin JJ, Gianaroli L, Trounson A (1996) The sperm centriole: its inheritance, replication and perpetuation in early human embryos. Hum Reprod 11(2): 345-56.

Schneider, S.,Roessli, D., and Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46(3): 613-23.

SeqLab (2000) Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisc.

Shigemoto M, Yoshimasa Y, Yamamoto Y, Hayashi T, Suga J, Inoue G, Okamoto M, Jingami H, Tsuda K, Yamamoto T, Yagura T, Oishi M, Tsujii S, Kuzuya H, Nakao K (1998) Clinical manifestations due to a point mutation of the mitochondrial tRNAleu(UUR) gene in five families with diabetes mellitus. Intern Med 37(3): 265-72.

Shitara H, Hayashi JI, Takahama S, Kaneda H, Yonekawa H
(1998) Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. Genetics 148(2): 851-7.

Shoffner JM, Lott MT, Lezza AM, Seibel P, Ballinger SW, Wallace DC (1990) Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA(Lys) mutation. Cell 61(6): 931-7.

Shoffner JM (1996) Maternal inheritance and the evaluation of oxidative phosphorylation diseases. Lancet 348(9037): 1283-8.

Shuster RC, Rubenstein AJ, Wallace DC (1988) Mitochondrial DNA in anucleate human blood cells. Biochem Biophys Res Commun 155(3): 1360-5.

Sigurgardottir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. Am J Hum Genet. 66(5): 1599-609.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129(2): 555-62.

Smith LC, Alcivar AA (1993) Cytoplasmic inheritance and its effects on development and performance. J Reprod Fertil Suppl 48: 31-43.

Sokal RR and Rohlf FJ (1981) Biometry: The Principles And Practice Of Statistics In Biological Research 2cd Edition. WH Freeman Publishing. San Francisco, CA.

Soodyall H, Jenkins T, Mukherjee A, du Toit E, Roberts DF, Stoneking M (1997) The founding mitochondrial DNA lineages of Tristan da Cunha Islanders. J Phys Anthropol. 104(2): 157-66.

Spuhler JN (1979) The First Americas: Origins, Affinities, and Adaptations, eds. Laughlin WS and Harper AL. Fischer, New York.

Stone AC, Stoneking M (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. Am J Hum Genet 62(5): 1153-70.

Sudoyo H, Marzuki S, Byrne E, Mastaglia F (1993) Phenotypic expression of mtDNA heteroplasmy in the skeletal muscle of patients with oculomyopathy: defect in mitochondrial protein synthesis. J Neurol Sci 117(1-2): 83-91.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (1999) Ubiquitin tag for sperm mitochondria. Nature 402(6760): 371-2.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. Biol Reprod 63(2): 582-90.

Suzuki S, Hinokio Y, Hirai S, Onoda M, Matsumoto M, Ohtomo M, Kawasaki H, Satoh Y, Akai H, Abe K, et al. (1994) Diabetes with mitochondrial gene tRNALYS mutation. Diabetes Care 17(12): 1428-32.

Suzuki Y, Tsukuda K, Atsumi Y, Goto Y, Hosokawa K, Asahina T, Nonaka I, Matsuoka K, Oka Y (1996) Clinical picture of a case of diabetes with mitochondrial tRNA mutation at position 3271. Diabetes Care 19(11): 1304-5.

Suzuki Y, Suzuki S, Hinokio Y, Chiba M, Atsumi Y, Hosokawa K, Shimada A, Asahina T, Matsuoka K (1997) Diabetes associated with a novel 3264 mitochondrial tRNA(Leu)(UUR) mutation. Diabetes Care 20(7): 1138-40.

Swofford, DL (2000) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Szuhai K, Ouweland J, Dirks R, Lemaitre M, Truffert J, Janssen G, Tanke H, Holme E, Maassen J, Raap A (2001) Simultaneous A8344G heteroplasmy and mitochondrial DNA copy number quantification in myoclonus epilepsy and ragged-red fibers (MERRF) syndrome by a multiplex molecular beacon based real-time fluorescence PCR. Nucleic Acids Res 29(3): E13.

