

DYF399S1: A Unique Three-Copy Short Tandem Repeat on the Human Y Chromosome

Gareth Ll. C. Henson

A recent paper (Kayser et al. 2004) reported 166 new Y Chromosome STR markers. The present paper reviews the data available for one of these markers and suggests that it would be a useful addition to the markers currently tested for genealogical purposes.

Introduction

Basic information concerning the short tandem repeat (STR) DYF399S1 is given on the on-line supplementary information file for the above paper, on line 168. Here it is supplemented with information from the Y chromosome reference sequence.

DYF399S1 is a tetranucleotide repeat with the following structure:



where n is the variable element of the repeat. The specified primers give a PCR product range of 277-305 on the samples tested, corresponding to a range for n of 14 to 21.

DYF399S1 is located in the AZFc region of the Y chromosome, the region containing numerous duplicated sections of DNA arranged into palindromes or mirror images (Kuroda-Kawaguchi et al. 2001). Uniquely amongst the known and novel STRs, there are normally 3 copies, reflecting the asymmetry between palindromes P1 and P2. The three copies are located in the "green" amplicon sections g1, g2 and g3. DYF399S1's nearest known STR neighbors are 3 of the 4 copies of DYS464 located in the "red" amplicon sections r1 r3 and r4. The fourth copy of DYS464 is in the r2 section which does not have an adjacent green section (see Fig. 1).

Test results for 8 samples and comparison with the Y reference sequence

The novel STRs were tested in 8 samples from 8

Received February 18, 2005

Address for correspondence: Gareth Ll. C. Henson, gareth.henson@breathe.com

different haplogroups and the published results for DYF399S1, with deduced repeat lengths, are as in Table 1 (sizes in ascending order, except the reference sequence copies which are in the order of the sequence). The repeat lengths are conjectural as multicopy STRs were not sequenced and so insertions/deletions in the surrounding DNA cannot be ruled out; however all the product sizes differ from each other by multiples of four which is consistent with all the variation being in the long repeat block.

The reference sequence copy in g1 has a base G insertion giving a STR sequence of



a product size of 302 and an allele of 20.1. There is also an A to G mutation immediately following the repeat block in the reference sequence copy in g2. Only further testing will determine whether these or similar variations are peculiarities of the reference sequence, are common in its own haplogroup or are recurring mutations across the range of the Y chromosome tree.

Table 1
Published Test Results for DYF399S1

Haplogroup	PCR product size	Alleles
A	289, 293, 297	17, 18, 19
B	285, 289, 293	16, 17, 18
C	281, 289, 293	15, 17, 18
E	285, 289, 293	16, 17, 18
I	277, 289, 297	14, 17, 19
J	277, 281, 285, 293	14, 15, 16, 18
K	281, 293	15, 18
R (sample)	281, 301, 305	15, 20, 21
R (ref seq)	302, 289, 293	20.1, 17, 18

