Characterization of the AZF region of the Y chromosome in Native American haplogroup Q

Authors: Evguenia Alechine, Werner Schempp, Daniel Corach
Submitted: 10. August 2016
Published: 10. August 2016
Volume: 3
Issue: 4
Keywords: AZF, haplogroup Q, male infertility, master thesis, Y chromosome
DOI: 10.17160/josha.3.4.219

MASTER THESIS

UNIVERSITY OF BUENOS AIRES

"CHARACTERIZATION OF THE AZF REGION OF THE Y CHROMOSOME IN NATIVE AMERICAN HAPLOGROUP Q"

M.Sc. Candidate

Biochemist Evguenia Alechine

Director

Prof. Dr. Daniel Corach

Co-Director

Prof. Dr. Werner Schempp

Master Thesis Dissertation for the degree of Master of Science (M.Sc.)
International Master Program in Biomedical Sciences

OCTOBER 2011

IMBS
International Master Program in Biomedical Sciences

ABSTRACT

The Y chromosome is a genomic niche for genes involved in male gamete production. The existence of an *azoospermia factor* (AZF) in its long arm is a key genetic determinant for spermatogenesis since its deletion is associated with infertility. Deletions in the AZFc region are the most frequent known genetic cause of male infertility. This region contains eight gene families involved in spermatogenesis, including *Deleted in Azoospermia* (DAZ) and the *Chromodomain Protein Y* (CDY) genes. AZFc displays significant variation across the male population; nonetheless the phenotypical consequences of some of these variants remain unclear.

Many Y-chromosome geographically differentiated haplogroups have been defined in the human population, however the information available on the Y chromosome sequence in GenBank belongs only to the European haplogroup R. Recent studies have shown that high mutation rates have driven extensive structural polymorphism among human Y chromosomes. Normal males belonging to different Y chromosome haplogroups showed large-scale differences from the reference sequence. In Native American populations the more frequent haplogroup is Q1a3a1 and is still present in the extant Argentinean male population.

The hypothesis of this Master Thesis was that individuals belonging to different Y-chromosome haplogroups might carry characteristic rearrangements on their Y chromosomes. Thus, the aim of this work was to detect microdeletions on the AZFc region of the Y chromosome and to quantify CDY and DAZ gene copy number in fertile men belonging to Native American haplogroup Q1a3a1.

RESUMEN EN ESPAÑOL

El cromosoma Y humano, debido a sus características haploides, a la falta de recombinación y a su herencia uniparental por vía paterna, lleva en su secuencia la historia de su linaje. En reproducción masculina, se identificó la región AZF como responsable de la regulación de la espermatogénesis. Específicamente, la región AZFc abarca el 60% de las microdeleciones descriptas y su deleción completa es la causa genética más frecuente de infertilidad masculina. Sin embargo, las funciones específicas de los genes localizados en la región AZFc aún son poco claras. Genes localizados en esta región, como DAZ (Deleted in azoospermia) y CDY (Chromodomain protein Y), cumplen funciones importantes en la espermatogénesis, sin embargo el efecto de su deleción sobre la fertilidad no es igual en todos los individuos.

Numerosos estudios indican que la presencia de ciertas microdeleciones en la región AZFc está relacionada con infertilidad masculina en determinadas poblaciones, mientras que las mismas deleciones no afectan la fertilidad en otras. Esto podría indicar que existen factores presentes en determinados haplogrupos del cromosoma Y que predisponen o protegen contra la infertilidad en presencia de las mismas microdeleciones.

La hipótesis de la presente Tesis de Maestría es que individuos pertenecientes al haplogrupo Nativo Americano Q1a3a1 podrían presentar polimorfismos estructurales en la región AZF del cromosoma Y sin signos de fertilidad disminuida. El objetivo de la misma es caracterizar el cromosoma Y de individuos de origen Nativo Americano pertenecientes al Hg Q1a3a1 para identificar posibles microdeleciones en la región AZFc y variaciones en el número de copias de los genes de CDY y DAZ, en hombres fértiles.

TABLE OF CONTENTS

ABBREVIATIONS

AFZ: azoospermia factor
BIO: biotin
BrdU: Bromodeoxyuridine
CDY: chromodomain protein Y gene
CE: capillary electrophoresis
CNVs: copy number variants
CODIS: Combined DNA Index System
DAPI: 4',6-diamidino-2-phenylindole
DAZ: deleted in azoospermia gene
DAZL: deleted in azoospermia like gene
DIG: digoxigenin
DNA: deoxyribonucleic acid
DTT: Dithiothreitol
dNTPs: Deoxyribonucleotides
DOP-PCR: Degenerate Oligonucleotid Primer Polimerase Chain Reaction
EBV: Epstein-Barr Virus
EDTA: Ethylenediaminetetraacetic acid
F: forward
FISH: fluorescence *in situ* hybridization
FITC: Fluorescein isothiocyanate
Hg: haplogroup
HRM: high resolution melting
ILS: Internal Lane Standard
MSY: male-specific region of the Y chromosome
NAHR: non-allelic homologous recombination
NRY: non-recombining region of the Y chromosome
NTC: non-template control
P: palindrome
PAR: pseudoautosomal region
PCR: polymerase chain reaction
PHA: phytohemaglutinin
PEG: Polyethylene glicol
qPCR: quantitative polymerase chain reaction
R: reverse
RRM: RNA recognition motif
SDS: Sodium dodecyl sulfate
SFVs: sequence family variants
SNPs: single nucleotide polymorphisms
SRY: sex-determining region Y
STRs: short tandem repeats
STS: sequence-tagged sites
TEC: Tris (10mM)- EDTA(1mM)- NaCl(100mM)
Tris: 2-Amino-2-hydroxymethyl-propane-1,3-diol
TRITC: rhodamine isothiocyanate
UBA: University of Buenos Aires
WHO: World Health Organization
Y-STSs: Y chromosomal sequence-tagged sites
Yq: long arm of the Y chromosome

“Reports of the demise of the Y chromosome and an impending extinction of men may have been exaggerated. The Y's full genome sequence reveals that we have underestimated its powers of self-preservation.

Instead of doubling up to protect its genetic cargo like other chromosomes, the lone Y safeguards its genes by having sex with itself, an international consortium has found.

"We're on a quest to bring respectability to the Y chromosome," says geneticist David Page of the Massachusetts Institute of Technology, leader of the sequencing team. The male-defining chromosome was previously thought of as a wasteland where genes go to die.

The Y's defenses are double-edged, however, sometimes leading to infertility. The sequence should help us to diagnose and treat such genetic mishaps.

Human chromosome pairs swap genes to minimize bad mutations. Y, which has no partner, faces being whittled away by mutation. Some estimate that the chromosome could be complete junk in about ten million years.

The finished sequence shows that the chromosome fights entropy with palindromes. About six million of its 50 million DNA letters reside in sequences that read the same, in opposite directions, on both strands of the double helix. The longest is nearly three million letters long. "The Y chromosome is a hall of mirrors," says Page.

These palindromes house many genes - which means that there is a copy at each end of the palindromic sequence. These provide back-ups should harmful mutations arise. The mirror-image structure also allows the arms to swap position when DNA divides. Genes are shuffled and bad copies are purged.

Page's team has calculated the amount of swapping needed in each generation to produce the near-perfect palindromes of the human Y. They estimate that every man's Y contains 600 DNA letters that differ from his father's. This is thousands of times more than the normal mutation rate.

"No one had contemplated that there would be this level of gene conversion in our own genome," says Huntington Willard of Duke University, Durham, North Carolina. "It gives us a glimpse of how the Y has protected itself."

Other researchers see swapping as an evolutionary accident, not a safeguard. "It's a daring suggestion, but I find it a bit difficult to believe," says geneticist Mark Jobling of the University of Leicester, UK.

Jobling is sceptical because the trick has a high cost: good genes are just as liable to be lost as bad. This is a major cause of male infertility, as most of the genes within the palindromes control testes development. One in every few thousand men is infertile because key genes have been deleted.

Genetic testing is already used to diagnose male infertility. A fuller understanding of the Y's make-up will help refine these tests, and improve doctors' advice to couples. "We have a greater knowledge of where the Y tends to break," says Page. "Testing needs to be updated to reflect our better understanding from the finished sequence."

The palindromes, and other forms of repeated DNA, made the Y chromosome very tricky to sequence. So the finished sequence comes from just one man's Y. Getting more sequences is essential, says Jobling, as the chromosome's structure, and hence biology, varies greatly around the world.

"We have a beautiful snapshot of the Y chromosome," he says. "Now we need to look in other lineages to build up a photo album of its diversity."

John Whitfield
Nature News, 19 June 2003

1. INTRODUCTION

1.1. THE Y CHROMOSOME

Just a few years ago, Jobling and Tyler-Smith said, "*The Y chromosome seemed to fulfill the role of juvenile delinquent among human chromosomes – rich in junk, poor in useful attributes, reluctant to socialize with its neighbors and with an inescapable tendency to degenerate*". Nowadays, we know that although the Y chromosome is not essential for life (males have it, but females survive perfectly without it), it is responsible for two important specific roles: sex determination and male ability to leave offspring.

Conversely to the rest of the human chromosomes, the hallmark of the Y chromosome is that it does not recombine with its partner, the X chromosome. This is true with the exception of the pseudoautosomal regions PAR1 and PAR2 located on the distal ends of both sexual chromosomes. The recombination between PAR1 and PAR2 is essential for the proper segregation of the sexual chromosomes during male meiosis. Thus, by avoiding recombination, the Y chromosome is transmitted from father to son and its haplotypes pass intact from generation to generation establishing a paternal lineage (Jobling and Tyler-Smith 2003).

The part of the Y chromosome that does not suffer recombination was named the non-recombining region of the Y chromosome (NRY). The NRY is built of three heterochromatic blocks (the centromeric region, the proximal Yq11.22 and the distal Yq12) and of three classes of euchromatic sequences. The heterochromatin encompasses at least six distinct sequence species forming long homogeneous tandem arrays and there is no evidence of transcription of these sequences. Within the 47 major branches of the Y genealogy, the distal Yq heterochromatin ranged in length from 29% to 54% of the metaphase Y chromosome (Repping, van Daalen et al. 2006). On the other hand, the euchromatic portion of the Y carries three classes of sequences, namely X-transposed, X-degenerate and ampliconic sequences. The X-transposed sequences exhibit 99% identity to the X chromosome. The X-degenerate sequences are remnants of ancient autosomes from which the modern X and Y chromosomes evolved. The ampliconic class includes large regions (about 30% of the euchromatin) where sequence pairs show more than 99.9% identity (Figure 1.1) (Skaletsky, Kuroda-Kawaguchi et al. 2003; Repping, van Daalen et al. 2006).

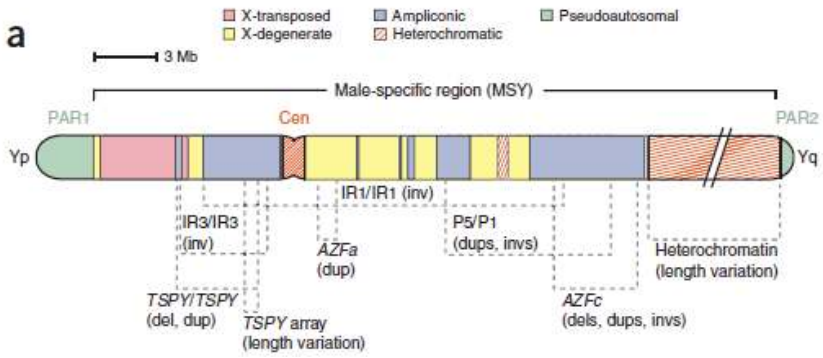


Figure 1.1. *The structure of the Y chromosome (Repping, van Daalen et al. 2006)*

The most prominent features of the Y chromosome, are the eight massive palindromes, most of them containing testis specific genes (Figure 1.2). The high level of similarity and the "mirror-like" palindromic structure makes this region prone to frequent intrachromosomal recombination and gene conversion events. Once the Y chromosome sequence was finished, the non-recombining region of the Y chromosome (NRY) was renamed to male-specific region of the Y chromosome (MSY) because it was evident that the denomination NRY was not reflecting the dynamic evolutionary reshuffling that this part of the genome was suffering (Rozen, Skaletsky et al. 2003; Skaletsky, Kuroda-Kawaguchi et al. 2003; Repping, van Daalen et al. 2006). As far as the Y chromosome does not recombine with a homologue partner, it will continue changing only by mutation, rather than more complex reshuffling that other chromosomes experience, and so will conserve a record of its history. The previously mentioned makes this part of the genome a useful tool for tracing human evolution (Jobling and Tyler-Smith 2003).

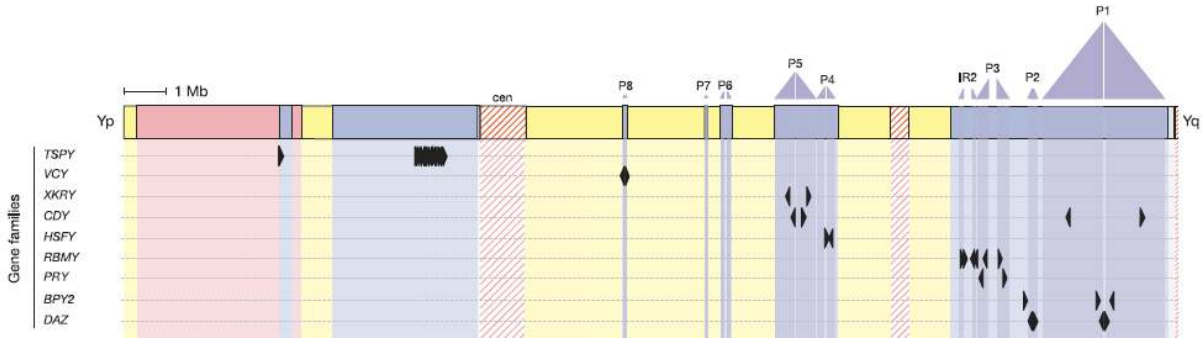


Figure 1.2. *Palindromic structure of the human Y chromosome (Skaletsky, Kuroda-Kawaguchi et al. 2003)*

Another characteristic of the Y chromosome is that, in the population as a whole, its effective population size is one quarter of any autosome making it more susceptible to genetic drift (Jobling and Tyler-Smith 2003). Therefore, particular populations carry their own characteristic lineages due

to the powerful influence of genetic drift on the Y chromosome, driven by its low effective population size (Jobling 2008).

In marked contrast to the functional specialization of the Y chromosome itself, is the eclectic nature of the research fields that encompass the study of Y diversity, including molecular reproductive genetics, genome dynamics as well as population and forensics genetics (Jobling and Tyler-Smith 2003).

One of the roles that was first associated with a particular region of the Y chromosome was the SRY (Sex-determining region Y) that is a sex-determining gene in the therians (placental mammals and marsupials). In mammals, the Y chromosome induces testis formation and thus male sexual development; in the absence of a Y chromosome, gonads differentiate into ovaries and female development ensues. Molecular genetic studies have identified the Y-located testis determining gene SRY as well as autosomal and X-linked genes necessary for gonadal development. The phenotypes resulting from mutation of these genes, together with their patterns of expression, provide the basis for establishing a hierarchy of genes and their interactions in the mammalian sex determination pathway. SRY is an intronless gene that encodes a transcription factor and is a member of the SOX (SRY-like box) gene family of DNA-binding proteins.

The existence of an *azoospermia factor* (AZF) in the long arm of the Y (Yq) has been detected already more than 30 years ago, since its deletion was associated with male infertility (Tiepolo and Zuffardi 1976). Due to the structural complexity of the Y chromosome, the characterization of the AZF region took almost 30 years. First, the sequencing of the AZFa and AZFc regions gave a glance to its structure (Kuroda-Kawaguchi, Skaletsky et al. 2001). Later on, the complete sequencing of the MSY allowed a better understanding of this male-specific portion of the genome and triggered the subsequent molecular investigations in the matter of Y chromosome male infertility (Skaletsky, Kuroda-Kawaguchi et al. 2003).

Large-scale molecular screening for Y chromosome microdeletions (i.e. those not identifiable via conventional cytogenetic techniques) revealed that such determinants displayed a tripartite organization. Thus, AZFa, AZFb and AZFc were established as the Y chromosome regions regulating spermatogenesis (Figure 1.3). Subsequent DNA sequencing approaches demonstrated that these regions harbor a total of 13 different genes/gene families. From the main genes described, USP9Y and DDX3Y mapped to AZFa, CYorf5, RPS4Y2, EIF1aY, KDM5D, XKRY, HSFY, PRY, RBMY1A1 and CDY2 mapped to AZFb, and BPY2, CDY1 and DAZ mapped to AZFc (Navarro-Costa, Goncalves et al. 2010; Navarro-Costa, Plancha et al. 2010).

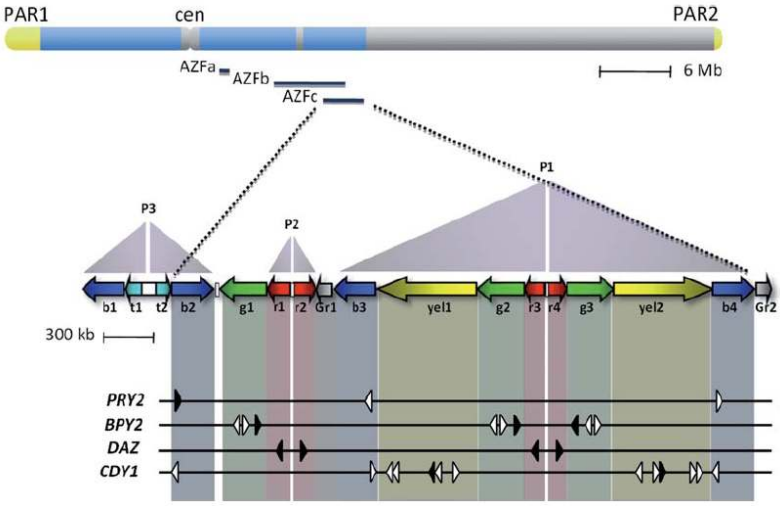


Figure 1.3. *Location of AZFa, b and c regions on the Y chromosome (Navarro-Costa, Goncalves et al. 2010)*

Due to its unusual genomic landscape and its dynamic structure, the Y chromosome is rich in copy number variants (CNVs). Because of its haploidy, the Y chromosome is unconstrained by pairing along its entire length, and this has led to the accumulation of high proportion of segmental duplications that can act as substrates for the generation of CNVs through a mechanism called non-allelic homologous recombination (NAHR)(Figure 1.4). As the Y chromosome sequence is composed mainly of large repeated-sequence blocks called "amplicons" organized into palindromic structures showing high sequence identity between the arms, such structures may undergo frequent inversion and gene conversion events, as well as duplications and deletions. A systematic survey has identified regions rich in CNVs in the TSPY1 array and in the AZFc region (Fernandes, Paracchini et al. 2004; Balaresque, Bowden et al. 2008; Jobling 2008)(Figure 1.5).

With the determination of the complete nucleotide sequence of AZFc by identifying and distinguishing between near-identical amplicons, six distinct families of amplicons have been described in this region. A color code was assigned to the amplicons for further studies (Figures 1.3 and 1.4). The amplicon units ranged in length from 115kb (gray) to 678kb (yellow). The turquoise, gray and yellow amplicons each occur twice in the region sequenced, whereas the green amplicon occurs three times and the blue and red amplicons each occur four times. Together, the six amplicon families account for 93% of the 4.5-Mb sequence (Kuroda-Kawaguchi, Skaletsky et al. 2001).