't Hart LM, Jansen JJ, Lemkes HH, de Knijff P, Maassen JA (1996) Heteroplasmy levels of a mitochondrial gene mutation associated with diabetes mellitus decrease in leucocyte DNA upon aging. Hum Mutat 7(3): 193-7.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3): 585-95.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10(3): 512-26.

Tang Y, Schon EA, Wilichowski E, Vazquez-Memije ME, Davidson E, King MP (2000) Rearrangements of human mitochondrial DNA (mtDNA): new insights into the regulation of mtDNA copy number and gene expression. Mol Biol Cell 11(4): 1471-85.

Thomas AW, Edward A, Sherratt EJ, Majid A, Gagg J, Alcolado JC (1996) Molecular scanning of candidate mitcohdrial tRNA genes in type 2 (non-insulin dependent) diabetes mellitus. J Med Genet 33: 253-255.

Thyagarajan B, Padua RA, Campbell C (1996) Mammalian mitochondria possess homologous DNA recombination activity. J Biol Chem 271(44): 27536-43.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53(3): 563-90.

Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci U S A 93(1): 196-200.

van den Ouweland JM, Lemkes HH, Ruitenbeek W, Sandkuijl LA, de Vijlder MF, Struyvenberg PA, van de Kamp JJ, Maassen JA (1992) Mutation in mitochondrial tRNA(Leu)(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness. Nat Genet 1(5): 368-71.

Vialettes B, Paquis-Flucklinger V, Bendahan D (1997) Clinical aspects of mitochondrial diabetes. Diabetes Metab 23 Suppl 2: 52-6.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253(5027): 1503-7.

Vionnet N, Stoffel M, Takeda J, Yasuda K, Bell GI, Zouali H, Lesage S, Velho G, Iris F, Passa P, et al. (1992) Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. Nature 356: 721-722.

Wainscoat JS, Hill AV, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, Lynch JR, Falusi AG, Weatherall DJ, et al (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319(6053): 491-3.

Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol. 37(6): 613-23.

Wallace DC (1986) Mitotic segregation of mtDNAs in human cell hybrids and the expression of chloramphenicol resistance. Somat Cell Mol Genet 12(1): 41-9.

Wallace DC (1994) Mitochondrial DNA sequence variation in human evolution and disease. Proc Natl Acad Sci U S A 91(19): 8739-46.

Ward RH, Frazier BL, Dew-Jager K, Paabo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci U S A 88(19): 8720-4.

Ward RH, Redd A, Valencia D, Frazier B, Paabo S (1993) Genetic and linguistic differentiation in the Americas. Proc Natl Acad Sci U S A 90(22): 10663-7.

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, and Santos RV (1996) Mitochondrial DNA Polymorphism in Three Brazilian Indian Tribes. American Journal of Human Biology 8: 317-323.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7(2): 256-76.

Watterson GA, Guess HA (1977) Is the most frequent allele the oldest? Theor Popul Biol. 11(2): 141-60.

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, et al. (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc 26: 375-400.

Wright S (1978) Evolution and the Genetics of Populations. Volume 4. Variability within and among Natural Populations. University of Chicago Press. Chicago.

Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. Tree 11: 367-372.

Yoneda M, Chomyn A, Martinuzzi A, Hurko O, Attardi G (1992) Marked replicative advantage of human mtDNA carrying a point mutation that causes the MELAS encephalomyopathy. Proc Natl Acad Sci U S A 89(23): 11164-8.

Zeviani M, Gellera C, Antozzi C, Rimoldi M, Morandi L, Villani F, Tiranti V, DiDonato S (1991) Maternally inherited myopathy and cardiomyopathy: association with mutation in mitochondrial DNA tRNA(Leu)(UUR). Lancet 338(8760): 143-7.

Zuerlein K, Martin JF, Vaughan L, Markow TA (1991) NIDDM: basic research plus education. Lancet 338: 1271.