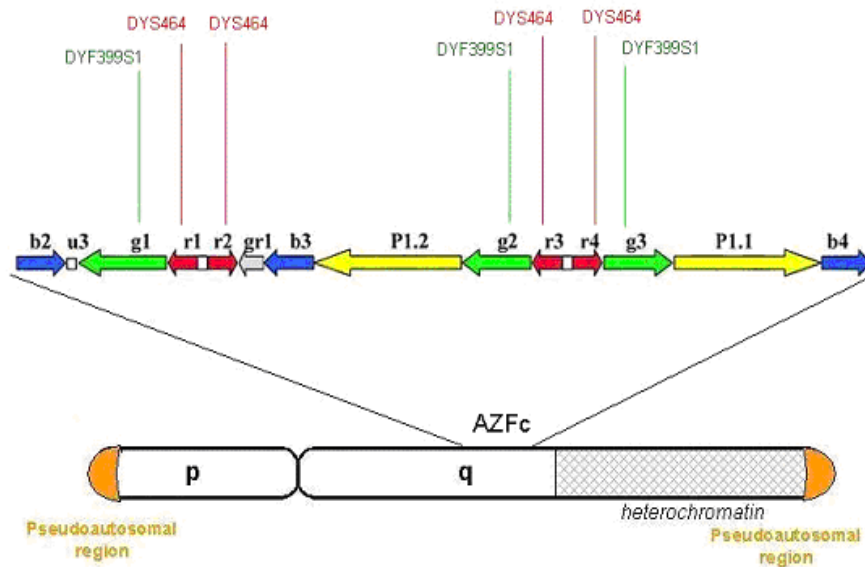


Figure 1 Positions of DYF399S1 and DYS464 Along the Y Chromosome. The arrows indicate the directionality of the duplicated sections. Adapted from Fig 1 of Fernandes et al 2004 and the diagram at <http://www.cstl.nist.gov/biotech/strbase/ystrpos1.htm>

Mutational Properties of DYF399S1

The available information, although limited, suggests DYF399S1 is a highly polymorphic marker, as all the compound results are distinct apart from 16,17,18 which is shared by haplogroups B and E. This compound type is one step away from that of the haplogroup C sample, all other pairwise comparisons (except possibly C/K) have distances of at least 2 steps. All samples except two have 3 distinct alleles, only the sample for haplogroup K has just 2 distinct values (which may mean one of these occurs twice, rather than a missing copy). The sample for hg J has 4 distinct alleles which implies an additional copy; this is likely to have been caused by a duplication event similar to those which produce additional copies for DYS464.

As well as being similar to DYS464 due to its location, DYF399S1 shares a similar repeat structure (Redd et al 2002). They both have tetranucleotide repeat motifs arranged in complex sequences comprising short invariant blocks and 1 longer variable block. The allele ranges of the two loci overlap, DYS464 having a range of 9 to 20 compared with DYF399S1's range of 14 to 21. This suggests DYF399S1 has a similar mutation rate. DYS464 is known to be a fast mutating polymorphic marker (Redd et al 2002, Berger et al

2003), well suited to genetic genealogy purposes; from the data published so far it seems likely that DYF399S1 would be similarly useful.

Another useful feature of DYS464 is that the compound haplotype usually gives a good indication of the haplogroup. The data for DYF399S1 is consistent with different haplogroups also having characteristic patterns for this marker. Curiously in both markers the highest values are found in haplogroup R. This may be coincidence or it may indicate something about the evolutionary history of this haplogroup (or the subgroup from which the samples came, probably R1b as both the tested sample and the reference sequence have DYS392 = 13).

Usefulness of Three Copies

DYF399S1 appears to be the only known marker which normally has 3 copies. This is useful because it means fewer ambiguous test results than with even number copy markers. For example with 2 copy markers (e.g. DYS385, YCAII, DYS459) it is often difficult to determine whether a sample with a single PCR product size has two identical copies or just one copy, the other having been deleted. Similarly with the 4 copy DYS464 where there are just one or two product sizes, it is difficult to distinguish an AABB pattern from AB00 (where 2

copies have been deleted). With a 3 copy marker if only 2 alleles are present, either one will be twice as common as the other (indicating an AAB or ABB pattern) or both will have the same frequency (indicating a deleted or possibly duplicated copy). Only when just one allele is detected will there be ambiguity between zero, one or two deletions.

When ambiguous results occur DYS464 and DYF399S1 can complement each other by indicating which of the possible interpretations is most likely. For example if DYS464 has just two observed values, then if DYF399S1 has its usual 3 values, it is unlikely that a deletion has occurred in DYS464 (and so AABB is the probable allele pattern). However if DYF399S1 has just two values with equal frequency (or just one value) it is more likely that a deletion has occurred in DYS464. Conversely if DYF399S1 has just two equally frequent values, then if DYS464 has more than its usual 4 values it is likely that a duplication has also occurred in DYF399S1 (i.e. an AABB pattern) but if DYS464 has fewer than 4 distinct values it is more likely that a deletion has occurred (i.e. an AB pattern).