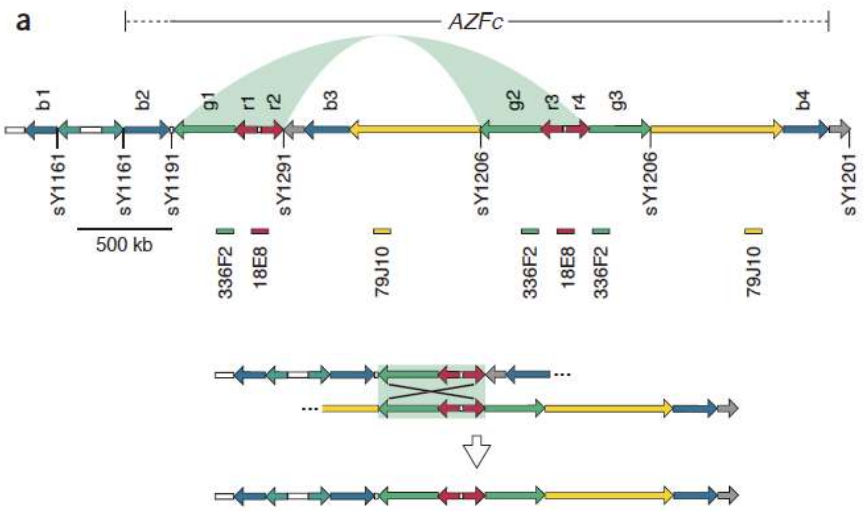


Figure 1.4. *Model of the homologous recombination mechanism in the AZFc region (Repping, Skaletsky et al. 2003)*

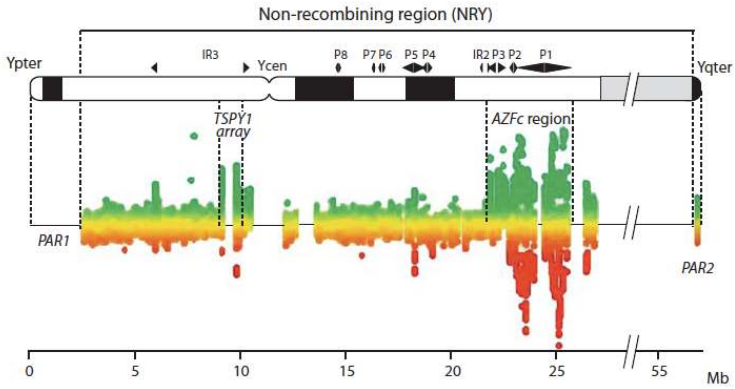


Figure 1.5*. Variation hotspots on the MSY (Jobling 2008)*

1.2. Y CHROMOSOME HAPLOGROUPS

Using binary polymorphisms with low mutation rates, such as single nucleotide polymorphisms (SNPs), a unique phylogeny can therefore easily be constructed (Jobling 2008). A chromosomal haplogroup refers to a group of chromosomes sharing a similar combination of binary states at multiple loci (Jobling and Tyler-Smith 2003). Many Y chromosome haplogroups have been defined in the human population that nowadays arises to 311 geographically differentiated Y haplogroups. The constant increase in number of these haplogroups demanded a unified nomenclature system for its use in further studies. Figure 1.6 depicts the old and new nomenclature systems used to designate Y chromosome haplogroups (YCC 2002; Karafet, Mendez et al. 2008).

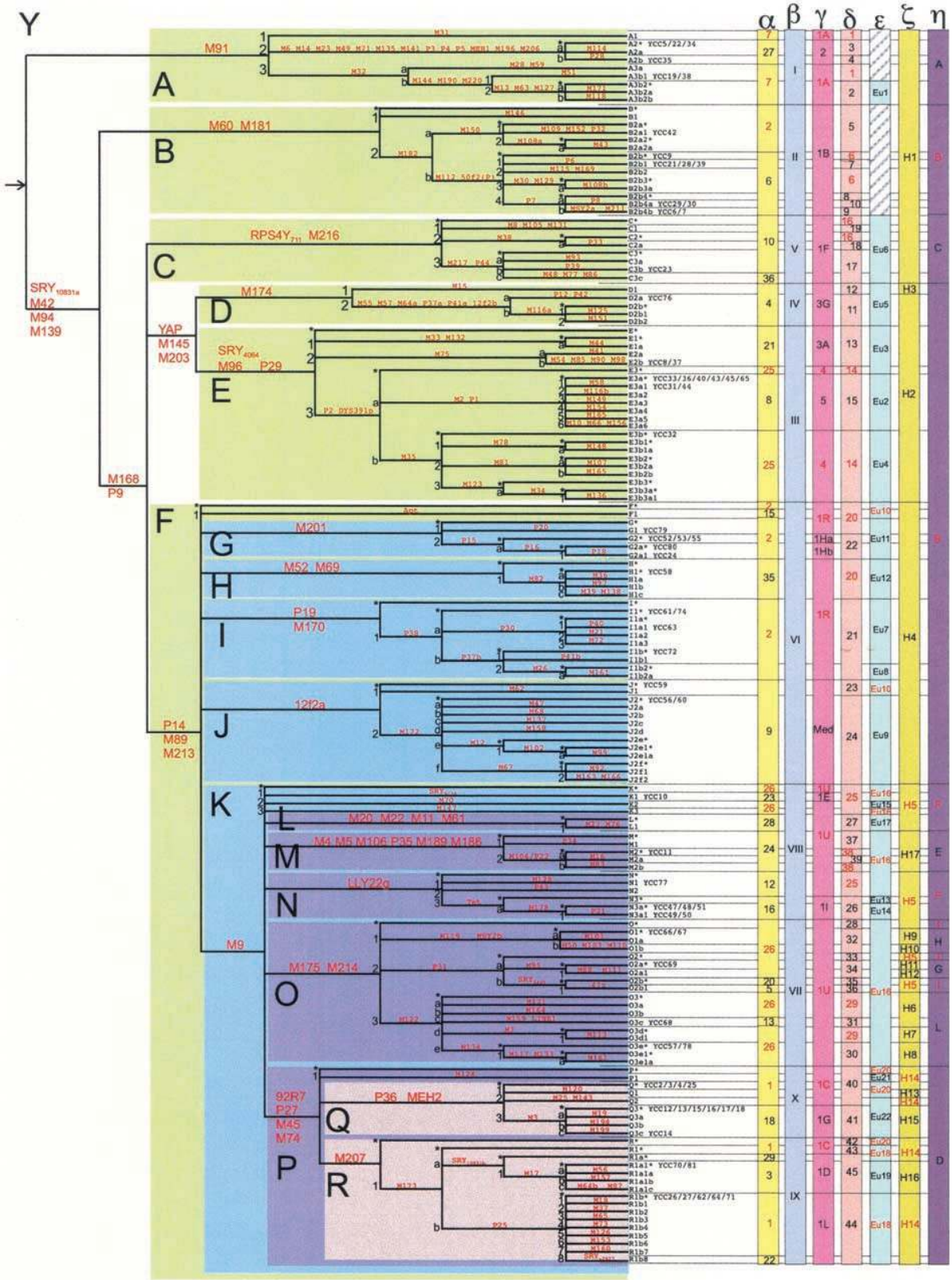


Figure 1.6. *The Y chromosome haplogroup tree (YCC 2002)*

An important remark is that the information available on the Y chromosome sequence in GenBank belongs to one single man's Y chromosome from the European haplogroup R (Jobling and Tyler-Smith 2003; Skaletsky, Kuroda-Kawaguchi et al. 2003). Haplogroup R is believed to have arisen approximately 27,000 years ago in Asia. The two currently defined subclades are R1 and R2. Sub-haplogroup R1b1a2 (M269C) is observed most frequently in Europe, especially Western Europe, but with notable frequency in southwest Asia. R1b1a2 is estimated to have arisen approximately 4,000 to 8,000 years ago in southwest Asia and to have spread into Europe from there (ISOGG 2011). In Argentina, haplogroup R1b1a2 was described as the more prevalent among the urban male population (Corach et al. 2010).

As we can appreciate from Figure 1.7, in the American continent, by far, the more prevalent is haplogroup Q1a3a1, indicated in the Figure as Q, and it is thought that it came from Asia and entered the continent through the Bering Strait (ISOGG 2011).

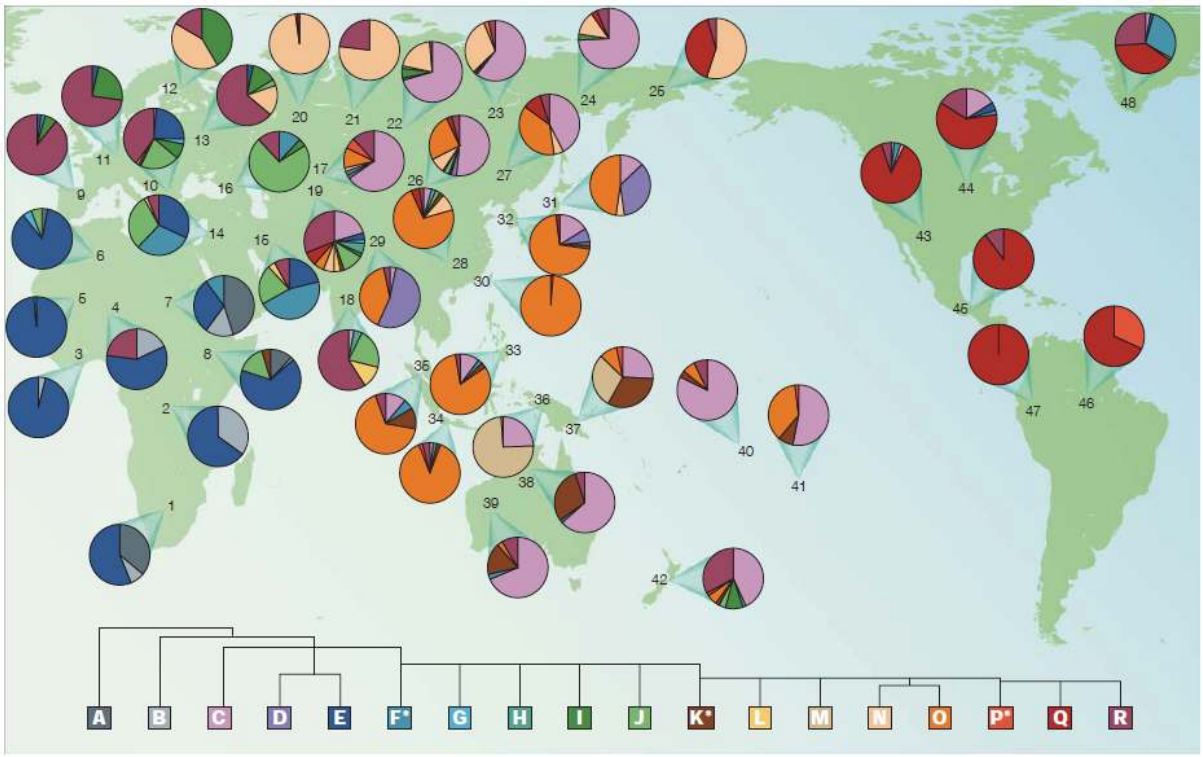


Figure 1.7. *Geographical distribution of Y chromosome haplogroups (Jobling and Tyler-Smith 2003)*

Nowadays, there are several methods to characterize Y chromosome haplogroups. Among others, a novel approach to perform a rapid screening of specific haplogroup-characterizing SNPs has been described. By means of Real Time PCR amplification and subsequent High Resolution Melting (HRM) analysis, it is possible to assign a peak to each SNP allele and therefore identify the haplogroup of each sample (Zuccarelli, Alechine et al. 2011).

1.3. MALE INFERTILITY

Infertility is the inability to conceive a child. A couple may be considered infertile if, after two years of regular sexual intercourse, without contraception, the woman has not become pregnant (and there is no other reason, such as breastfeeding or postpartum amenorrhea). Primary infertility is infertility in a couple who have never had a child. Secondary infertility is failure to conceive following a previous pregnancy. Infertility may be caused by infection in the man or woman, but often there is no obvious underlying cause. Male factor infertility is the sole cause of infertility in approximately 20% of infertile couples, and in 30% to 40% of couples both male and female factors contribute. Thus half of all infertility can be attributed in part or completely to the male factor (WHO).

Male infertility can be classified by quantitative abnormalities (i.e. azoospermia, cryptozoospermia or oligozoospermia) or by qualitative abnormalities (i.e. asthenospermia, teratozoospermia or necrospermia) (Hargreave 2000; Ferlin, Arredi et al. 2007). The etiology of male infertility could be due to disorders related to motility or sperm function, disorders related to obstructive lesions, disorders related to spermatogenic failure, sexual dysfunction disorders of erection and ejaculation or endocrine dysfunction. Both environmental and genetic factors, combined or separate, should be considered (Vogt 2005). About 50% of the cases of male infertility are classified as "idiopathic", evidencing no previous history of fertility problems and with normal findings on their physical examination. The majority of these cases find their etiology in genetic abnormalities (Vogt 2004).

Genetic causes account for 10-15% of severe male infertility. These genetic abnormalities include chromosome translocations, aneuploidies, Y chromosome microdeletions and androgenic receptor gene mutations. Thus, genetic causes are either detected at the cytogenetic and/or at the molecular levels. In addition to gene mutations and polymorphisms, damage to the chromatin and epigenetic abnormalities also may contribute to male infertility (Vogt, Edelmann et al. 1996; Ferlin, Arredi et al. 2006).

Regarding the Y chromosome, both large rearrangements and microdeletions have been accused to play a significant role in male infertility. For diagnostic identification of these Y chromosome variations, it is possible to screen the Y chromosome for large rearrangements by FISH and microdeletions by plus/minus sequence-tagged site (STS) markers on the Y chromosome, as it has been widely used and described (Repping, Skaletsky et al. 2003; Repping, van Daalen et al. 2004; Repping, van Daalen et al. 2006).

The role of the genes on the Y chromosome in male infertility has been widely studied and the AZF complete deletion was associated with male infertility (Tiepolo and Zuffardi 1976). As previously discussed, AZFc deletions account for approximately 60% of all recorded AZF deletions and, therefore, the genetic disruption of AZFc is generally perceived as a key concern for male reproductive health (Navarro-Costa, Goncalves et al. 2010). From the most frequently found

deletions on AZFc (Figure 1.8), the gr/gr deletion (sY1291 deletion) and b2/b3 or g1/g3 deletion (sY1191 deletion) have been identified in men with impaired as well as normal espermatogenesis and these deletions have been found to predispose to spermatogenic failure in some populations but not in others (Repping, van Daalen et al. 2004; Yang, Ma et al. 2010).

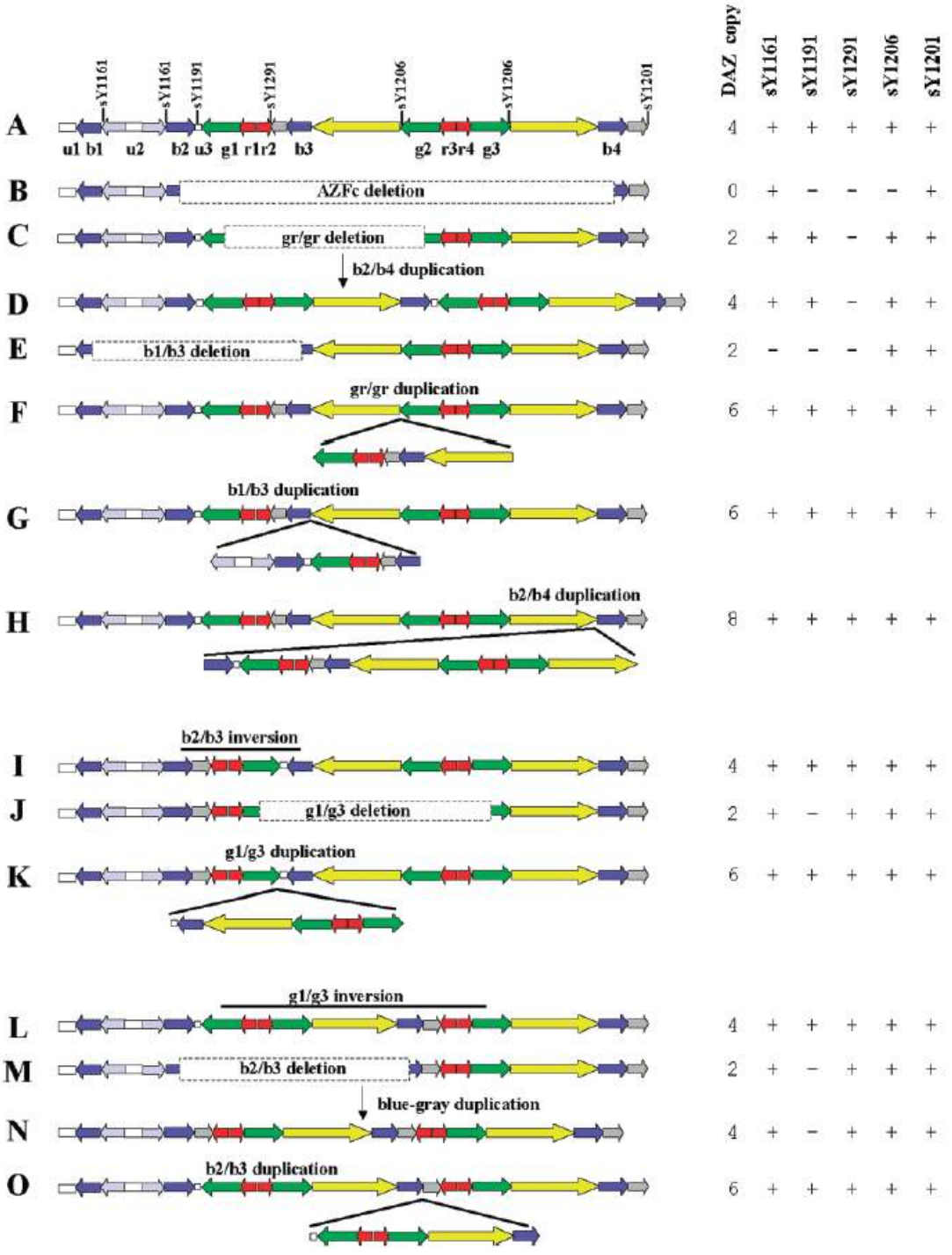


Figure 1.8. *Most frequently described AZFc deletions (Lin, Hsu et al. 2006)*

In clinical terms, the importance of the AZF regions is paramount as deletions in these domains are one of the most frequent genetic causes of spermatogenic failure. AZFc displays significant variation in terms of genomic architecture across the male population. Yet, the

phenotypical consequences for spermatogenesis of some of these variants are unclear (Navarro-Costa, Goncalves et al. 2010). An association study established that the gr/gr deletion, a 1.6-Mb deletion, is a significant risk factor for spermatogenic failure. However, as the gr/gr deletion has lower penetrance than other characterized Y-chromosomal deletions can be therefore transmitted from father to son. By studying the distribution of gr/gr-deleted chromosomes across different branches of the Y genealogical tree, it was suggested that the existence of this deletion as a polymorphism reflects a balance between haploid selection and homologous recombination (Repping, Skaletsky et al. 2003).

1.4. AZF GENES: DAZ & CDY

The AZFc region contains eight gene families expressed only in testis tissue, three of which code for proteins, but the importance of each individual gene for fertility is not yet understood (Fernandes, Paracchini et al. 2004).

Deleted in Azoospermia (DAZ) is a member of the DAZ gene family and encodes a RNA binding protein that plays a prominent role in the establishment and maintenance of the male germ line. Its expression is restricted to pre-meiotic germ cells, particularly in spermatogonia, with a cytoplasmatic localization. Four copies of this gene (DAZ1, DAZ2, DAZ3 and DAZ4) are found on chromosome Y within palindromic duplications; one pair of genes is part of the P2 palindrome and the second pair is part of the P1 palindrome (Figure 1.9). Each gene contains a 2.4 kb repeat including a 72-bp exon, called the DAZ repeat; the number of DAZ repeats is variable and there are several variations in the sequence of the DAZ repeat. Each copy of the gene also contains a 10.8 kb region that may be amplified; this region includes five exons that encode an RNA recognition motif (RRM) domain. This gene contains three copies of the 10.8 kb repeat. However, no transcripts containing three copies of the RRM domain have been described; thus the reference sequence for this gene contains only two RRM domains. The DAZ gene has an autosomal homologue, the DAZL gene that is considered that gave origin to the Y chromosomal DAZ gene through transposition, thus the loss of Y-borne DAZ copies may be compensated by DAZL expression (Navarro-Costa et al., 2010; NCBI).

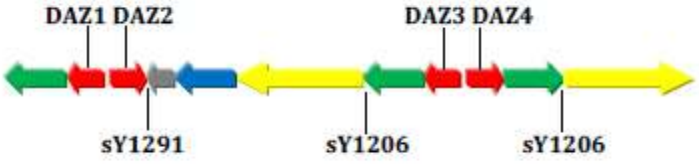


Figure 1.9. *DAZ copies location on the reference Y chromosome (adapted from Navarro-Costa et al., 2010)*

Chromodomain Protein Y (CDY) gene encodes a protein containing a chromodomain and a histone acetyltransferase catalytic domain. Chromodomain proteins are components of heterochromatin-like complexes and can act as gene repressors. This protein is localized to the nucleus of late spermatids where histone hyperacetylation takes place. Histone hyperacetylation is thought to facilitate the transition in which protamines replace histones as the major DNA-packaging protein. The human chromosome Y has two identical copies of this gene within the palindrome P1 (CDY1A/CDY1B) and a pair of closely related genes in the palindrome P5 (CDY2A/CDY2B)(Figure 1.10). Two protein isoforms are encoded by transcript variants of this gene. Additional transcript variants have been described, but their full-length nature has not been determined (Navarro-Costa et al., 2010).

