**DAZ gene copies: evidence of Y chromosome evolution**

Ana Teresa Fernandes¹,⁵, Susana Fernandes², Rita Gonçalves¹, Rosália Sá²,³, Paula Costa², Alexandra Rosa¹, Cristina Ferrás², Mário Sousa²,³,⁴, António Brehm¹ and Alberto Barros²,⁴

¹Laboratory of Human Genetics, University of Madeira, Funchal, Madeira, ²Department of Genetics, Faculty of Medicine, ³Lab Cell Biology, ICBAS, University of Porto and ⁴Centre for Reproductive Genetics Alberto Barros, Porto, Portugal

²To whom correspondence should be addressed at: Laboratory of Human Genetics, University of Madeira, Campus da Penteada, 9000-390 Funchal, Madeira, Portugal. E-mail: atgf@uma.pt

The DAZ gene, a contributing factor in infertility, lies on the human Y chromosome’s AZFc region, whose deletion is a common cause of spermatogenic failure. Y chromosome binary polymorphisms on the non-recombining Y (NRY) region, believed to be a single occurrence on an evolutionary scale, were typed in a sample of fertile and infertile men with known DAZ backgrounds. The Y single-nucleotide polymorphisms (Y-SNPs) with low mutation rates are currently well characterized and permit the construction of a unique phylogeny of haplogroups. DAZ haplotypes were defined using single-nucleotide variant (SNV)/sequence tagged-site (STS) markers to distinguish between the four copies of the gene. The variation of 10 Y chromosome short tandem repeat (STRs) was used to determine the coalescence age of DAZ haplotypes in a comparable time frame similar to that of SNP haplogroups. An association between DAZ haplotypes and Y chromosome haplogroups was found, and our data show that the DAZ gene is not under selective constraints and its evolution depends only on the mutation rate. The same variants were common to fertile and infertile men, although partial DAZ deletions occurred only in infertile men, suggesting that those should only be used as a tool for infertility diagnosis when analysed in combination with haplogroup determinations.

**Key words:** DAZ/infertile/phylogeny/Y chromosome

**Introduction**

The human Y chromosome is strictly paternally inherited and, in most of its length, does not engage in pairing and crossing over during meiosis (Lahn and Page, 1997). Deletions in the Y chromosome AZFc region are associated with male infertility and occur de novo at a frequency of about 1/4000, being the most common known cause of spermatogenic failure (Reijo et al., 1995; Vogt et al., 1996). Within the AZFc region, there are six distinct families of massive repeat units (amplicons) organized in three palindromes (Kuroda-Kawaguchi et al., 2001; Skakletsky et al., 2003).

The DAZ gene family, located in the AZFc region, is organized into two clusters and contains a variable number of copies (Glaser et al., 1998; Saxena et al., 2000; Fernandes et al., 2004; Lin et al., 2005). Several mechanisms were suggested to explain the differences concerning the relative number of DAZ copies, such as intrachromosomal recombination events between amplicons in case of deletions (Repping et al., 2003; Vogt and Fernandes, 2003), gene conversion within palindromes to justify the duplications (Jobling and Tyler-Smith, 2003; Graves, 2004) and inversions like those originating in palindromic sequences (Kuroda-Kawaguchi et al., 2001; Repping et al., 2004). Gene conversion events, considered to be frequent when compared with base substitutions, might also influence the probability of rearrangements (Skakletsky et al., 2003), therefore favouring restoration of the original sequence and protecting the Y chromosome from the degeneration potentiated by haplody (Hawley, 2003; Rozen et al., 2003; Graves, 2004).