Known Deletion/Duplication Patterns and Expected Results for DYF399S1

Descriptions of deletions in the AZFc region are given by Repping et al 2003 and 2004 and Fernandes et al. 2004. The expected results for DYF399S1 and DYS464 are shown in Table 2. Most of these deletion patterns are rare. For example Repping et al (2003) found 4 gr/gr deletions in 215 control samples from various Y haplogroups.

Duplications appear to be more common than deletions – additional copies of DYS464 appear in the genealogical databases Y-Search and Y-Base. Details of duplication structures have not been published but it is to be expected that where 5, 6 or 7 copies of DYS464 occur there will also be additional copies of DYF399S1.

It is not normally possible to determine the exact order of multi-copy alleles along the Y chromosome as they are generally embedded in large sections of duplicated DNA, with the exception of DYS385 which lies close to the non-duplicated centre of the P4 palindrome (Kittler et al 2003). Nevertheless an analysis of deletion and duplication patterns within haplogroups and genealogical pedigrees may provide some information as to where particular alleles are located, at least relative to each other.

Summary

- DYF399S1 is the only known three-copy STR marker
- It is likely to be highly polymorphic making it suitable for genetic genealogical testing
- As well as being a useful marker in its own right it will be a complement to DYS464, each marker helping to resolve ambiguous test results for the other and together providing information about deletion/duplication patterns

Table 2
Expected STR Copy Numbers in Known Deletion Types

Deletion type	Copies of DYF399S1	Copies of DYS464
None (reference sequence)	3	4
gr/gr	2	2
gr/gr + b2/b4 dup	4	4
b1/b3	2	2
b2/b3 inv + g1/g3 del	1	2
gr/rg inv + b2/b3 del	1	2
b2/b4 (whole of AZFc)	0	0

Inv = inversion, dup = duplication

Additional Note

This paper was originally prepared before the publication in the *Journal of Medical Genetics* of a letter, "Inadvertent diagnosis of male infertility through genealogical DNA testing" by King et al. The author is aware that the views expressed in the letter regarding *DYS464* apply equally to *DYF399S1*. He believes that the way forward is for genetic genealogists to develop their understanding of structural variations in the Y chromosome and their implications, and that *DYF399S1* will be a useful marker in this process.

Electronic-Database Information

www.ensembl.org	blast search of reference sequences
www.ysearch.org	genetic genealogy database
www.ybase.org	genetic genealogy database
http://www.cstl.nist.gov/biotech/strbase/ystrpos1.htm	Y chromosome diagram

References

- Berger B, Niederstatter H, Brandstatter A, Parson W (2003) Molecular Characterisation and Austrian Caucasian data of the multi-copy Y-chromosomal STR *DYS464*. *Forensic Sci Int* 137:221-230
- Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH (2004) A large AZFc deletion removes *DAZ3/DAZ4* and nearby genes from men in Y Haplogroup N. *Am J Hum Genet* 74:180-187
- Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyiddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C (2004) A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet* 74:1183-1197 ([Supplementary data](#))
- King TE, Bosch E, Adams SM, Parkin EJ, Rosser ZH, Jobling MA (2005) Inadvertent diagnosis of male infertility through genealogical DNA testing. *J Med Genet* 42:366-368
- Kittler R, Erler A, Brauer S, Stoneking M, Kayser M (2003) Apparent intrachromosomal exchange on the human Y chromosome explained by population history. *Eur J Hum Genet* 11:304-314
- Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genetics* 29:279-286
- Redd AJ, Agellon AB, Kearney VA, Contreras VA, Karafet T, Park H, de Knijff P, Butler JM, Hammer MF (2002) Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci Int* 130:97-111
- Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet* 35:247-251
- Repping S, van Daalen SKM, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2004) A family of human Y chromosomes has dispersed throughout northern Europe despite a 1.8-Mb deletion in the Azoospermia Factor c Region. *Genomics* 83:1046-1052