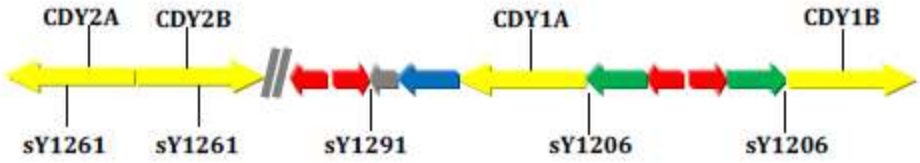


Figure 1.10. *CDY copies location on the reference Y chromosome (adapted from Navarro-Costa et al., 2010)*

To detect variation in gene copy number, several techniques have been described. A recently developed approach uses quantitative Real Time PCR to assess gene copy number. This approach amplifies, in the same PCR run, one reference gene i.e. single copy gene and the gene of interest, both for the unknown sample and a calibrator sample of known copy number, to determine the copy number of the gene of interest (Munch, Kirsch et al. 2008; Noordam, Westerveld et al. 2011; Greve, Alechine et al., submitted).

1.5. PREVIOUS STUDIES

Many association studies between deletions on the AZFc region and spermatogenic failure in different populations have been performed and are described in Table 1.1.

Table 1.1. *Reports on the association of AZFc deletions with male infertility (Krausz, Giachini et al. 2009; Navarro-Costa, Goncalves et al. 2010; Puzuka, Pronina et al. 2011; Shahid, Dhillon et al. 2011)*

Association	Pub. Date	Authors	Population
	2003	Repping et al.	Dutch
	2005	de Llanos et al.	Spanish
	2005	Ferlin et al.	Italian

	2005	Giachini et al.	Italian
	2005	Lynch et al.	Australian
	2007	Navarro-Costa et al.	Portuguese
	2008	Yang et al.	Chinese
	2008	Giachini et al.	Italian
	2011	Shahid et al.	Indian
	2011	Puzuka et al.	Latvian
	2004	Fernandes	European
No association	2004	Machev et al.	French
	2004	Hunchlenbroich et al.	German
	2006	Ravel et al.	Admixed
	2006	Fernando et al.	Sri Lanka
	2006	Zhang et al.	Chinese
	2006	Carvalho et al.	Brazilian
	2006	de Carvalho et al.	Japanese
	2007	Zhang et al.	Chinese
	2007	Wu et al.	Chinese
	2007	Lin et al.	Chinese (in Taiwan)
	2007	Lardone et al.	Chilean
	2007	Imken et al.	Moroccan
	2008	Stouffs et al.	Admixed
	2009	Ravel et al.	Admixed
	2009	Krausz	European

Recent studies have shown that high mutation rates have driven extensive structural polymorphisms among human Y chromosomes. Normal males belonging to different Y chromosome haplogroups showed large-scale differences from the reference sequence in several haplogroups. In particular, haplogroups Q1a3a1 and N showed a large deletion in the AZFc region that apparently does not affect fertility (Repping et al., 2006). On the other hand, in haplogroup N, a common variant of the human Y chromosome lacking the DAZ3/DAZ4 and BPY2.2/BPY2.3 doublets in distal AZFc in men with no signs of reduced fertility was described. These data confirm the expectation that the human Y chromosome sequence and gene complement may differ substantially between individuals and more variations are to be expected in different Y chromosomal haplogroups (Fernandes et al., 2004).

Furthermore, as a common variation in the *AZFc* region, the gr/gr deletion is regarded as a significant risk factor for spermatogenic impairment, whereas the association of the deletion's phenotypic expression with Y-chromosomal background was studied in several haplogroups but is still a subject of debate (Repping, Skaletsky et al. 2003; Fernandes, Paracchini et al. 2004; Repping, van Daalen et al. 2004; Repping, van Daalen et al. 2006; Yang, Ma et al. 2008; Yang, Ma et al. 2010; Shahid, Dhillon et al. 2011). In a previous study regarding haplogroup E, it was found that its particular background seems to be more prone to rearrangements, in particular to b2/b4 microdeletions, than other haplogroups. One would have expected that selection would have acted

to lower its frequency, but this does not seem to be the case (Arredi, Ferlin et al. 2007). The contribution of the deletion to spermatogenic impairment in different Y-chromosomal haplogroups was studied and, the gr/gr deletion was the most common rearrangement found, and its frequency was significantly higher in men with azoospermia or oligozoospermia relative to normozoospermia, while all *DAZ3/DAZ4+CDY1b+BPY2.2 or 2.3* deletions were found only in haplogroup Q1 without any distribution difference between the azoospermic and oligozoospermic and normozoospermic groups. This study provided further evidence for the existence of multiple subtypes of gr/gr deletion and indicates that gr/gr-*DAZ1/DAZ2* deletion is a significant risk factor. However, the association of the phenotypic variation of gr/gr deletion with Y-chromosomal haplogroups is not definite yet, because of the limited amounts of deletions observed in each haplogroup and the lack of quantitative trait such as sperm density analysis. The fact that a common gr/gr copy deletion haplotype was found exclusively in the Y haplogroup Q1, without pathogenic consequences, underscore the importance of haplogrouping and copy number typing (Yang, Ma et al. 2010).

Y chromosomes in apparently normal males also have been described as carrying null and duplicated alleles at the microsatellite DYS448, which lies in the proximal part of the azoospermia factor c (AZFc) region and is important in spermatogenesis. This "ampliconic" repeats act as substrates for non-allelic homologous recombination (NAHR). These rearrangements appear to owe their heterogeneous population frequencies to two major factors. First, having once occurred, a neutral rearrangement can be successfully propagated in a particular population due to its social organization. Second, haplogroups in general are highly geographically differentiated, and some haplogroups appear to be either predisposed to, or protected against, rearrangements (Balaresque, Bowden et al. 2008).

1.6. ARGENTINEAN POPULATION

The most frequent haplogroup exhibited by Native American populations is Q1a3a1. Although there is general agreement that America was first settled from Asia by people who migrated across Beringia, the pattern of migration, its timing, and the place of origin in Asia of the people that migrated to the Americas is unclear. Recently, mitochondrial DNA data proposed a model for the peopling of the New World in which Native American ancestors diverged from the Asian gene pool prior to 40,000 years ago and experienced a gradual population expansion as they moved into Beringia. After a long period of little change in population size in greater Beringia, Native Americans rapidly expanded into the Americas ~15,000 years ago (Cavalli-Sforza and Feldman 2003; Kitchen, Miyamoto et al. 2008)(Figure 1.11). The observation of one Y chromosome haplogroup at high frequencies in native populations was interpreted as possible single founder Native American Y lineage. This putative single founder Native American lineage is characterized by a C>T mutation at locus DYS199 within the P-M45 Y lineage. The derivate state T of M3 is the main characteristic of the

Native American Y haplogroup Q1a3a1 (Underhill, Jin et al. 1996; Bortolini, Salzano et al. 2003; Karafet, Mendez et al. 2008). It is still unknown whether America was settled from a single founder lineage or if the haplogroup Q1a3a1 had higher fitness than other migrating lineages and was the only paternal lineage that survived through thousands of years in this continent.

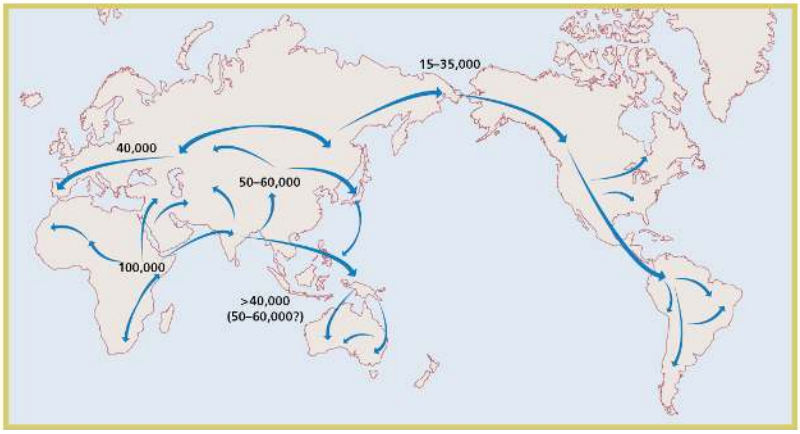


***Figure 1.11**. Human migration map (Cavalli-Sforza & Feldman, 2003)*

Regarding specifically the Argentinean population, the first inhabitants of the Argentinean territory were Native American communities. After the discovery of the Americas in 1492, migratory waves into the nowadays territory of Argentina started in the 16th century with the Spanish colonization. Also, the forced introduction of Africans brought as working force as slaves in the colony between the 17th and 19th centuries has been described. In the 19th century, a second important immigration event took place promoted by the Argentinean government. Most immigrants came from Europe and Minor Asia, however the big majority were Italian. During the era of rapid growth in the late 19th century (from 1880 onwards up to 1930) and the first half of the 20th century, another important urban immigration happened, mainly coming from Europe. Finally, in the late 20th century and early 21st century, economic migrants arrived from Sub Saharan Africa, Korea, China, Central America, Bolivia, Peru, Paraguay, Colombia and from Central and Eastern Europe. All this migratory events had a important impact on the structure of the Argentinean population, mainly due to the admixture with the Native American population.

The nowadays Argentinean population, although influenced by several admixture events, still maintains its Native American genetic contribution. Recent studies have shown that, regarding the Y chromosome lineages, the population of Argentina has 45% of males belonging to haplogroup R1b1a2, 8% to haplogroup I (both European) and 5% to Native American haplogroup Q1a3a1 (Corach et al. 2010)(Figure 1.12).

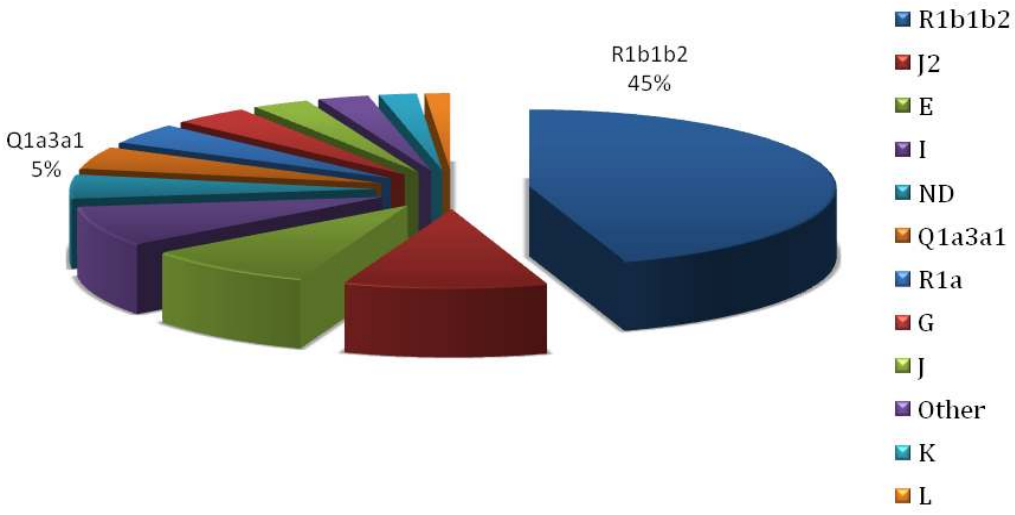


Figure 1.12. *Y chromosome haplogroup distribution in the Argentinean population (Corach et al. 2010)*

With newly developed genetic tools, by Real Time PCR, it is possible to screen the urban population looking for the individuals belonging to haplogroup Q1a3a1, as well as haplogroup I and the GenBank`s reference haplogroup R (Zuccarelli, Alechine et al. 2011).

Until now, the Argentinean population has not been studied for association between rearrangements on the AZF region and Y chromosome haplogroups, neither the haplogroup Q1a3a1, defined by M3 marker, has been deeply studied for this matter so far (Repping, Skaletsky et al. 2003; Fernandes, Paracchini et al. 2004; Repping, van Daalen et al. 2004; Repping, van Daalen et al. 2006).

2. HYPOTHESIS & AIMS

HYPOTHESIS

Individuals belonging to Native American haplogroup Q1a3a1 might have suffered rearrangements on the AZF region of the Y chromosome and therefore carry characteristic polymorphisms that differ from the reference sequence not affecting normal male fertility.

AIM

Characterize fertile men belonging to Native American haplogroup Q1a3a1 regarding AZFc microdeletions and CDY and DAZ gene copy number.

OBJECTIVES

- ✓ Sample collection from men involved in paternity testing as alleged fathers
- ✓ Paternity confirmation by autosomal STR analysis
- ✓ FISH for DAZ and CDY on the sample to be used as calibrator for gene dosage assessment
- ✓ Development of a Real Time PCR quantification tool for DAZ and CDY copy number assessment
- ✓ Y chromosome haplogroup screening (Q1a3a1, R1b1a2 or I) by multiplex Real Time PCR and High Resolution Melting analysis
- ✓ Y-STSs analysis by multiplex PCR and Capillary Electrophoresis
- ✓ Copy number analysis of DAZ and CDY genes by Real Time qPCR

3. MATERIALS & METHODS

3.1. STUDIED POPULATION

Samples for this study were selected from men involved as alleged fathers in paternity testing analysis and belonged to the urban population of 10 provinces of Argentina. Namely, the encompassed provinces were: Buenos Aires, Rio Negro, Chubut, Misiones, Corrientes, Salta, Formosa, Chaco, Santa Fe and Neuquén. Samples from fathers belonging to the Native American Pilagá community from Las Lomitas (Formosa) have been also included in this study (Figure 3.1).

Urban population samples have been collected at the Genetic Fingerprinting Service, School of Pharmacy and Biochemistry, UBA, between January–July 2011 for genetic paternity testing analysis. Only samples from individuals with confirmed paternity by autosomal STR markers were used for following analysis.

Figure 3.1. *Map of Argentina showing the sampled provinces*

3.2. SAMPLES

Blood samples from finger puncture or buccal swabs have been collected from the studied population described above. Buccal swabs have been air-dried and blood samples have been spotted onto Whatman 3MM or FTA® paper. All sample donors signed a written consent statement and the samples were treated anonymously for the purpose of this study.

3.3. MATERIALS

3.3.1. CHEMICALS & REAGENTS

Table 3.3.1. *Molecular Biology, Cell Culture and FISH Reagents*

Reagent	Supplier
DNA IQ™ Casework Pro Kit for Maxwell® 16	Promega Corp., Madison, USA
Ultra Pure™ SDS	Invitrogen™
EDTA 0.5M pH=8.0	Promega Corp., Madison, USA
Proteinase K	Promega Corp., Madison, USA
DTT	Promega Corp., Madison, USA
Tris Base	Promega Corp., Madison, USA
ClNa	AnalytiCals Carlo Erba
ClH	UCB
Plexor® HY System	Promega Corp., Madison, USA
PowerPlex® 16 System	Promega Corp., Madison, USA
5X GoTaq® Reaction Buffer	Promega Corp., Madison, USA
$MgCl_2$ 25mM	Promega Corp., Madison, USA
dNTPs 100mM (dATP, dCTP, dGTP, dTTP)	Promega Corp., Madison, USA
GoTaq® DNA Polymerase	Promega Corp., Madison, USA
SYTO® 9	Invitrogen™
QuantiTec SYBR Green PCR Kit	QIAGEN
QIAGEN Multiplex PCR Kit	QIAGEN
Hi-Di™ Formamide	Applied Biosystems™
ILS 600	Promega Corp., Madison, USA
Big Dye® Terminator v1.1 Cycle Sequencing Kit	Applied Biosystems™
ExoSAP-IT®	USB®
Ethanol, Isopropyl alcohol and Methanol	Merck
Trypton	Becton, Dickinson and Company
Colcemid	Biochrom
Acetic Acid	Merck
Formamide	Merck
QIAGEN Plasmid Maxi Kit	QIAGEN
Anti-DIG mouse antibody	Sigma-Aldrich®
Anti-mouse goat antibody-FITC	Sigma-Aldrich®
Avidin-TRITC	Sigma-Aldrich®
DAPI	Invitrogen™
$AcNH_4$	Mallinckrodt®
Human Cot DNA	Invitrogen™
Dextran Sulfate	Sigma-Aldrich®
PEG 8000	Promega Corp., Madison, USA
dUTP-Biotin	Roche Diagnostics
dUTP-Digoxigenin	Roche Diagnostics
DOP-PCR Random Primers	Biomers
Agarose	Sigma-Aldrich®
Iscoves Medium	Biochrom
PHA	Remel
BrdU	Roth
Yeast	Becton, Dickinson and Company
Kanamycin	Serva

3.3.2. DISPOSABLES

Table 3.3.2. *Major disposables used for the study*

Disposable	Supplier
2mL Eppendorf tubes	Axygen®
1.5mL Eppendorf tubes	Axygen®
0.5mL Eppendorf tubes	Axygen®
0.2mL thin-wall 8 strip PCR tubes	Axygen®
0.1mL 4 strip tube for Corbett RG-6000	Labnet International, Inc.
1000µL filtered tips	Axygen®
200µL filtered tips	Axygen®
100µL filtered tips	Axygen®
10µL filtered tips	Axygen®

3.3.3. EQUIPMENT

Table 3.3.3.1. *Equipment from the Genetic Fingerprinting Service, School of Pharmacy and Biochemistry, UBA*

Equipment	Supplier
Maxwell® 16 System	Promega Corp., Madison, USA
VorTemp™	Labnet International, Inc.
Rotor Gene 6000	Corbett Life Science
GeneAmp® PCR System 9600	Perkin Elmer
GeneAmp® PCR System 9700	Applied Biosystems
DNA Thermal Cycler	Perkin Elmer
3100-Avant Genetic Analyzer	Applied Biosystems
Spectrafuge	Labnet International, Inc.
Vortex	Labnet International, Inc.
Mini Centrifuge	Labnet International, Inc.