The DAZ gene was derived from the autosomal homologue DAZL (DAZ gene-like) by transposition of the 3p24 chromosomal section to the Y chromosome about 40 million years ago (Reijo et al., 1995; Saxena et al., 1996; Seboun et al., 1997), subsequent to the splitting of the Old and New World monkeys (Seboun et al., 1997; Kumar and Hedges, 1998). Unlike DAZL, a gene that remains as a single copy, DAZ appears with a variable number of copies with 99.9% homology (Kuroda-Kawaguchi et al., 2001; Skakletsky et al., 2003). The variation of each DAZ gene copy seems to be essentially due to a tandem amplification of exon 7 (Saxena et al., 2000; Vogt and Fernandes, 2003). Deletions in DAZ2, DAZ3 and DAZ4 copies, found in fertile and infertile men, are described as familial variants inherited from father to son (Vogt et al., 1996; Saxena et al., 2000; Fernandes et al., 2002, 2004), although DAZ1/2 deletions are restricted only to infertile men (Fernandes et al., 2002; Vogt and Fernandes, 2003; Ferlin et al., 2005). Comparisons between the two DAZ sequences available from Genbank show that only DAZ1 exhibits a conserved structure, suggesting that it might be essential for human spermatogenesis (Vogt and Fernandes, 2003; Ferrás et al., 2004), although one case of a fertile man with a DAZ1 deletion was recently described (Machev et al., 2004). DAZ haplotypes were previously defined using single-nucleotide variants (SNVs/sequence tagged-site (STSs) to distinguish the four copies of the DAZ gene (Fernandes et al., 2002), and it has been shown to be a reliable method to identify each DAZ gene copy (Figure 1).

There are now more than 200 well-characterized non-recombining Y (NRY) biallelic markers, all having low mutation rates. These are assumed to be unique events in the evolutionary process, whose hierarchical structure allows the construction of a unique phylogeny of haplogroups (Y Chromosome Consortium, 2002). NRY markers are considered to be neutral, evolving only because of mutation rate (Jobling and Tyler-Smith, 2003). The largest fraction of extant European Y chromosome pool is composed of the I and R haplogroups thought
to be surviving lineages of Paleolithic origin. These haplogroups are believed to have expanded in the post-glacial period, after a severe bottleneck during the Last Glacial Maximum (20 000 years ago) (Semino et al., 2000). Haplogroups E and J, which represent approximately 20% of Europeans, were introduced in southern Europe from the Near East by immigrant farmers, during the Neolithic expansion 10 000 years ago (Semino et al., 2004). A study suggested that a weak negative selection due to partial deletion of genes needed for spermatogenesis could act on some haplogroups, namely haplogroup D2 (Repping et al., 2003). Otherwise, the frequencies of matriarchal lineages defined by mitochondrial DNA and the paternal counterpart identified by Y chromosome haplogroups are similar in Europe (Richards et al., 2000; Semino et al., 2000), corroborating the theory of no selection on the Y chromosome structure.

Materials and methods

Samples
DNA samples from 97 infertile men presenting with varying clinical syndromes ranging from Sertoli-cell-only syndrome (SO) to maturation arrest (MA) and hypospermatogenesis (HP) were analysed. All patients displayed a spectrum ranging from Sertoli-cell-only syndrome (SO) to maturation arrest (MA) and were introduced in southern Europe from the Near East by immigrant farmers, during the Neolithic expansion 10 000 years ago (Semino et al., 2004). A study suggested that a weak negative selection due to partial deletion of genes needed for spermatogenesis could act on some haplogroups, namely haplogroup D2 (Repping et al., 2003). Otherwise, the frequencies of matriarchal lineages defined by mitochondrial DNA and the paternal counterpart identified by Y chromosome haplogroups are similar in Europe (Richards et al., 2000; Semino et al., 2000), corroborating the theory of no selection on the Y chromosome structure.

Y chromosome haplogroups

The STSs containing Y-SNPs were assayed as described (Underhill et al., 2000, 2001). Genotyping was performed using both native and engineered restriction fragment length polymorphism (RFLP) methods. The SNPs analysed are shown in their phylogenetic order defining the haplogroup status of each Y chromosome. The binary markers M4, M61, M147, LLY22g, M175 and P36 were assayed, but derived alleles were not observed. The haplogroup nomenclature and phylogeny used was the one proposed by the Y chromosome Consortium (2002).