Table 3.3.3.2. *Equipment from the Institut für Humangenetik, Universitats Klinikum, Albert-Ludwigs University, Freiburg*

Equipment	Supplier
Laminar Flow Cabinet Heraeus	Thermo Scientific
Incubation Shaker	Infors HT
Microcentrifuge	Labnet International, Inc.
PTC 100 Thermal Cycler	MJ Research
NanoDrop	Thermo Scientific

Table 3.3.3.3. *Equipment from the Neuro and Cytogenetic Lab, San Martin University, Buenos Aires, Argentina*

Equipment	Supplier
Ultra Freezer Fh-80v	Presvac
NanoDrop	Thermo Scientific
PRISM-R Cooled Microcentrifuge	Labnet International, Inc.
Novus Electric Oven	ORL
Masson Waterbath	Vicking
BX-60 Olympus IX-71 Microscope	Olympus

3.3.4. PCR PRIMERS & COSMIDS

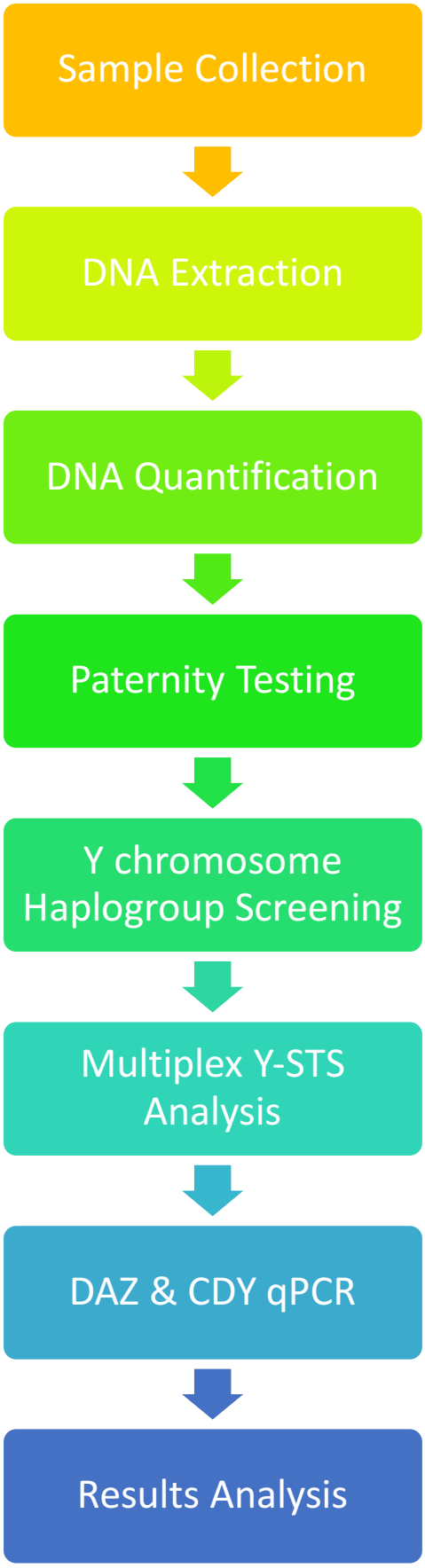
Table 3.3.4.1. *PCR primers´ sequences from 5´→ 3´*

Primer	Sequence	Provider
DYS199/M3 F	CCTGACAATGGGTCACCTCT	IDT
DYS199/M3 R	TTCATTTTAGGTACCAGCTCTTCC	IDT
M269 F	AAGGGGAATGATCAGGGTTT	IDT
M269 R	CCTTCTGAGGCACATATGATAA	IDT
U179 F	GGCTGTAATTCATAGCCTACCA	IDT
U179 R	GTTGCCAGCTCCTCTTTTCA	IDT
DAZ F	TCTGTGCCTGCCTCTCTGTA	IDT
DAZ R	GCCTTATCCTCGGTTTTCCT	IDT
CDY F	TGGAGAGGTTCAGGCACATG	IDT
CDY R	TCCAGCTCTTCACCAGGTTC	IDT
SRY F	GCCACACACTCAAGAATGGA	IDT
SRY R	CCAATGTTACCCGATTGTCC	IDT
sY1191 F	**ACATTTTGCTGCCGGTCA**-CCAGACGTTCTACCCTTTCG	IDT
SY1191 R	GAGCCGAGATCCAGTTACCA	IDT
sY1206 F	**ACATTTTGCTGCCGGTCA**-ATTGATCTCCTTGGTTCCCC	IDT
sY1206 R	GACATGTGTGGCCAATTTGA	IDT
sY1291 F	**ACATTTTGCTGCCGGTCA**-TAAAAGGCAGAACTGCCAGG	IDT
sY1291 R	GGGAGAAAAGTTCTGCAACG	IDT
sY1201 F	**ACATTTTGCTGCCGGTCA**-CCGACTTCCACAATGGCT	IDT
sY1201 R	GGGAGAAAAGTTCTGCAACG	IDT
sY1261 F	**ACATTTTGCTGCCGGTCA-** AAGGAGCTTGCCTCATACAATG	IDT
sY1261 R	TTAGAGCTTGCAAGAAGAGTCTAGTAC	IDT
M13-FAM	**FAM**-TGTAAAACGACGGCCAGT	Operon
DOP-PCR Random Primers (F and R)	CCGACTCGAGNNNNNNATGTGG	Biomers

Table 3.3.4.2. *Cosmids*

Cosmid	Gene	Provider	Publication
6B7	DAZ	Kay Taylor , London	Taylor et al. 1996
2A49	CDY	Pauline Yen, Taiwan	Kühl et al. 2001

3.4. Workflow


Sample Collection
DNA Extraction
DNA Quantification
Paternity Testing
Y chromosome Haplogroup Screening
Multiplex Y-STS Analysis
DAZ & CDY qPCR
Results Analysis

3.5. METHODS

3.5.1. DNA EXTRACTION

DNA from both buccal swabs and blood samples has been extracted by means of a semi-automated platform (Maxwell® 16 System, Promega Corp., Madison, USA) using DNA IQ™ Casework Pro Kit for Maxwell® 16 as described by the manufacturer. Briefly, biological samples on buccal swabs, Whatman 3MM or FTA® paper were incubated over-night in 500µL TEC-SDS 0.5% (10mM Tris, 1M ClH pH=8, 10mM EDTA pH=8, 100mM NaCl) and 10µL protease (Proteinase K, Promega) at 56°C and 750rpm. Afterwards, 250µL of Lysis Buffer with 1mM DTT was added and put into the Maxwell cartridge and extracted by means of magnetic particles that bind DNA. After purification, DNA was eluted in 30µL of Elution Buffer.

3.5.2. DNA QUANTIFICATION

Sample DNA quantification was carried out with Plexor® HY System quantification kit (Promega Corp. Madison, USA) and Rotor-Gene 6000 Real Time PCR equipment (Corbett Life Science, Australia) as described by the manufacturer. Briefly, all reactions were carried out with 10µL Master Mix, 1µL Primer Mix and 2µL of DNA sample in a final volume of 20µL. Standard curve ranged 50-0,016 ng/µL and female, male and Non-template controls (NTC) were included in each run. PCR cycling conditions were 95°C 2 min followed by 35 cycles of 95°C 10 sec – 60°C 20 sec and signal detection was performed on the green, yellow and orange channels. Assessment of autosomal (green) and male (yellow) DNA concentration, and presence of PCR inhibitors (orange) was achieved by interpolation of the sample Ct (Threshold cycle) in the corresponding standard curve. A melting curve analysis was performed after each run to check product specificity (Figure 3.5.2).

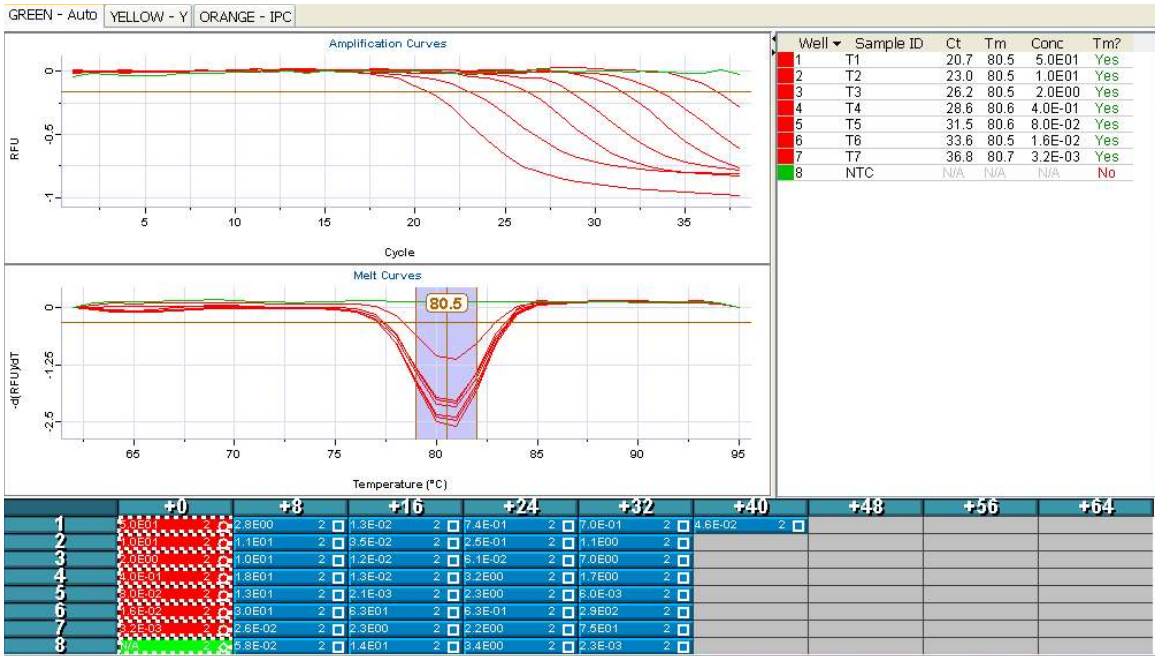


Figure 3.5.2. *Plexor® HY DNA quantification profile.*

3.5.3. PATERNITY TESTING

For paternity assessment, PowerPlex® 16 System (Promega Corp. Madison, USA) was used as described by the manufacturer. PowerPlex® 16 System amplifies 15 autosomal STRs and a sex marker (Amelogenin) and 13 of the 15 markers are those recommended by CODIS (Figure 3.5.3.1). Briefly, all multiplex reactions were carried out with 2µL Master Mix, 1µL Primer Mix and 1µL template DNA (0.5-2 ng/µL) in a final volume of 10µL. PCR cycling conditions were 95°C 1 min, 10 cycles of 94°C 30 sec – 60°C 30 sec (29% ramp) – 70°C 45 sec (23% ramp) and 22 cycles of 90°C 30 sec – 60°C 30 sec (29% ramp) – 70°C 45 sec (23% ramp). Amplification of the samples was carried out in a GeneAmp® PCR System 9600 or 9700. In all cases, samples belonging to the mother, child and alleged father have been amplified for comparison of the genetic profiles (Figure 3.5.3.2). 1µL of the PCR products were prepared and denatured in 9.5µL Hi-Di Formamide (high deionized) and 0.5µL ILS 600 (Internal Lane Standard 600, Promega), as internal standard. Resolution and detection was performed in a 3100-*Avant* Genetic Analyzer (Applied Biosystems) using a 36cm array and POP4 polymer. Capillary electrophoresis conditions were 60°C, 15kV/180 sec (pre-run), 6.5kV/1 sec (injection) and 15kV/1400sec (run). Analysis of the results was performed with GeneMapper ID-X software. In all threesomes, sharing of one allele with the mother and the other allele with the father was analyzed for 15 STR markers and the paternity probability was higher than 99.99% (Figure 3.5.3.3).

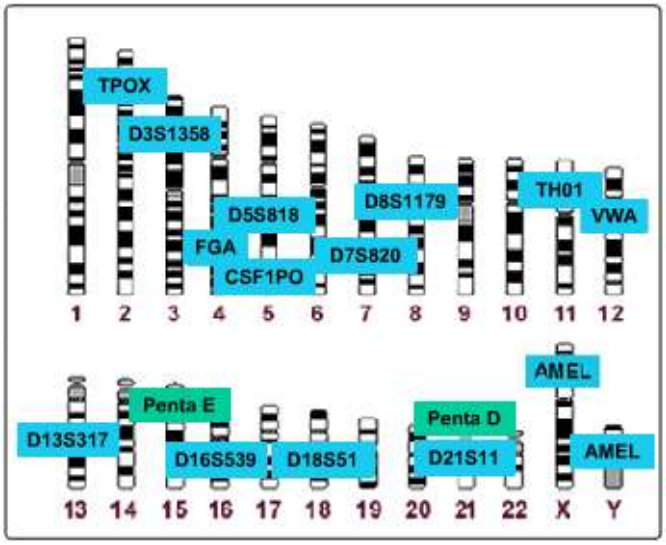


Figure 3.5.3.1. *Autosomal STR markers used for paternity testing*

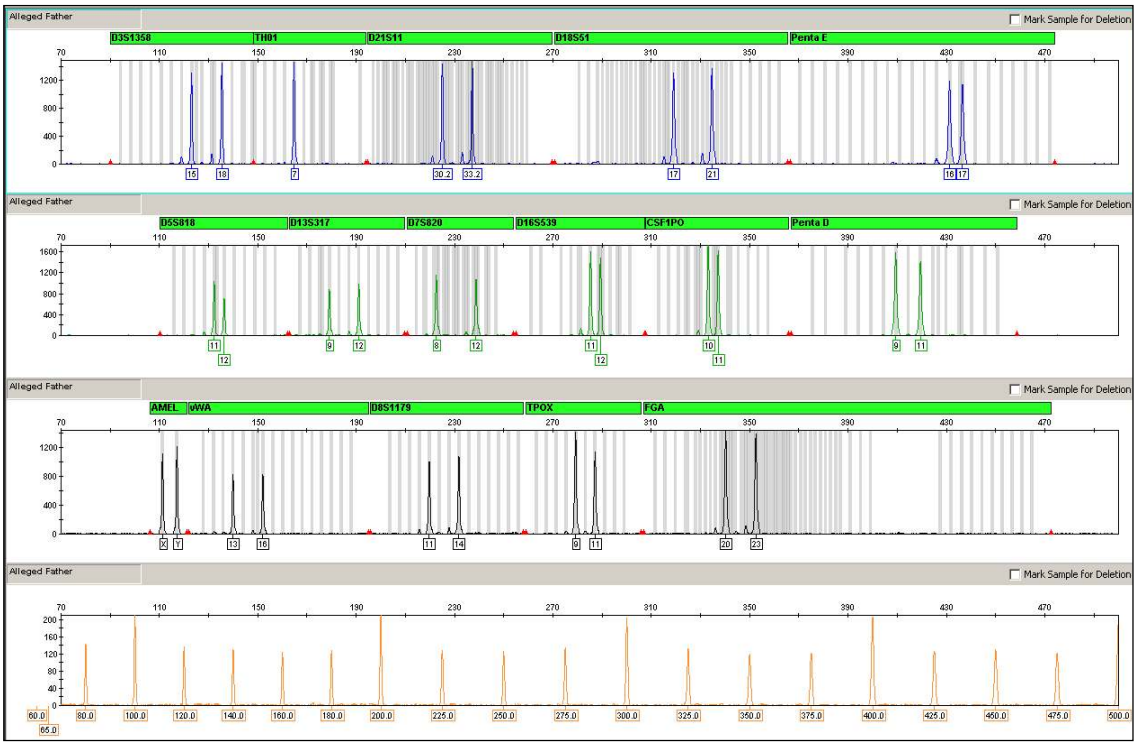


Figure 3.5.3.2. *PowerPlex® 16 profile for a male sample*

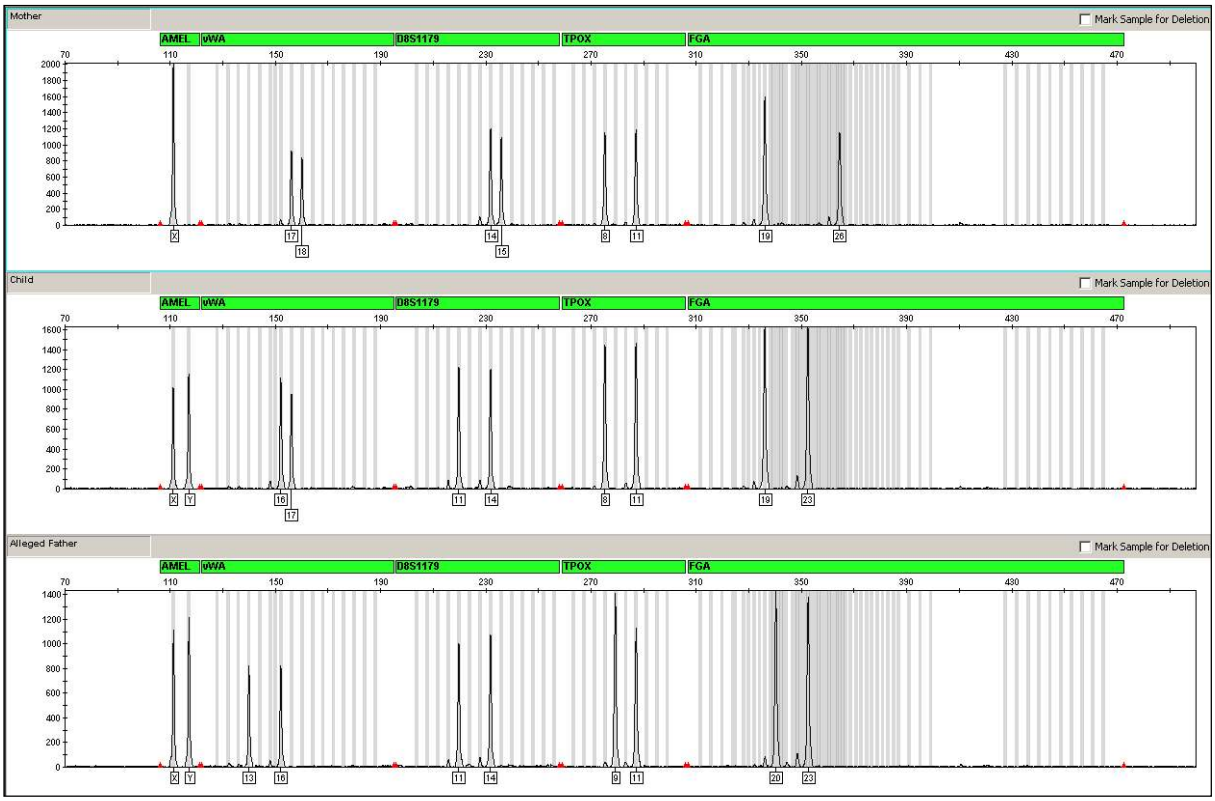


Figure 3.5.3.3. *Yellow channel (5/16 markers) showing 3 genetic profiles in a paternity testing analysis. From top to bottom: mother (a), child (b) and alleged father (c). In all depicted markers, the child shares one allele with the mother and the other with the father.*

3.5.4. Y-CHROMOSOME HAPLOGROUP SCREENING (Q, R AND I)

Screening for haplogroup Q1a3a1, R1b1a2 and I was performed with multiplex Real Time PCR and High Resolution Melting Analysis. For this purpose, all samples have been normalized to a concentration of 10ng/µL to standardize the amplification and the melting curve analysis. All reaction were carried out in a total volume of 25µL, containing 0.75µL Syto® 9 (50µM), 0.4µL GoTaq® Polymerase (5 U/µL), 5µL Buffer 5X (Cl_2Mg 1.5mM) and 0.2µL dNTPs (25mM). Known haplogroup status samples were used as controls. The cycling profiles were performed as follows: 94°C 1 min – 94°C 5 sec and 55°C 10 sec (x33) – 72°C 2 min. Primer sequences are listed in Table 3.3.4.1 and primer concentrations in Table 3.5.4.

Table 3.5.4. *Primers used for Y chromosome haplogroup screening*

Hpg	Marker	SNP	Amplicon length (bp)	Primer concentrations
Q1a3a1	M3	C/T	58	10pmol
R1b1a2	M269	T/C	76	Fwd 20pmol / Rv 40pmol
I	U179	G/A	93	10pmol

All Real Time PCR reactions were performed on a Rotor Gene 6000 (Corbett Life Science, Australia). After amplification, the products were denatured by increasing the temperature from

69°C to 80°C and the florescence was measured each 0.2°C. From the acquired data, a melting curve was plotted for each sample by calculating the derivative dF/dT for each data point using the Rotor-Gene 6000 software. The results, depicted in Figure 3.5.4, allowed assignment of the corresponding haplogroup to all samples by comparing the position of the peak (indicated with the corresponding bin) with the control samples for Hg R1b1a2, Q1a3a1 and I. This approach for haplogroup assessment was recently published in *Forensic Science International: Genetics* by Zuccarelli, Alechine et al (2011).

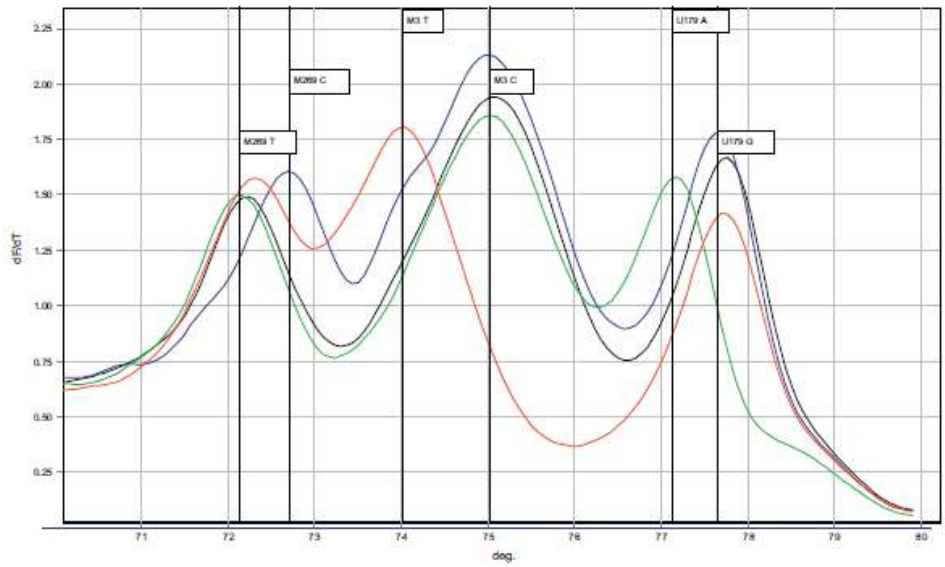


Figure 3.5.4. *Y chromosome haplogroup screening melting profile showing Q1a3a1 (red), R1b1a2 (blue) and I (green) controls (Zuccarelli, Alechine et al. 2011)*

3.5.5. BLOOD CULTURE

1mL of blood sample with heparin was cultured in a flask with 10mL Iscoves Medium, 100µL PHA and 2 drops of heparin for 71.5h in a 37°C oven and CO_2 atmosphere. 6.5h before harvesting 100µL of BrdU were added, mixed thoroughly and put back in the 37°C oven. 30min before harvesting 100µL Colcemid were added, mixed and put back in the 37°C oven. The blood culture was then centrifuged for 10 min at 1000rpm and the supernatant was removed. The tube was filled up to 3/4 with hypotonic solution and the blood cells were resuspended very smoothly, but completely, and incubated for 30 min in water bath at 37°C. The tube was filled with fixative, resuspended and centrifuged 10 min at 1000rpm. The supernatant was removed with vacuum and washed vigorously 2-3 times with fixative. The suspensions were checked on the slides and left filled with fixative overnight at -20°C. The next day the fixation step was repeated once more and the slides were made using the "splash" technique. Dilutions were made if necessary.

3.5.6. FISH (FLUORESCENCE IN SITU HYBRIDIZATION)

Specific probes for FISH analysis were prepared from cosmids cos6B7 for *DAZ* (Taylor, Hornigold et al. 1996) and cos2A49 for *CDY* (Kuhl, Rottger et al. 2001)(Table 3.3.4.2). For cosmid purification, Qiagen Plasmid Maxi Kit has been used as described by the manufacturer. The bacteria with the cosmids were cultured in 200mL LB medium (10g Trypton, 10g NaCl and 5g Yeast in 1L H_2O, pH=7.2-7.4 with 30 µg/mL Kanamycin) for 12-16 hs at 37°C. The culture was centrifuged at 6000rpm for 15min at 4°C. The supernatant was discarded and the pellet resuspended in 10mL P1 Buffer. 10mL of P2 Buffer were added, mixed by inversion and incubated for 5 min at room temperature. Then, 10mL of P3 Buffer were added, mixed and incubated for 15 min on ice. The previous was centrifuged for 30min at 12.000rpm at 4°C. The supernatant was filtered in a 50mL falcon tube. The column was prepared with 5mL QBT Buffer, and the supernatant has been passed through the column. Afterwards, the column was washed 2 times with 10mL QC Buffer. The DNA was eluted from the column to a new tube with 5mL QF Buffer. 3.5mL of Isopropyl alcohol was added, inverted and centrifuged for 30min at 10.000rpm at 4°C. The supernatant was discarded and the pellet resuspended in 2mL 70% Ethanol and passed to a 2mL Eppendorf tube. This was again centrifuged for 10min at 12.000rpm at 4°C, the supernatant was discarded and the pellet left to dry for 5-10min. Finally, the DNA was resuspended in 50µL TE Buffer. DNA concentration was measured by NanoDrop.

For labeling of the probes with biotin (BIO) or digoxigenin (DIG), DOP-PCR was performed in a final volume of 50µL, 5µL of 10X Buffer, 2.5µL of 50mM Cl_2Mg, 8µL of 1.25mM dNTPs (dATP, dCTP and dGTP), 6µL of 1.25mM dTTPs, 5µL of 1mM Biotin-16-UTP or 1mM Digoxigenin-11-UTP, 1µL of 100pmol/µL DOP-primers, 0.5µL of Taq Polymerase and 200ng of purified DNA. Cycling conditions were: initial step 94°C 5 min followed by 35 cycles of 94°C 20 sec - 56°C 1 min - 72°C 3 min and a final extension at 72°C for 10 min. The PCR reaction was carried out in a PTC 100 Thermal cycler (MJ research).

For probes´ precipitation, 30µL of CDY or DAZ PCR product, 90µL of Isopropyl alcohol and 15µL $AcNH_4$ 7.5M was incubated for 2 hours at -70°C. After 25-minute centrifugation at 13.500rpm, supernatant was discarded and the pellet was washed with 70% Ethanol. After centrifugation for 5 minutes at 13.500rpm, supernatant was discarded and the pellet was resuspended in 5µL of H_2O. DNA concentration was measured with NanoDrop.

Prior to *in situ* hybridization, the slides were denatured in 70% Formamide 2XSSC (NaCl 15mM and NaCitrate 1.5mM pH=7.0) at 72°C and then dehydrated in 70%, 90% and 100% ethanol, each for 2 min, and then air-dried. All FISH-assays were performed on metaphase spreads. Hybridization of the probes was performed with 50% Formamide, 10% Dextran sulfate, 30% PEG 8000, 1 µg Cot DNA (1µg/µl) in 2X SSC (Hybridization mix). 1µL of each probe was mixed with 4µL of the hybridization mix and denatured for 5 minutes at 95°C, and then incubated with the slides

over-weekend at 37°C in a wet chamber. After washing for 2 minutes in 0.4XSSC 0.3% Tween at 42°C and 2 minutes in 2XSSC 0.1% Tween at room temperature, the slides were incubated in blocking solution (5% powder milk and 1% BSA in 2XSSC) for 10 minutes at room temperature. Afterwards, washing of the slides in 2XSSC at 42°C was performed. Biotinylated probes were detected with TRITC-conjugated Avidin and DIG-labeled probes using anti-DIG mouse antibodies (Sigma) followed by FITC-conjugated goat anti-mouse antibodies (Sigma). First, the slides were incubated with 20µL of anti-DIG mouse antibody (dilution 1/200, as recommended by the manufacturer) for 20 minutes at 37°C in a wet chamber and then washed once in 0.1% Tween 2XSSC and twice in 2XSSC for 2 minutes each at 42°C. For detection, slides were incubated with 20µL FITC-conjugated anti-mouse antibody (0.1 mg/mL) and TRITC-conjugated Avidin (0.02 mg/mL) for 20 minutes at 37°C in a wet chamber. Slides were then washed once in 0.1% Tween 2XSSC and twice in 2XSSC for 2 minutes each at 42°C, counterstained with DAPI (4',6-diamidino-2-phenylindole; 60 ng/mL). All images were obtained using a BX-60 Olympus IX-71 Microscope (Olympus, Japan) and digitalized with an Optronics camera. DAPI, FITC and TRITC images were merged using Adobe PhotoShop program.

3.5.7. MULTIPLEX Y-STSS ANALYSIS

Two multiplex PCR amplification reactions were performed for the following STS markers located in the AZFb and AZFc region: sY1191, sY1206, sY1261, sY1291 and sY1201 (Repping et al., 2003). Figure 3.5.7 depicts the location on the Y chromosome of each of the STS markers used. Accession numbers, amplification product sizes, copy number in the reference sequence and primer concentrations in the multiplex reactions are summarized in Table 3.5.7.

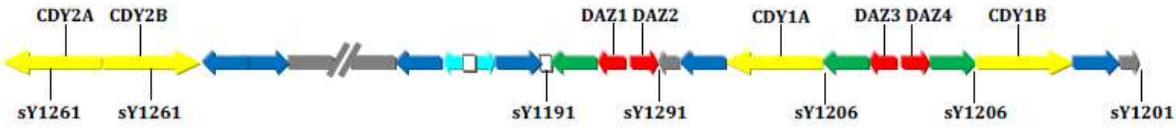


Figure 3.5.7. *Y-STSs location on the reference Y chromosome (adapted from Navarro-Costa et al., 2010)*

Amplification of sY1191, sY1206 and sY1261 was performed in triplex with 5µL Multiplex PCR Master Mix (QIAGEN), 3pmol FAM-labeled M13 primer, 5pmol sY1191, 10pmol sY1206, 10pmol sY1261 and 1µL template DNA (1ng/µL approximately) in a total volume of 10µL. The cycling conditions were the following: 15 minutes for Taq polymerase activation at 95°C, 30 cycles of 95°C 30 sec, 60°C 30 sec and 72°C 30 sec, and 8 cycles of 95°C 30 sec, 53°C 30 sec and 72°C 30 sec, followed by 30 minute extension at 72°C. Duplex PCR amplification was performed for sY1202 and sY1291 with 5µL Multiplex PCR Master Mix (QIAGEN), 3pmol FAM-labeled M13 primer, 30pmol sY1291, 30pmol sY1261 and 1µL template DNA (1ng/µL approximately) in a total volume of 10µL. The cycling

conditions were the following: 15 minutes for Taq polymerase activation at 95°C, 30 cycles of 95°C 30 sec, 58°C 30 sec and 72°C 30 sec, and 8 cycles of 95°C 30 sec, 53°C 30 sec and 72°C 30 sec, followed by 30 minute extension at 72°C. All PCR reactions were performed in a GeneAmp® PCR System 9700 (Applied Biosystems). PCR primers´ sequences are listed in Table 3.3.4.1 and have been designed with M13 tails for product detection with M13 primers labeled with FAM™, as previously described (Oklander, Zunino et al. 2007). After PCR, products were prepared with 7.5µL Hi-Di Formamide, 1µL ILS 600 (Internal Lane Standard 600), 0.5µL of the triplex PCR (sY1191, sY1206 and sY1261) and 1µL of the duplex PCR (sY1291 and sY1201), and denatured for 3 min at 95°C. PCR products´ resolution and detection was performed with a 3100-*Avant* Genetic Analyzer (Applied Biosystems) using a 36cm array and POP4 polymer. Capillary electrophoresis conditions were 60°C, 15kV/180 sec (pre-run), 4kV/1 sec (injection) and 15kV/1800sec (run). Analysis of the results was performed with GeneMapper ID-X software.

Table 3.5.7. *Primers used for Y chromosome microdeletion screening by STS amplification*

STS	Accession Number	Amplicon	Copies in Reference Sequence	Primer Pair Concentration
sY1191	G73809	385bp	1	5pmol
sY1206	G67171	394bp	2	10pmol
sY1291	G72340	527bp	1	10pmol
sY1201	G67170	677bp	1	30pmol
sY1261	BV703585	481bp	2	30pmol

In order to confirm a deletion of a specific STS marker, a second PCR reaction was performed. This duplex analysis was performed with the deleted marker and with a second (not deleted) marker as internal PCR control. Amplification and detection conditions were the same as for the multiplex reactions described above.

3.5.8. SEQUENCING ANALYSIS

In order to confirm STS results, sequencing analysis was performed after multiplex amplification. First, PCR products were incubated for 15 min at 37°C with ExoSAP-IT enzyme to hydrolyze primers, followed by 15 min at 80°C for enzyme inactivation. Afterwards, the products were sequenced by means of Big Dye Terminator v1.1 kit with 2µL Premix, 6µL 2.5X Reaction Buffer, 10pmol F or R primer and 2µL of PCR product in a total volume of 20µL. Sequencing was performed in a GeneAmp® PCR System 9700 (Applied Biosystems) with the following cycling conditions: 25 cycles of 96°C 15 sec – 50°C 5 sec – 60°C 2 min. Purification of the obtained sequences was performed with Isopropyl alcohol 75% precipitation, washing with Ethanol 70% and air dried. Denaturalization was performed

in 22µL of Hi-Di Formamide for 3 min at 95°C. Sequences resolution and detection was performed in a 3100-*Avant* Genetic Analyzer (Applied Biosystems) using a 36cm array and POP4 polymer. Capillary electrophoresis conditions were 55°C, 15kV/180 sec (pre-run), 1kV/7 sec (injection) and 15kV/1700sec (run). Sequencing analysis was carried out with SecScape and Sequencher software.

3.5.9. COPY NUMBER ANALYSIS OF DAZ AND CDY GENES

Copy number analysis was performed by a quantitative Real Time PCR approach previously described (Munch, Kirsch et al. 2008; Noordam, Westerveld et al. 2011). qPCR was carried out using primers for *DAZ*, *CDY* and *SRY* genes. Primers were designed using human Y chromosome sequence obtained from GenBank (NC_000024.9) (Skaletsky, Kuroda-Kawaguchi et al. 2003)and Primer3 software. In order to obtain a more standardized amplification level, amplicon length was set at 110-150bp. Homodimers and heterodimers, potential hairpins, and secondary structure formations were checked using Oligo Analyzer software; primer specificity was checked using UCSC In-Silico PCR on-line software. Amplification products on the reference sequence are shown on Figures 3.5.9.2 - 3.5.9.8 and can be also accessed from http://tinyurl.com/3slt2bk. Primer sequences used for *DAZ*, *CDY* and *SRY* are summarized in Table 3.3.4.1. All qPCR reactions were carried out with 5µL QuantiTec SYBR® Green PCR Master Mix (Qiagen), 2.5µL of each primer pair (0.25 µM final concentration) and 2.5µL of template DNA (0.4-20ng/µL) in a total volume of 10µL. All assays were performed in duplicate and a calibrator sample was included in every run (also in duplicate). Cycling conditions were 50°C 2 min, 95°C 15 min and 35 cycles of 95°C 15 sec, 55°C 30 sec and 72°C 30 sec. A melting curve analysis was performed routinely following amplification to detect non-specific products. All qPCR reactions were carried out in a Rotor-Gene 6000 Real Time PCR equipment (Corbet Life Science, Australia).

qPCR validation was performed according to Livak and Schmittgen (2001) to ensure that amplification efficiencies of target and reference genes are approximately equal. For this purpose, a standard curve was performed using 2-fold serial dilutions of the control sample DNA ranging from 50 to 0.4ng/µL. Results of the calibration are showed in Figure 3.5.9.1 and the expected slope for DAZ/SRY and CDY/SRY standard curves was close to zero (DAZ slope=-0.25; CDY slope=-0.29).

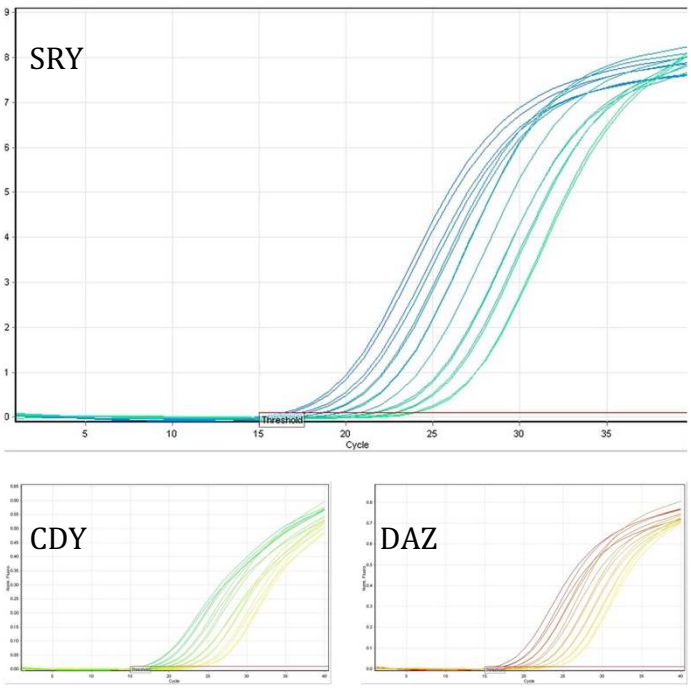


Figure 3.5.9.1. *qPCR results for the validation run (dilutions ranging 50-0.4ng/μL)*

The control sample used as calibrator belongs to hg R1b1a2 and was characterized by paternity testing and FISH analysis. The number of *DAZ* and *CDY* genes was calculated by interpolation of the Ct value in the corresponding standard curve using Rotor-Gene 6000 software (Corbett Life Science). In all cases, the copy number of *DAZ* or *CDY* was normalized to *SRY*, used as reference single copy gene located on the Y chromosome. Mean values were calculated for each sample from the normalized *DAZ* and *CDY* copy number values.

This approach was first set up for the Y chromosome genes with primate samples that have been also analyzed with FISH for DAZ and CDY genes (Greve et al., 2011).

Table 3.5.9. *Characteristics of the qPCR amplification products*

Gene	Product size (bp)	N° products HUMAN
SRY	124	1
DAZ	121	4
CDY	111	4

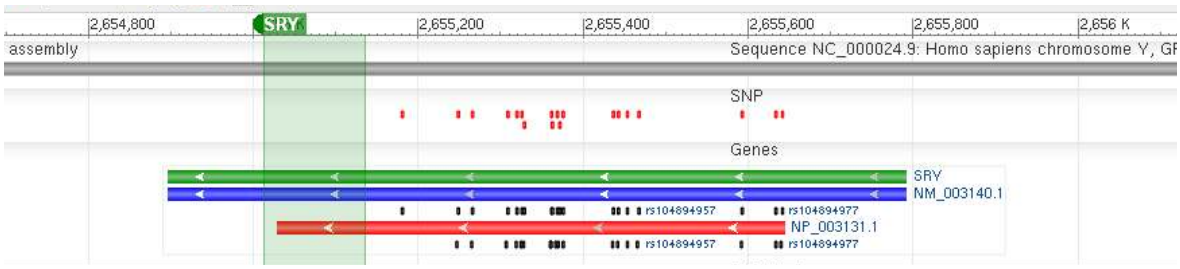


Figure 3.5.9.2. *Predicted SRY amplification product for the reference sequence*

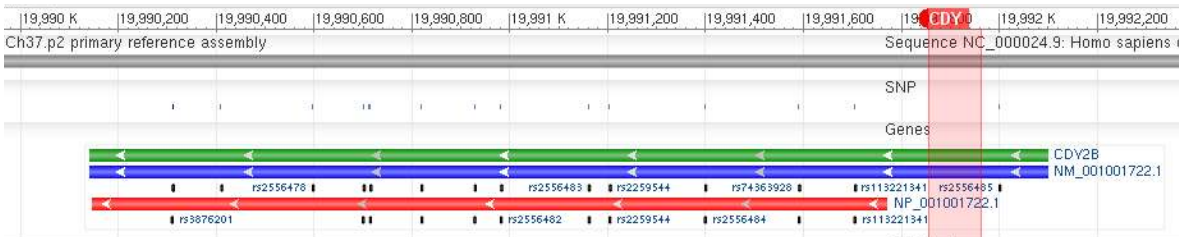


Figure 3.5.9.3. *Predicted CDY2B amplification product for the reference sequence*

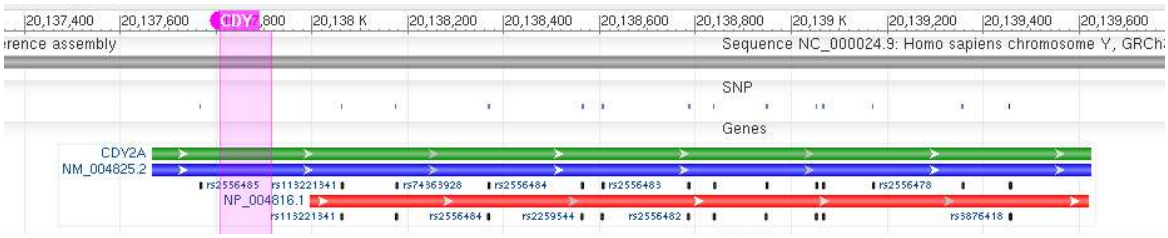


Figure 3.5.9.4. *Predicted CDY2A amplification product for the reference sequence*

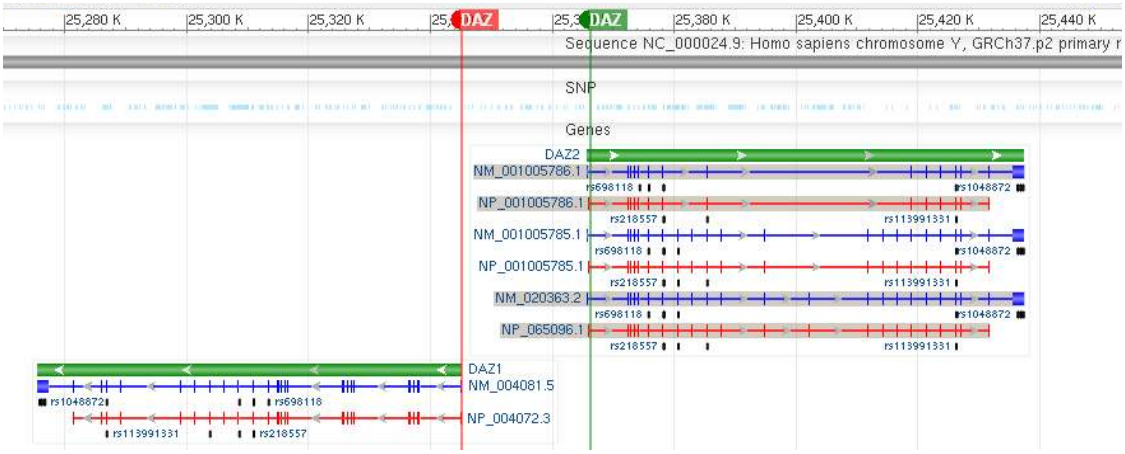


Figure 3.5.9.5. *Predicted DAZ1 and DAZ2 amplification products for the reference sequence*