SNV/STSs PCR for partial DAZ deletions

DAZ haplotypes, using SNVs/STSs to distinguish among the four copies of DAZ gene, were determined in all fertile and infertile samples through the analysis of six DAZ-single-nucleotide variants (SNV1–VI) and two DAZ-STS (DAZ-RRM3 and Y-DAZ3), as previously described (Fernandes et al., 2002).

Coalescence time of DAZ haplotypes

To define the coalescence time within each DAZ haplotype, we assayed 10 short tandem repeat (STR) (Powerplex Y system kit, Promega, Madison, WI, USA) in chromosomes belonging to the different DAZ types, using standard methodology. The coalescence time of STR variation was estimated as the average squared difference in the number of repeats between all current chromosomes and the founder haplotype (assumed to be modal), divided by $w = 6.9 \times 10^{-4}$ per 25 years (Zhivotovsky et al., 2004). The coalescence time was expressed in terms of thousands of years into the past (Kya), and the minimum sample size considered for coalescence age calculation was six individuals.

Statistical analysis

Frequency differences were calculated according to Fisher’s exact test using the Arlequin software (Schneider et al., 2000), for fertile, infertile and also other samples from northern Portugal (no accession to the fertility state) (Gonçalves et al., 2005). The same procedure was used to compare the frequency of R haplogroup on fertile and infertile men, as well as those with complete AZFc deletion, and to compare the number of deletions within haplogroups in infertile men. Probability values of $P < 0.05$ were considered as statistically significant.

Results

The frequency distribution of Y chromosome haplogroups in fertile and infertile men is shown in Figure 2. Comparisons of haplogroup frequencies within fertile and infertile men showed no significant differences. Similar results were obtained when both groups were compared with the reference population from the same region of northern Portugal. Our data indicate a clear association of DAZ haplotypes and Y chromosome haplogroups (Table I) in accordance with previous studies, suggesting a possible association between AZFc partial deletions and Y chromosome variant haplogroups (Repping et al., 2003; Vogt and Fernandes, 2003; Machev et al., 2004; Repping et al., 2004; Vogt, 2005). As an example, the 2*4d pattern was exclusively found in haplogroup I, in fertile and infertile individuals. Moreover, the 4d pattern is common to the phylogenetically close haplogroups G, K, J1 and P, placing its occurrence at a time before divergence. All fertile individuals from haplogroups I and J2 display a characteristic DAZ haplotype.

The coalescence time for the more common DAZ haplotypes is shown in Figure 3 and is in accordance with that given for the associated haplogroups (Semino et al., 2000).

Of 19 individuals with complete AZFc deletion, the frequency of the R haplogroup was determined to be similar to that found in fertile and infertile men.

When we compared the number of deletions in infertile men within each haplogroup, significantly more deletions were found in haplogroup J1 than in haplogroup R ($P < 0.05$).
Figure 2. Phylogenetic tree of Y chromosome haplogroups. Haplogroup defining mutations assayed in this study are shown along the branches. The grey boxes show the relative position of DAZ mutations which leads to the DAZ haplotype.

Table I. Pattern of DAZ haplotype within each haplogroup for fertile ($n = 91$) and infertile men ($n = 97$)

<table>
<thead>
<tr>
<th>Y chromosome haplogroup</th>
<th>Fertile men $n$</th>
<th>DAZ haplotype</th>
<th>Infertile men $n$</th>
<th>DAZ haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1b</td>
<td>No deletion 53</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>No deletion 43</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>R1*</td>
<td>No deletion 3</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>No deletion 2</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>P</td>
<td>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>K</td>
<td>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>G</td>
<td>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>I</td>
<td>2<em>3</em>4d 1<em>2</em>3</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>2<em>3</em>4d 1<em>2</em>3</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>J1</td>
<td>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>J2</td>
<td>2p<em>4d 2p</em>4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>2d 4p 2d*4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>E3b</td>
<td>2d<em>3</em>4d 2d*4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>2d 4p 2d*4d 4d 4d 4d</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
<tr>
<td>A</td>
<td>0 3*4</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
<td>0 3*4</td>
<td>2d 4p 2d<em>4d 2</em>3<em>4d 2d</em>4d</td>
</tr>
</tbody>
</table>
**Discussion**