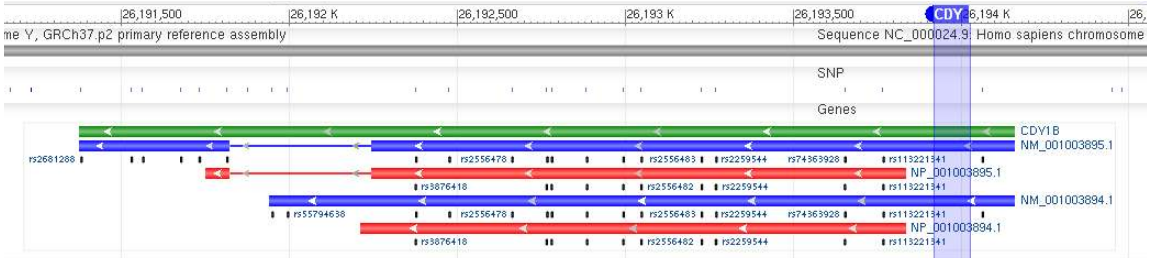


Figure 3.5.9.6. *Predicted CDY1B amplification product for the reference sequence*

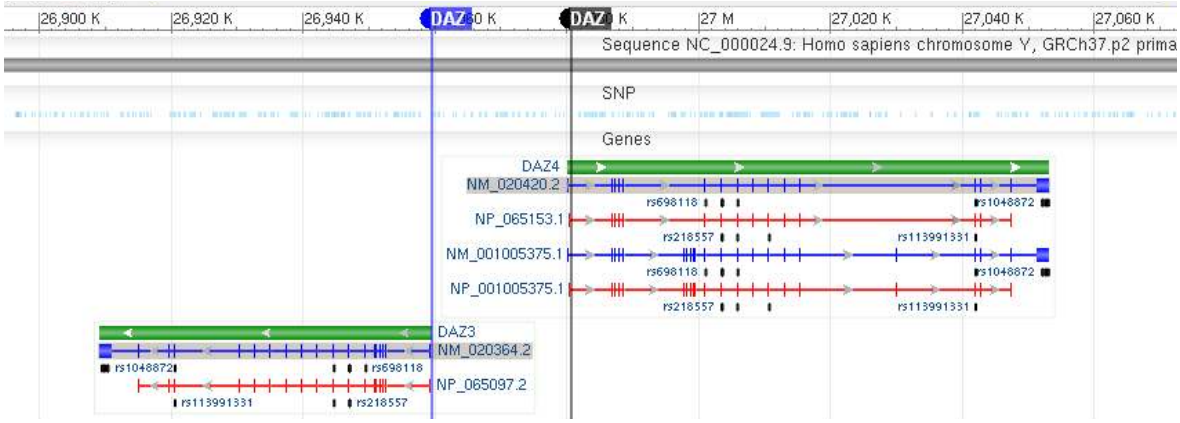


Figure 3.5.9.7. *Predicted DAZ3 and DAZ4 amplification products for the reference sequence*

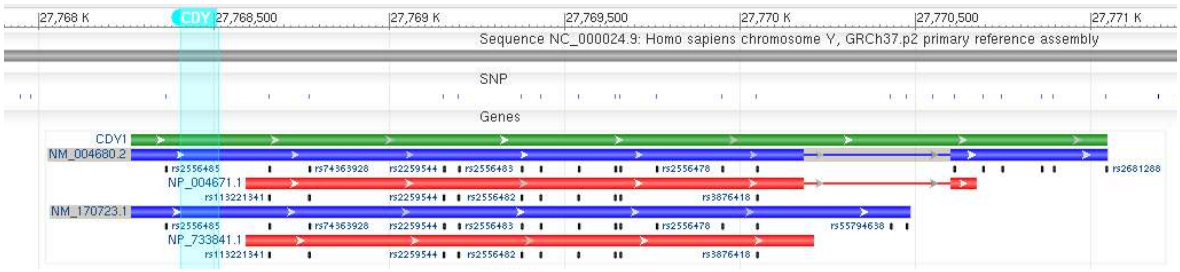


Figure 3.5.9.8. *Predicted CDY1(A) amplification product for the reference sequence*

3.6. STATISTICAL DATA ANALYSIS

Chi-Square Calculator with nine rows and columns was used for statistical analysis of DAZ and CDY copy number between haplogroup Q1a3a1 and R1b1a2 (Kirkman, T.W. 1996).

4. RESULTS

4.1. SAMPLE SELECTION

For paternity casework, after DNA extraction and quantification, DNA typing analysis was performed to determine whether the analyzed alleged father was the biological father or he was excluded as a first degree relative. Only unrelated male individuals who were confirmed for paternity by autosomal STRs were considered for the ongoing analyses.

Afterwards, all samples have been submitted to Y haplogroup screening by Real Time PCR and HRM, as described above. Figure 4.1 depicts the result of the haplogroup screening assay performed. From these results it is possible to classify the samples as belonging to haplogroup Q1a3a1 or R1b1a2 by comparison with the control sample peak (melting temperature). Samples belonging to haplogroup I or any other haplogroup were not used for following analyses.

As a result of the screening process, 37 samples from confirmed fathers belonging to haplogroup Q1a3a1 and an equal number of samples from fathers belonging to haplogroup R1b1a2 have been selected.

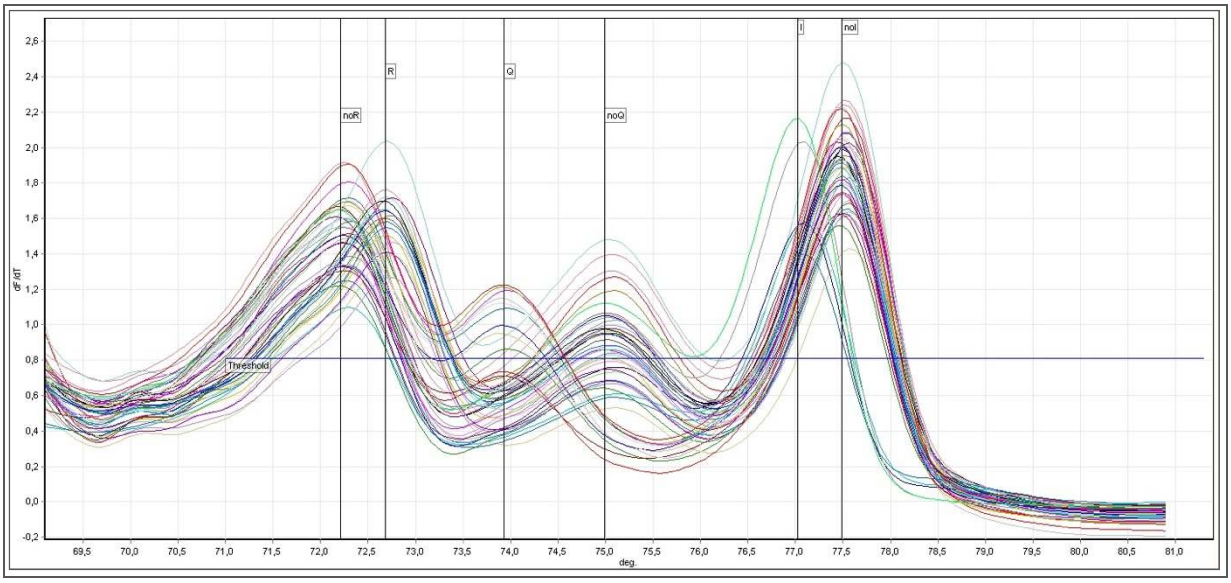


Figure 4.1. *HRM profile for the analyzed samples*

4.2. FISH ANALYSIS FOR THE CONTROL SAMPLE

In order to be able to use the control sample as calibrator for the following qPCR analyses, it was necessary to confirm (i) that the sample belongs to haplogroup R1b1a2 (as described above) and therefore GenBank´s reference sequence information can be applied, (ii) that the individual was fertile and the paternity was confirmed by autosomal STR analysis (as described above) and (iii) that there were not cytogenetic abnormalities on the Y chromosome.

To accomplish the latest, FISH analysis with probes for DAZ and CDY genes was performed on the control sample. As it is shown in Figure 4.2, in all analyzed metaphases it was possible to distinguish 1 signal for DAZ and 2 signals for CDY in each chromatide. These results were in concordance with previous results on the human Y chromosome. It was not possible to distinguish the four copies of DAZ and they were seen as only 1 signal by FISH analysis. This effect is due to the fact that the DAZ copies are located so close to each other (1Mb difference between duplets) that this cytogenetic technique could not separate them as four signals and is only seen as 1 big signal under the microscope. In the case of CDY, it was possible to distinguish 2 signals one corresponding to CDY1A/CDY1B (distal) and the other to CDY2A/2B (proximal) with a separation of 6Mb. In this case, it is not possible to distinguish between the A and B copies, but it was possible to evidence the presence of CDY1 and CDY2.

From these results, and in accordance with previously described human Y chromosomes (Rottger, Pasantes et al. 2000; Kuhl, Rottger et al. 2001; Wimmer, Kuhl et al. 2002), it is possible to conclude that the control sample shows a normal Y chromosome, for the matter of this study, and that it can be used as calibrator.

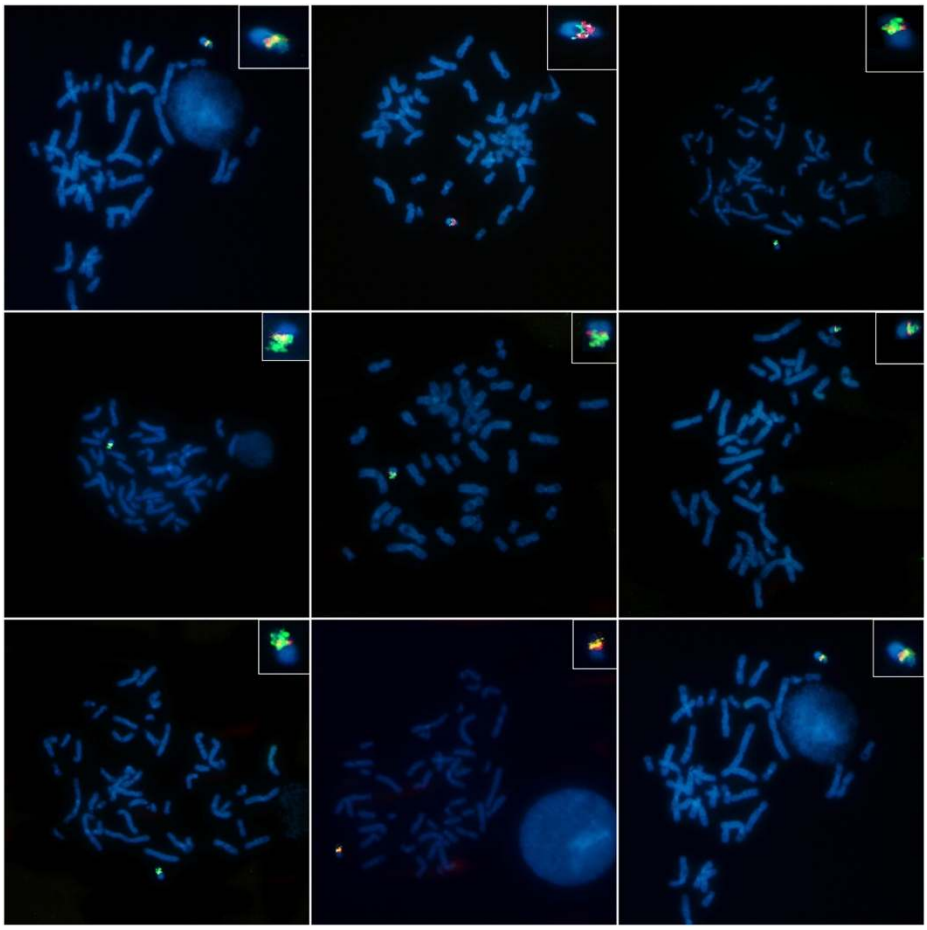


Figure 4.2.1. *FISH analysis of the control sample belonging to the reference Hg R (DAZ signals are shown in red and CDY signals in green)*

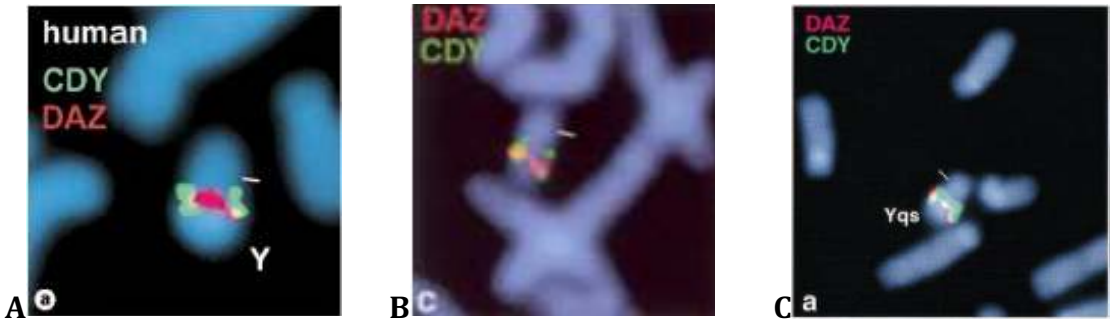


Figure 4.2.2. *Previously published FISH results on human Y chromosomes (Rottger, Pasantes et al. 2000; Kuhl, Rottger et al. 2001; Wimmer, Kuhl et al. 2002)*

4.3. Y-STSs ANALYSIS

The assay for identification of microdeletions on the AZFc region by means of plus/minus amplification of STSs and CE, was successfully set up. Figure 4.3.1 shows the results for the positive, negative female and non-template controls.

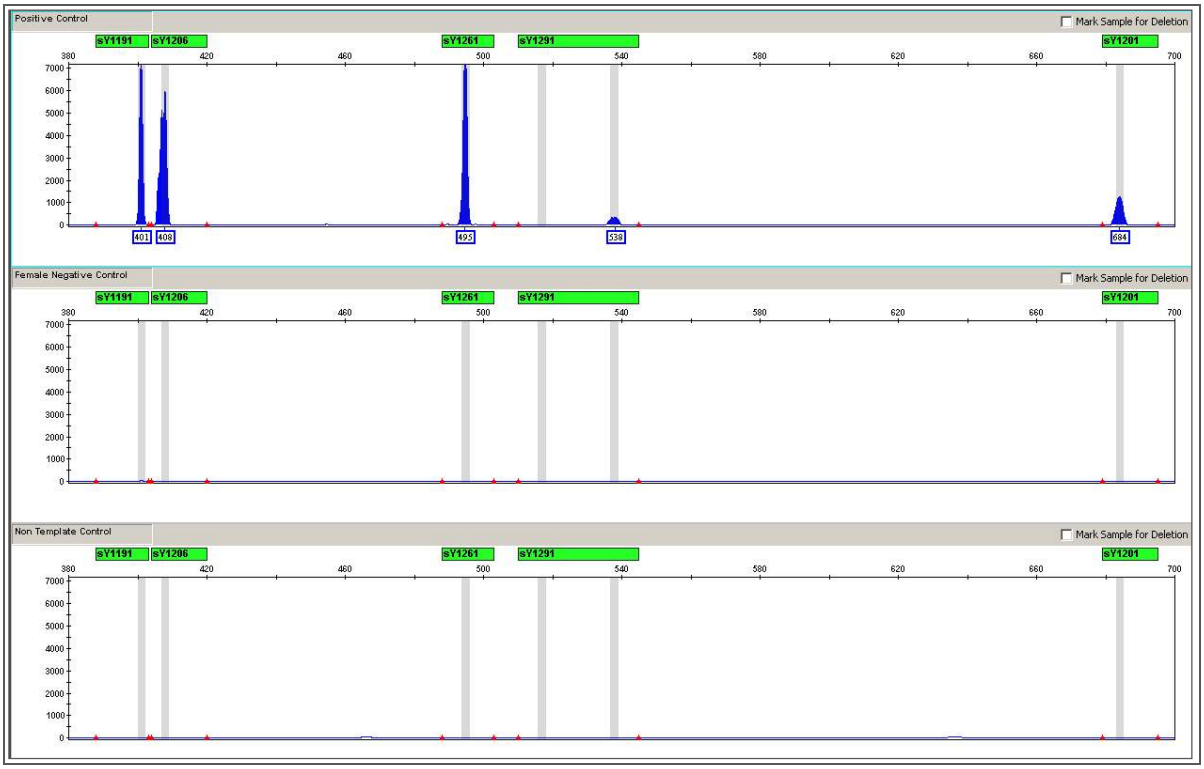


Figure 4.3.1. *Results of Y-STS for the positive, female and non-template controls*

The analysis of 37 samples belonging to haplogroup R1b1a2 and 37 samples belonging to haplogroup Q1a3a1 showed no evidence of any differential deletion of Y-STS markers between the two haplogroups. Nevertheless, the sY1291 marker has a differential size between the haplogroups: 517pb in Q1a3a1 and 538pb in R1b1a2. This size difference was consistent within all the analyzed

samples from both haplogroups. Figure 4.3.2 shows the results for R1b1a2 samples and Figure 4.3.3 for Q1a3a1 samples.