The concordance of DAZ haplotypes with the branching and inner variability is therefore an indicator of neutrally driven evolution on the basis of mutation rate alone. If DAZ haplotypes were to be affected by selective constraints, as a strong linkage to infertility or founder effect and/or bottlenecks phenomena, their expected level of variability at the fast-evolving microsatellite (and consequently the coalescence time) would be smaller.

For the occurrence of specific DAZ partial deletions only in infertile men, two scenarios are possible: (i) they are connected to male infertility or (ii) they are within the normal variation range for the haplogroup but escaped the sampling. Care is needed when associating DAZ haplotypes with a causal factor of male infertility. For example, haplotype 2d*4d, present in infertile men from haplogroup G and J1, is in fact part of the normal variation found in both E3b case-study groups. Therefore, we suggest that a higher reliability for an association of DAZ haplotype-infertility demands the knowledge of their haplogroups. Nevertheless, there must be some weakness in the Y chromosome structure; if not, the number of deletions in infertile men from different haplogroups would be the same and no differences such as the ones existing between J1 and R haplogroups would have been observed.

On the contrary, the frequency of R haplogroup compared with those individuals with complete AZFc deletions was similar to both groups of fertile and infertile men. Therefore, the chromosome structure on the R haplogroup does not seem to protect against the deletion, despite palindromic sequences being more stable because of the reduced possibility of recombination (Kuroda-Kawaguchi et al., 2001; Fernandes et al., 2004).

Although not in the context of our study, it is possible that the distribution and frequencies of each DAZ haplotype may vary with geography and the demographic history of populations, as observed for the population perspective of the Y chromosome genetic system. The high frequency of the variant DAZ ‘no deletions’ can reflect the exclusive relation to haplogroup R, the most common in Europe (Semino et al., 2000; Jobling and Tyler-Smith, 2003). It therefore seems clear that the Y chromosome inversion described (Kuroda-Kawaguchi et al., 2001), based on the Genebank sequence of the subject with reference RPCI-11 (Vogt and Fernandes, 2003; Vogt, 2005), only exists in individuals belonging to the haplogroup R. Consequently, the DAZ haplotype ‘no deletions’ can no longer be regarded as a correct designation because it only represents a reference and not an ancestral. The oldest DAZ haplotype that represents the absence of the distal part of the DAZ gene copy 4 is the DAZ 4d haplotype. The distal part is shown to be present only in the sequence of the subject with reference RPCI-11 and not in the sequence of CTA/CTB men (Saxena et al., 2000; Vogt and Fernandes, 2003). The same might be observed in mitochondrial DNA lineages, where the Cambridge reference sequence (CRS) is not the ancestral but the reference obtained by the complete sequence (Anderson et al., 1981), revised in 1999 (Andrews et al., 1999).

In conclusion, we demonstrated that there is an association between the DAZ haplotype and the haplogroups of the Y chromosome. As DAZ partial deletions might be a polymorphic event associated with a specific haplogroup or an individual cause of infertility, those should only be analysed for infertility diagnosis when analysed in combination with haplogroup determinations. The possibility that a mutation defining a haplogroup could be more than only a single mutation—for instance, being associated with Y chromosome rearrangements—therefore remains open for debate.

**Acknowledgements**

We are grateful to David Shaproski and Christoph Roehrig for their help in preparing the final version of this manuscript. This study was partially supported by FCT (SRFH/BPD/23616/05, 6664/01; POCI/SAU-MM/06709/04, 60555/04, 59997/04; UMIB).

**References**


Submitted on February 7, 2006; resubmitted on May 8, 2006; accepted on May 11, 2006.