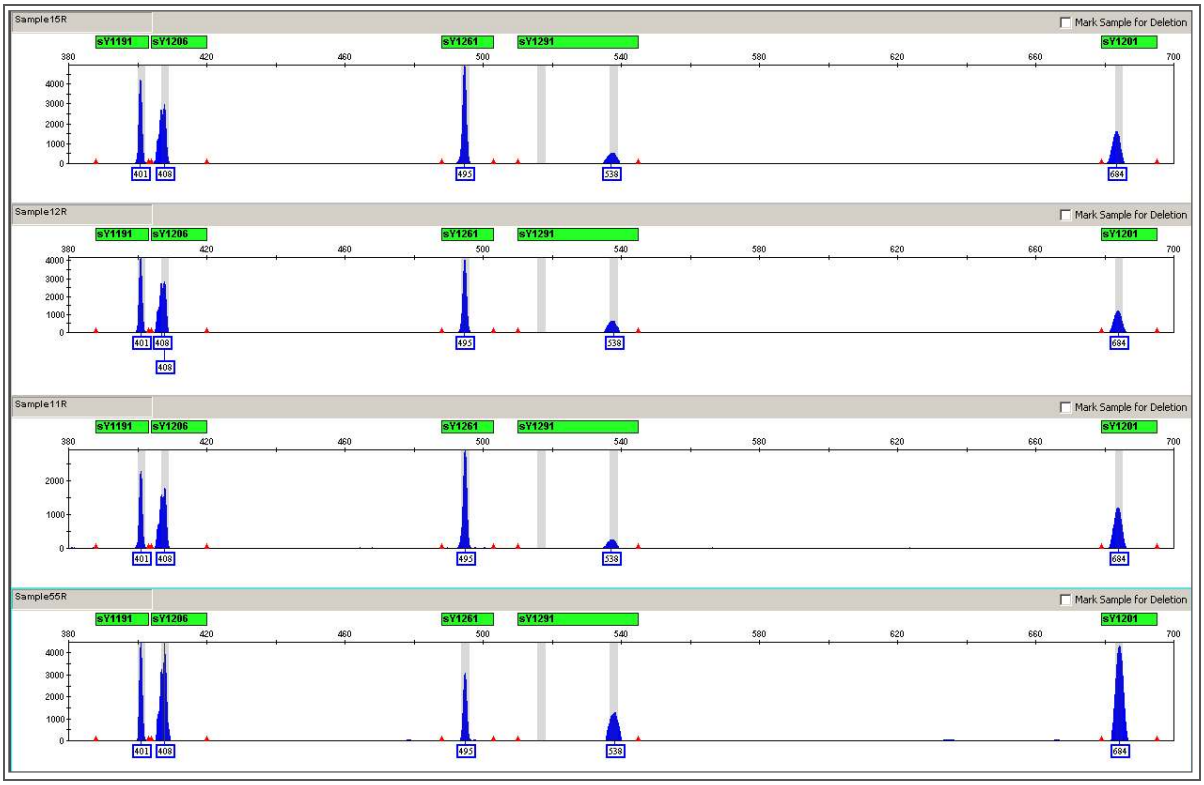


Figure 4.3.2. *Y-STS results for haplogroup R1b1a2 samples*

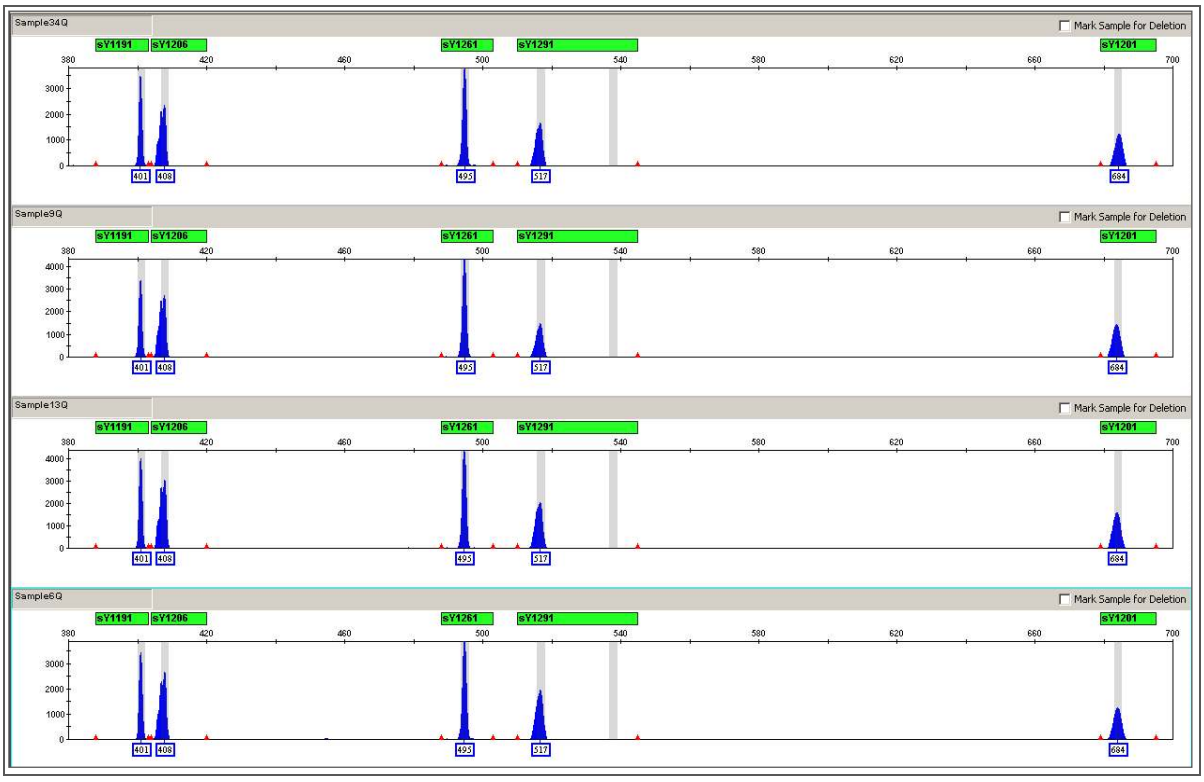


Figure 4.3.3. *Y-STS results for haplogroup Q1a3a1 samples*

From the 74 samples analyzed, only 2 showed a deletion of sY1291 marker. One sample belonged to haplogroup R1b1a2 (50R) and the other to Q1a3a1 (27Q). Figure 4.3.4 shows the results for the 2 deleted samples.

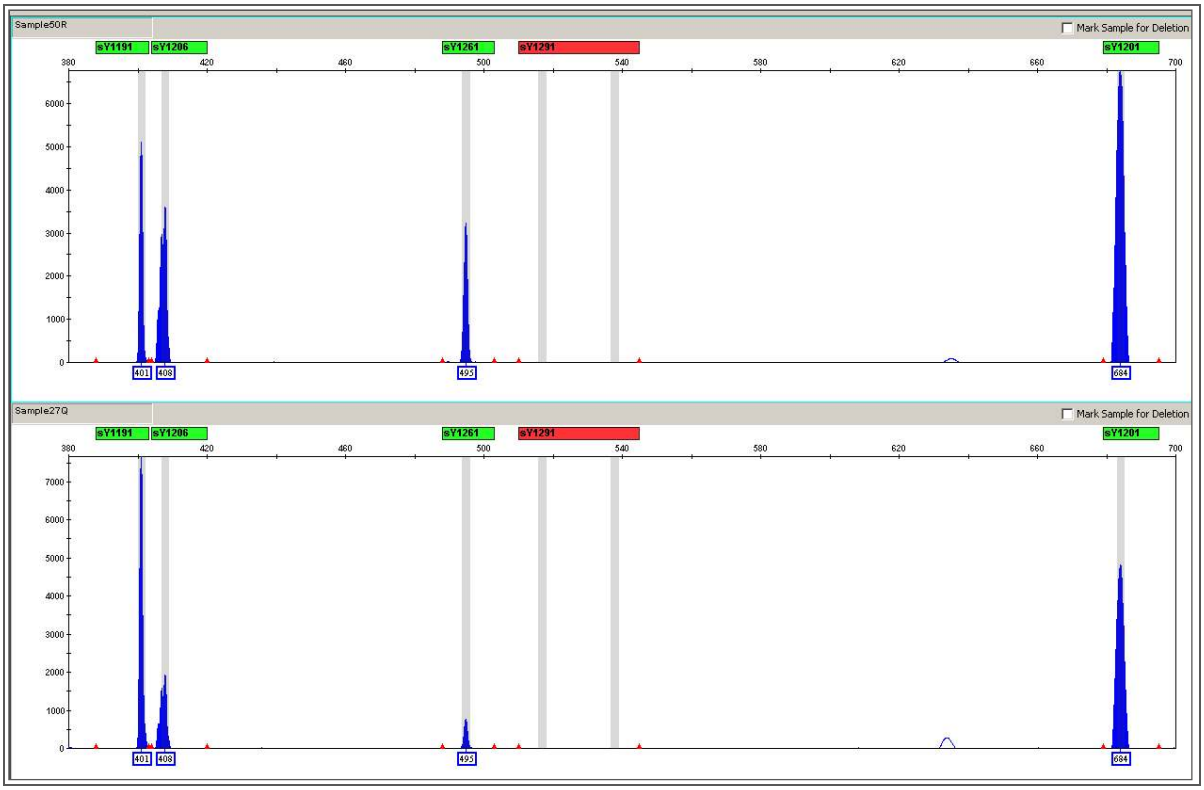


Figure 4.3.4. *Y-STS results for samples 50R (top) and 27Q (bottom) with a sY1291 deletion*

From all the set of samples belonging to haplogroup Q1a3a1, characterized by the derived state for M3, none of the analyzed samples showed the sY1191 deletion previously described by Repping et al (2004, 2006). To compare this results with the other haplogroup in which this deletion was also described, a sample belonging to haplogroup N was also analyzed for Y-STSs. The result of this analysis is depicted in Figure 4.3.5 and shows that the sY1191 deletion is present in haplogroup N and absent in haplogroup Q1a3a1.

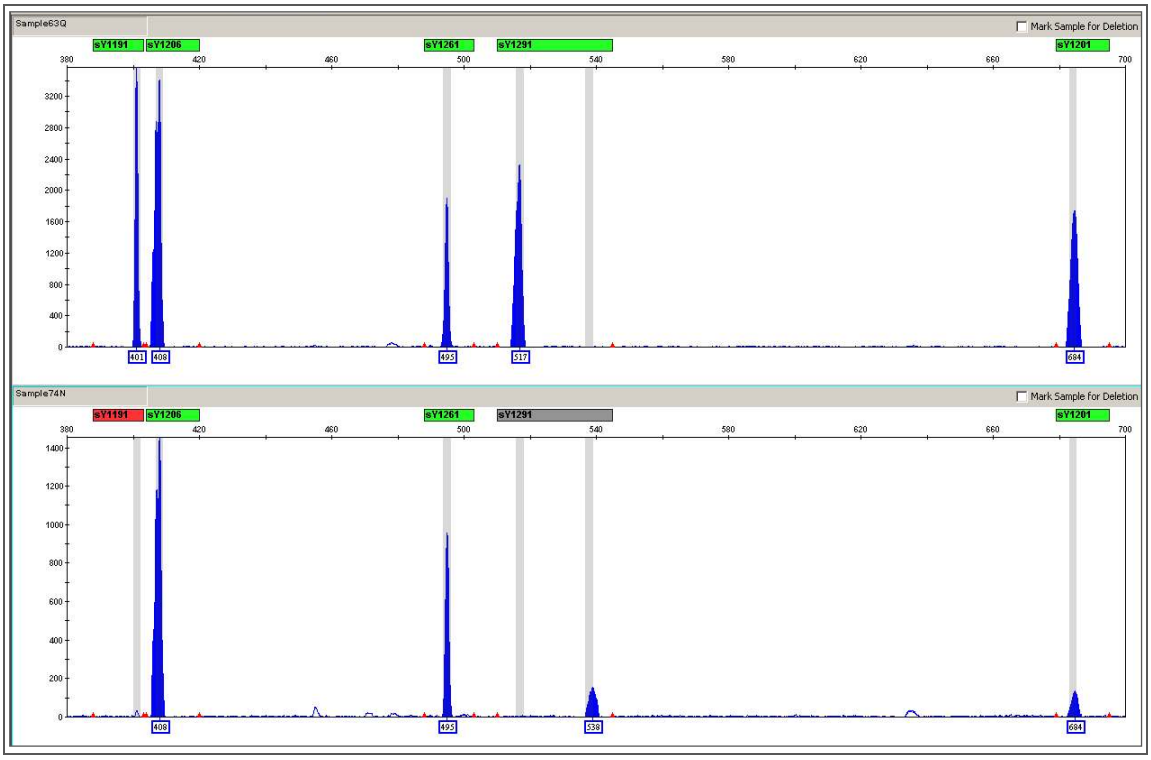


Figure 4.3.5. *Y-STS haplotype showing a deletion of sY1191 in haplogroup N (74N) and no deletion in haplogroup Q1a3a1 (63Q)*

4.4. SEQUENCING OF SY1291

For confirmation of the size difference seen in sY1291 marker between haplogroup R1b1a2 and Q1a3a1, sequencing analysis was performed. The results, shown in Figure 4.4, allow identifying that the size difference was due to a homopolymeric T region that has 16T in the haplogroup Q1a3a1 samples and 39T in the R1b1a2 samples.

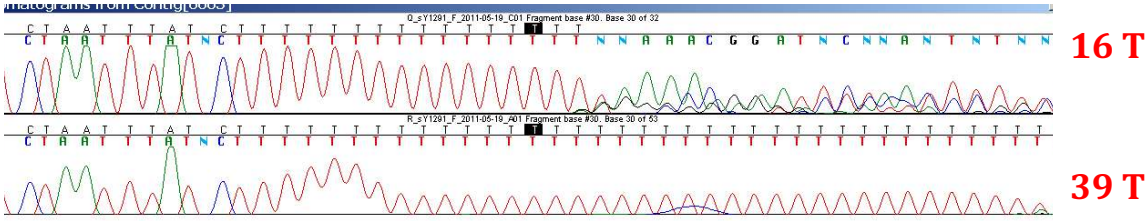


Figure 4.4. *Results of the sequencing analysis of sY1291 showing the homopolymeric T region in haplogroup Q1a3a1 (top) and R1b1a2 (bottom)*

4.5. SEQUENCING OF SY1206

Sequencing of sY1206 marker was performed in order to identify which of the two copies of sY1206 was present (Figure 3.5.7). Two polymorphic positions have been analyzed: positions 122 and 311 (from the 394pb sequence defined by GenBank primers; Accession Number G67171). On the reference sequence, in the position 122 (rs1297904), T is present in the proximal copy of sY1206 and G in the distal copy. In the position 311 (rs805793), A is present in the proximal copy and G in the distal. By sequencing the products from the Y-STSs amplification, is was possible to identify that, meanwhile position 311 remains invariable (Figure 4.5.2), position 122 showed that in haplogroup

Q1a3a the T variant was present in both proximal and distal copies of sY1206 indicating that in the distal copy a sequence variation G>T has occurred (Figure 4.5.1 and 4.5.3).

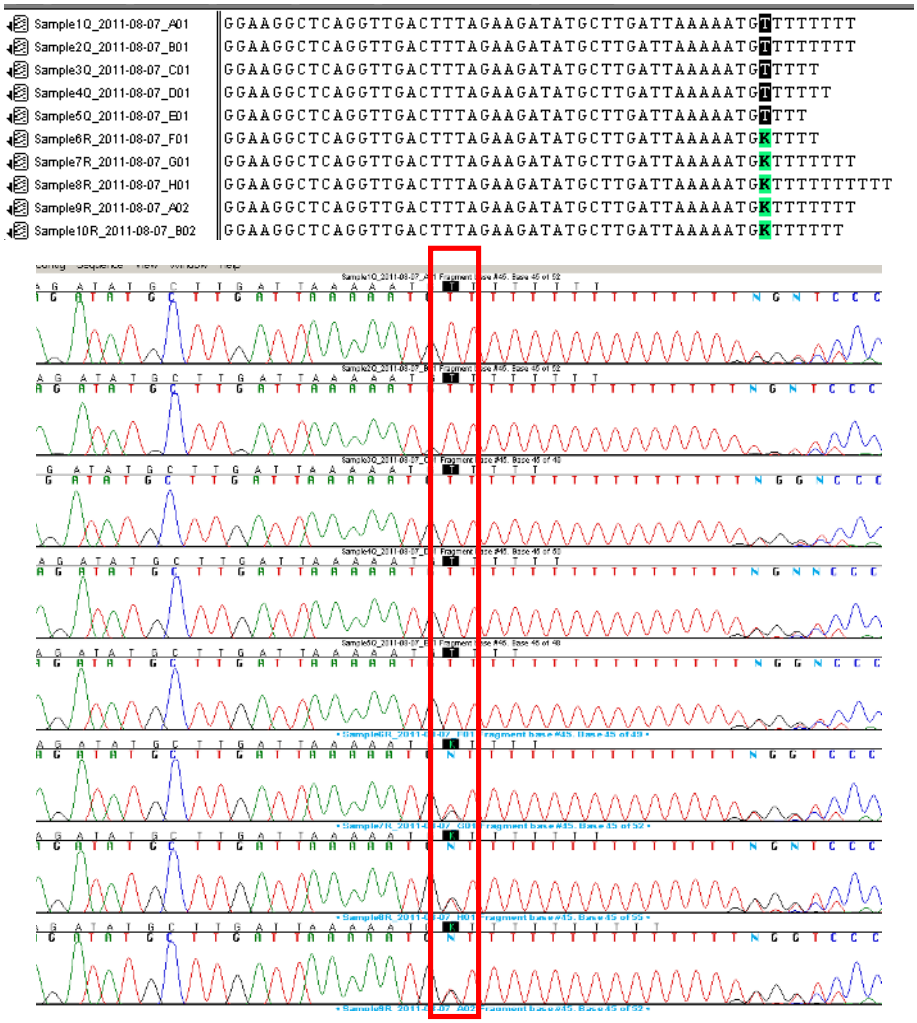


Figure 4.5.1. *Polymorphic position 122 showing T variant in haplogroup Q1a3a1 (first 5 from the top) and T/G (K according to IUPAC) heteroplasmy in haplogroup R1b1a2 (first 3 from the bottom)*

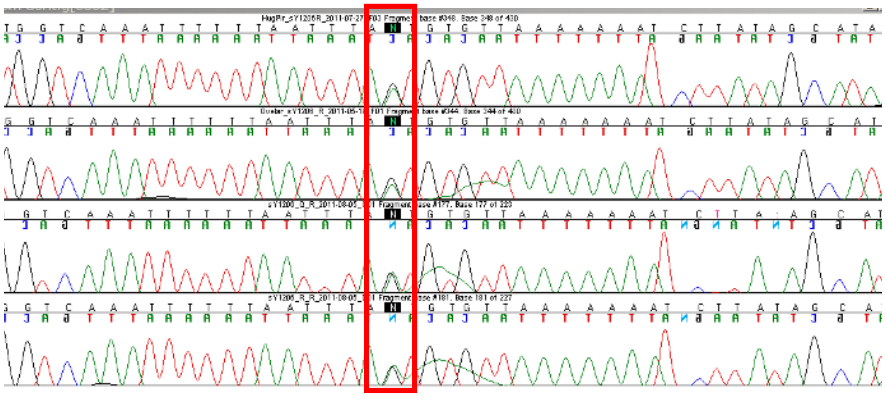


Figure 4.5.2. *Polymorphic position 311 for haplogroups R1b1a2 and Q1a3a*

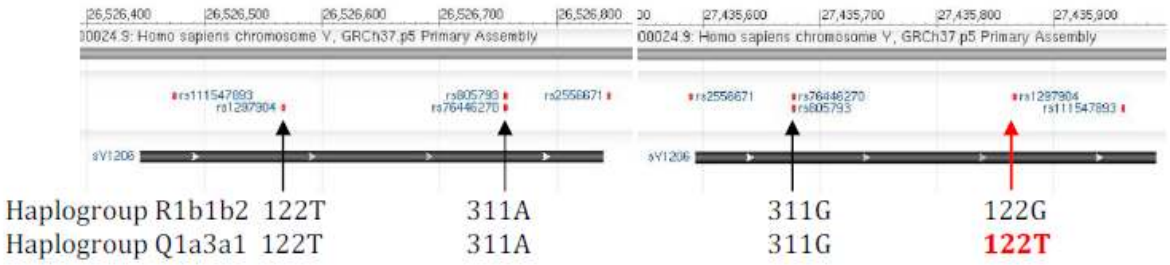


Figure 4.5.3. *Scheme of sY1206 polymorphisms*

From the sequencing analysis, one of the samples belonging to haplogroup Q1a3a1, showed a heteroplasmic site in positions 290-291 (C/A-G/T) (Figure 4.5.4).

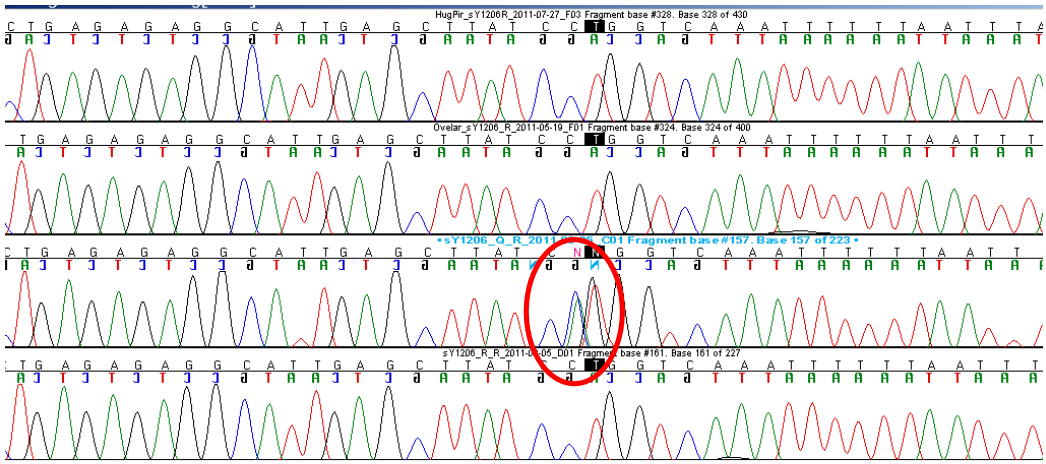


Figure 4.5.4. *Polymorphic positions 290-291 in sample 16Q (circled in red)*

4.6. Sequencing of sY1191

To confirm the results for sY1191, the amplification product was sequenced with both F and R primers, and the obtained sequence was aligned with GenBank´s reference sequence. Samples belonging to haplogroup R1b1a2 and Q1a3a1 have been sequenced. Although only 319/385bp had an optimal quality, the obtained sequence presented 100% identity with the reference sequence for this Y-STS marker (Figure 4.6).

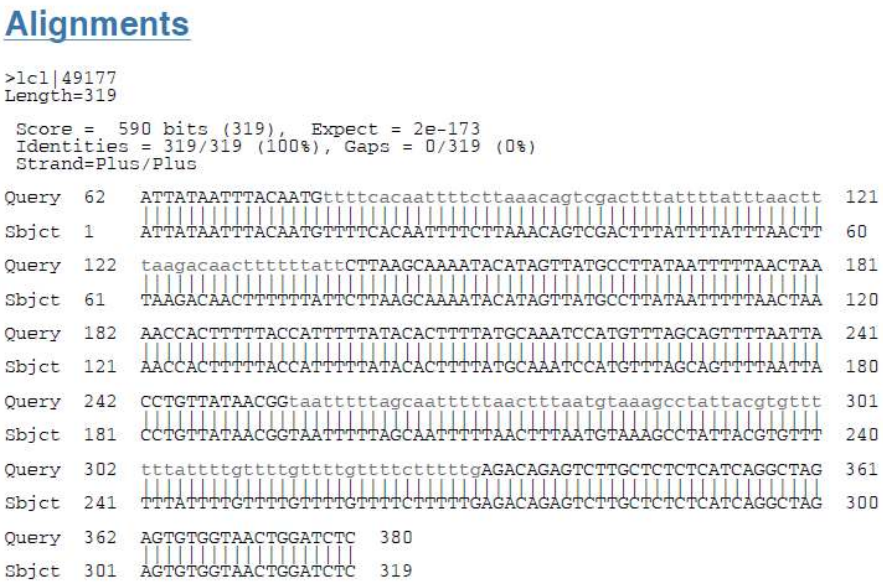


Figure 4.6. *Sequence alignment between the obtained consensus sequence for both haplogroups and the reference sequence for sY1191*

4.7. qPCR FOR DAZ & CDY COPY NUMBER ANALYSIS

A total of 55 samples (30 from Hg Q1a3a1 and 25 from Hg R1b1a2) have been analyzed for DAZ and CDY gene dosage analysis. The obtained results are depicted in Figure 4.7. Although high copy number variants (6-7 DAZ and 5-6 CDY copies) were only present in Hg Q1a3a1 and not in R1b1a2, statistical analysis showed no significant differences between both haplogroups regarding overall DAZ and CDY copy number distribution.

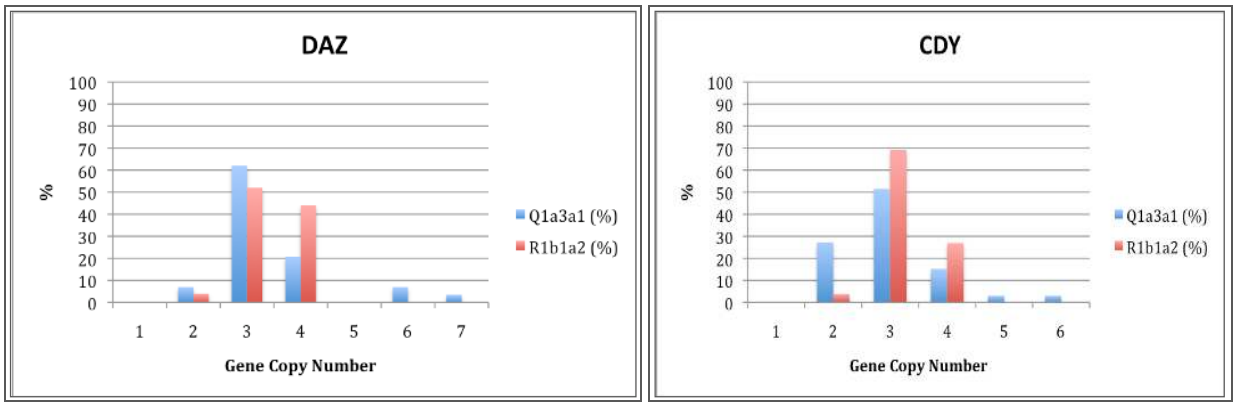


Figure 4.7. *Gene Copy Number for DAZ and CDY in haplogroup Q1a3a1 (blue) and R1b1a2 (red)*

5. DISCUSSION

The results included in this Master Thesis Dissertation allowed characterizing fertile men belonging to Native American haplogroup Q1a3a1 regarding AZFc microdeletions and CDY and DAZ gene copy number.

From the unrelated male samples collected in the Genetic Fingerprinting Service, along the sampling period considered (from January to July 2011) approximately 500 men have been analyzed for paternity assessment. The analyzed samples were submitted to Y chromosome haplogroup screening as a routine to identify, mainly, samples belonging to Native American haplogroup Q1a3a1. Simultaneously, samples belonging to the GenBank´s reference haplogroup R have been also screened and selected. As previously said, in the Argentinean population the percentage of men belonging to haplogroup Q1a3a1 is about 5%. It should be noted that from the paternity casework performed in our population, not all alleged fathers turn out to be the biological fathers. With all this in mind, from our sample reservoir, it was possible to select 37 samples from confirmed fathers and simultaneously belonging to haplogroup Q1a3a1. The present was the first study of the kind performed in Argentinean population and also the only study, until the present, were fertile controls have been characterized by genetic paternity testing analysis with autosomal STRs.

The development of the Real Time quantification assay for DAZ and CDY copy number was first accomplished with primate samples. With the approach described above, it was possible to confirm the FISH results in primate evolution studies. The developed technique become part of a manuscript that is already submitted (Greve et al.) and was the trigger for this Master Thesis work on human Y chromosomes belonging to haplogroup Q1a3a1, which has not been studied in deep until the present. The hypothesis of this work was based on the fact that in primates and in humans the MSY was described to suffer microdeletions and copy number variants that are not always related to male infertility and this variation can be found also in normal fertile males (Fernandes, Paracchini et al. 2004; Yang, Ma et al. 2010).

To be able to assess copy number of DAZ and CDY genes, it was necessary to use a reference single copy gene (SRY) that was amplified in each reaction and a calibrator sample with known copy number of these genes. For this matter, a sample belonging to haplogroup R1b1a2 was used under the assumption that both DAZ and CDY are found in 4 copies on the reference Y chromosome. The calibrator sample was also tested for paternity confirmation and DAZ and CDY genes were screened with FISH showing a normal Y chromosome hybridization pattern compatible with previously published results (Wimmer et al 2002, Röttger et al 2000, Kühl et al 2001).

The results obtained by qPCR for DAZ and CDY evidenced that the distribution of copy numbers of these genes is not significantly different between the studied haplogroups, nevertheless it is mandatory to confirm these results by further studies using another molecular biology technique (i.e. Sequence Family Variants to identify which copy of the gene is present), because unexpected

results came out indicating that DAZ could be present not only in an even number of copies, as it is deduced from its duplet palindromic structure in the reference human Y chromosome.

To identify whether microdeletions are found in the AZF region of the Y chromosome, sY1261, sY1191, sY1291, sY1206 and sY1201 STSs have been used. Almost all the analyzed STSs are located in the AZFc region with the exception of sY1206 that is located in palindrome P5 together with 2 copies of CDY2 gene. Y-STSs analysis of 37 samples belonging to haplogroup R1b1a2 and 37 samples belonging to haplogroup Q1a3a1 showed no evidence of any differential deletion of these Y-STSs markers between the two haplogroups.

Nevertheless, one intriguing characteristic was found in sY1291 marker: it evidenced a differential size between haplogroup Q1a3a1 (517pb) and R1b1a2 (538pb). This size difference of 21bp was consistent within all the analyzed samples from both haplogroups and was confirmed by sequencing analysis. This finding was in concordance with the results described by Lin, Hsu et al. (2006) where a length difference was first characterized for the sY1206 marker but not linked to any haplogroup or population.

Another highlighting result was that in none of the 37 samples from haplogroup Q1a3a1 the b2/b3 deletion was detected as described by Repping et al. (2004, 2006) for this haplogroup. According to the mentioned studies, haplogroup Q1a3a1 presents a c35 architecture (b2/b3 deletion) in which a deletion of sY1191 was detected in the AZFc region. To confirm our results, the amplification products for sY1191 have been sequenced and showed complete homology to the reference sequence. As described by Repping et al. this deletion was also found in haplogroup N. Owing this information, a sample belonging to haplogroup N (from a previous study, Corach et al. 2010) was also amplified and the sY1191 deletion described was confirmed for this haplogroup. It is important to highlight that both studies used the same primer sequences to amplify this marker, but differences rely on the number of Q1a3a1 (derivative state for M3) samples used (37 vs. 3) and the paternity confirmation (samples analyzed by Repping et al. has an unknown spermatogenic phenotype and DNA was from EBV-transformed lymphoblastoid cell lines). For this matter it is important to further characterize other haplogroups in order to identify the branch of the Y haplogroup tree were this deletion occurred.

On the other hand, an unexpected finding in this study was a SNPs variant found on the distal copy of the sY1206 marker that showed a G>T variation (according to the reference sequence) found in haplogroup Q1a3a1 and not in R1b1a2. This SNP has already been described in GenBank (rs1297904), but the described variation was referred to a sequence variation between the proximal and distal copies of sY1206 (Figure 4.5.3). The present study was the first, to our knowledge, to identify the T variant in the distal copy and to associate it with Native American haplogroup Q1a3a1. Nevertheless, it will be interesting to analyze whether the G>T variation occurred in Hg Q1a3a1 or the T>G variation occurred in Hg R1b1a2.

One particular result was found in only one sample from haplogroup Q1a3a1 and it was the identification of two heteroplasmic sites in marker sY1206 in the positions 290 and 291. This result may indicate that in one of the copies of this marker a variation has occurred or that, by chance, a variation in one copy occurred next to the variation in the other copy. Further analysis will be needed to elucidate which of the two mechanisms produced this particular heteroplasmic profile.

From the 74 samples analyzed, only 2 had the gr/gr deletion detected by the absence of sY1291 marker and this deletion was found equally in both haplogroups. Taking into account that all analyzed samples had certainly fathered a child, these results indicate that the gr/gr deletion is not deleterious to male fertility according to what has been previously described (Repping et al 2003, Hucklenbroich 2005, Shahid 2011, Carvalho 2006).

6. CONCLUSIONS & FUTURE PERSPECTIVES

Our initial hypothesis affirming that individuals belonging to Native American haplogroup Q1a3a1 might have suffered variations on the AZF region of the Y chromosome and therefore present characteristic polymorphisms was confirmed, but still needs further analysis. Surprisingly, differences among the studied haplogroups were not the ones expected, but they arose from further characterization of the analyzed STS markers.

This study was the first to analyze the possible association between Y chromosome haplogroups and rearrangements on the MSY in normal fertile Argentinean male population and it should, and will, be followed by analysis of infertile patients together with the characterization of semen parameters.

As Argentina has an admixed population, with European and Native American genetic components, thus association studies must be carried out considering the Y chromosomal origin of the individual. Regarding infertility diagnosis and proper counseling, haplogroup assessment and following proper diagnosis should be developed. Meanwhile there have been numerous association studies in European and other populations between Y chromosome rearrangements, haplogroup and patient´s phenotype, no such studies have been performed in Native American population.

To continue with the present research, it is needed to identify SFVs (sequence family variants) of DAZ and CDY genes in order to clarify which of the copies of DAZ/CDY has been deleted or duplicated (Machev, Saut et al. 2004; Yang, Ma et al. 2010). For this purpose, a molecular genetics approach, using Real Time PCR and High Resolution Melting (HRM) analysis, aiming to identify the SNPs in these genes can be easily set up. Similar diagnostic methods have been already described (Shahid, Dhillon et al. 2011) but neither by means of HRM analysis.

On the other hand, more Y-STSs markers should be analyzed to narrow the deletion´s site and to identify other polymorphic sites.

Furthermore, although DAZ and CDY are the more studied genes in male fertility, other genes located in the AZF region may play an important role and had not been studied yet.

The main conclusion of this Master Thesis Dissertation is that this is just "the tip of the iceberg" and there is a lot of work to do in this field and a multidisciplinary context should be suggested for this matter.

“I never see what has been done;

I only see what remains to be done.”

Buddha

ACKNOWLEDGMENTS

The present Master Thesis work was performed almost entirely in the Genetic Fingerprinting Service, School of Pharmacy and Biochemistry, University of Buenos Aires; except for the probe preparation and qPCR development that took place in the Zytogenetik Lab, Institut für Humangenetik, University of Freiburg, and the FISH analysis that took place in the Neuro and Cytogenetic Lab, University of San Martín, Buenos Aires, Argentina. I would like to acknowledge Christine Hodler from the Institut für Humangenetik, and Dr. Claudia Perandones, Dr. Martin Radrizzani, and Veronica Farini, from the Neuro and Cytogenetic Lab of the San Martín University, for their support with FISH experiments. I would also like to acknowledge Luciano M. for being the control sample donor. Finally, I would like to acknowledge the University of Buenos Aires, the German Academic Exchange Service (DAAD), and Biothera-Foundation for the financial support during my participation in the Master Program.

REFERENCE BIBLIOGRAPHY

Arredi, B., A. Ferlin, et al. (2007). "Y-chromosome haplogroups and susceptibility to azoospermia factor c microdeletion in an Italian population." J Med Genet **44**(3): 205-8.

Balaresque, P., G. R. Bowden, et al. (2008). "Dynamic nature of the proximal AZFc region of the human Y chromosome: multiple independent deletion and duplication events revealed by microsatellite analysis." Hum Mutat **29**(10): 1171-80.

Bortolini, M. C., F. M. Salzano, et al. (2003). "Y-chromosome evidence for differing ancient demographic histories in the Americas." Am J Hum Genet **73**(3): 524-39.

Cavalli-Sforza, L. L. and M. W. Feldman (2003). "The application of molecular genetic approaches to the study of human evolution." Nat Genet **33 Suppl**: 266-75.

Corach, D., O. Lao, et al. (2010). "Inferring continental ancestry of argentineans from Autosomal, Y-chromosomal and mitochondrial DNA." Ann Hum Genet **74**(1): 65-76.

Ferlin, A., B. Arredi, et al. (2006). "Genetic causes of male infertility." Reprod Toxicol **22**(2): 133-41.

Ferlin, A., B. Arredi, et al. (2007). "Molecular and clinical characterization of Y chromosome microdeletions in infertile men: a 10-year experience in Italy." J Clin Endocrinol Metab **92**(3): 762-70.

Fernandes, S., S. Paracchini, et al. (2004). "A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N." Am J Hum Genet **74**(1): 180-7.

Greve, G., E. Alechine, et al. (submitted). "Y-Chromosome Variation in Hominids: Intraspecific Variation is Limited to the Polygamous Chimpanzee."

Hargreave, T. B. (2000). "Genetic basis of male fertility." Br Med Bull **56**(3): 650-71.

ISOGG (2011). "International Society of Genetic Genealogy (2011). Y-DNA Haplogroup Tree 2011."

Jobling, M. A. (2008). "Copy number variation on the human Y chromosome." Cytogenet Genome Res **123**(1-4): 253-62.

Jobling, M. A. and C. Tyler-Smith (2003). "The human Y chromosome: an evolutionary marker comes of age." Nat Rev Genet **4**(8): 598-612.

Karafet, T. M., F. L. Mendez, et al. (2008). "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree." Genome Res **18**(5): 830-8.

Kitchen, A., M. M. Miyamoto, et al. (2008). "A three-stage colonization model for the peopling of the Americas." PLoS One **3**(2): e1596.

Kirkman, T.W. (1996) Statistics to Use.http://www.physics.csbsju.edu/stats/ (16 Aug 2011)

Krausz, C., C. Giachini, et al. (2009). "Phenotypic variation within European carriers of the Y-chromosomal gr/gr deletion is independent of Y-chromosomal background." J Med Genet **46**(1): 21-31.

Kuhl, H., S. Rottger, et al. (2001). "Loss of the Y chromosomal PAR2-region in four familial cases of satellited Y chromosomes (Yqs)." Chromosome Res **9**(3): 215-22.

Kuroda-Kawaguchi, T., H. Skaletsky, et al. (2001). "The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men." Nat Genet **29**(3): 279-86.

Lin, Y. W., C. L. Hsu, et al. (2006). "A two-step protocol for the detection of rearrangements at the AZFc region on the human Y chromosome." Mol Hum Reprod **12**(5): 347-51.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-8.

Machev, N., N. Saut, et al. (2004). "Sequence family variant loss from the AZFc interval of the human Y chromosome, but not gene copy loss, is strongly associated with male infertility." J Med Genet **41**(11): 814-25.

Munch, C., S. Kirsch, et al. (2008). "Evolutionary analysis of the highly dynamic CHEK2 duplicon in anthropoids." BMC Evol Biol **8**: 269.

Navarro-Costa, P., J. Goncalves, et al. (2010). "The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility." Hum Reprod Update **16**(5): 525-42.

Navarro-Costa, P., C. E. Plancha, et al. (2010). "Genetic dissection of the AZF regions of the human Y chromosome: thriller or filler for male (in)fertility?" J Biomed Biotechnol **2010**: 936569.

Noordam, M. J., G. H. Westerveld, et al. (2011). "Gene copy number reduction in the azoospermia factor c (AZFc) region and its effect on total motile sperm count." Hum Mol Genet **20**(12): 2457-63.

Oklander, L. I., G. E. Zunino, et al. (2007). "Isolation, characterization and evaluation of 11 autosomal STRs suitable for population studies in black and gold howler monkeys Alouatta caraya." Molecular Ecology Notes **7**(1): 117-120.

Puzuka, A., N. Pronina, et al. (2011). "Y chromosome--a tool in infertility studies of Latvian population." Genetika **47**(3): 394-400.

Repping, S., H. Skaletsky, et al. (2003). "Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection." Nat Genet **35**(3): 247-51.

Repping, S., S. K. van Daalen, et al. (2006). "High mutation rates have driven extensive structural polymorphism among human Y chromosomes." Nat Genet **38**(4): 463-7.

Repping, S., S. K. van Daalen, et al. (2004). "A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region." Genomics **83**(6): 1046-52.

Rottger, S., J. Pasantes, et al. (2000). "Familial mosaicism of del(Y) and inv del(Y)." Cytogenet Cell Genet **91**(1-4): 208-11.

Rozen, S., H. Skaletsky, et al. (2003). "Abundant gene conversion between arms of palindromes in human and ape Y chromosomes." Nature **423**(6942): 873-6.

Shahid, M., V. S. Dhillon, et al. (2011). "Associations of Y-chromosome subdeletion gr/gr with the prevalence of Y-chromosome haplogroups in infertile patients." Eur J Hum Genet **19**(1): 23-9.

Skaletsky, H., T. Kuroda-Kawaguchi, et al. (2003). "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes." Nature **423**(6942): 825-37.

Taylor, K., N. Hornigold, et al. (1996). "Mapping the human Y chromosome by fingerprinting cosmid clones." Genome Res **6**(4): 235-48.

Tiepolo, L. and O. Zuffardi (1976). "Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm." Hum Genet **34**(2): 119-24.

Underhill, P. A., L. Jin, et al. (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.

Vogt, P. H. (2004). "Genomic heterogeneity and instability of the AZF locus on the human Y chromosome." Molecular and Cellular Endocrinology **224**(1-2): 1-9.

Vogt, P. H. (2005). "AZF deletions and Y chromosomal haplogroups: history and update based on sequence." Hum Reprod Update **11**(4): 319-36.

Vogt, P. H., A. Edelmann, et al. (1996). "Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11." Hum Mol Genet **5**(7): 933-43.

WHO "WHO: http://www.who.int/topics/infertility/en/."

Wimmer, R., H. Kuhl, et al. (2002). "Comparative mapping of CDY and DAZ in higher primates." Cytogenet Genome Res **96**(1-4): 287-9.

Yang, Y., M. Ma, et al. (2010). "Differential effect of specific gr/gr deletion subtypes on spermatogenesis in the Chinese Han population." Int J Androl **33**(5): 745-54.

Yang, Y., M. Ma, et al. (2008). "Y chromosome haplogroups may confer susceptibility to partial AZFc deletions and deletion effect on spermatogenesis impairment." Hum Reprod **23**(9): 2167-72.

YCC (2002). "A nomenclature system for the tree of human Y-chromosomal binary haplogroups." Genome Res **12**(2): 339-48.

Zuccarelli, G., E. Alechine, et al. (2011). "Rapid screening for Native American mitochondrial and Y-chromosome haplogroups detection in routine DNA analysis." Forensic Sci Int Genet **5**(2): 105-8.