David N. Ewing, M.D.
July 13, 2008

DYF399X Report

Note

This page concerns only DYF399X in the Ewing Project and presupposes a working knowledge of the project and how its groups are organized. Links to our other Results Pages are available in the Results Directory, which can be reached through the link at the top of this page. If you are not acquainted with the Ewing Project, please start with the Help and Introduction links.

Background

DYF399X usually consists of three alleles of a very rapidly mutating marker in a palindromic region of the Y-chromosome, which makes it subject to recLOH events and much complicates interpretation. If you don't understand this first sentence of the report, don't feel like the Lone Ranger—me neither, really. At the end of this page under **References**, I have quoted an email from Thomas Krahn, a scientist with FtDNA, which gives a number of references for the more technically minded among you.

For our purposes, it is sufficient to start with the understanding that DYF399X provides rapidly mutating markers that may help us differentiate branches within the large closely related group of Ewings. DYF399X will only be helpful to men who have already identified a group of genetically closely related men for which they are trying to get better genetic branch resolution. For the Ewing Project, this means that it could be helpful for members of Groups 1, 3, 4, 5 and 7 to have this test. It would not be much help for members of Groups 2, 8 and 9—at least not until some larger clusters have been identified within these groups. The six men in Group 6 might find this test helpful, but we do not think there is enough confusion in Group 6 to justify the small additional expense (about $25), yet. (If you do not know about how these groups are organized and constituted, please click on the Introduction link.)

Results

So far, we have DYF399X results on twenty-four men in the large closely related group of Ewings. (The twenty-fifth, John McEwan in Group 8, is a professional geneticist who is very interested in this stuff and he has had practically every available test, including DYF399X. Note that he has only two DYF399X alleles and is quite different than the men in the closely related group, as would have been expected.)

Group	ID	DYF399X		
	Modal	21t	25/26c	27.1t
1	FI	21t	24c	24t
1	WC3	21t	25c	27.1t
1	WE3	21t	25c	27.1t
1	WR3	21t	25c	27.1t
1	DC2	21t	26c	26.1t
1	JC	21t	26c	26.1t
1	RA	23t	26c	28.1t
1	CA	21t	26c	27.1t
1	MT	21t	26c	27.1t
3	SR	22t	27c	27.1t
3	GR	21t	24c	27.1t
3	DN	21t	24c	25.1t
3	HW	21t	26c	25.1t
4	GW	21t	25c	27.1t
4	RB	21t	25c	27.1t
4	RD	21t	25c	28.1t
5	JM2	21t	26c	27.1t
5	RC	21t	26c	27.1t
5	AL	21t	26c	27.1t
5	JL	22t	26c	27.1t
5	TG	22t	25c	27.1t
7	DL2	21t	25c	27.1t
7	SC	21t	25c	27.1t
7	PA2	21t	25c	25.1t
8	JMc	null	23c	26.1t

Let us see what we can make of these results. First of all, as we hoped and expected, there is quite a bit of diversity in the results on this test—the idea of testing an especially rapidly moving marker is to get more diversity so that we can distinguish otherwise very similar haplotypes. The last time we updated this report a month ago, we said, "…there are seven men that have 25c and nine that have 26c; if we test several more men, the modal could easily shift." Now, there are ten of each, and again we have no modal. It looks like this neck-and-neck race is going to go on indefinitely, and we will need to adduce other considerations than a modal value to speculate about the ancestral value at the middle marker.

- **Group 4** Two of the men in Group 4 (GW & RB) have identical results and a third (RD) has a one step difference from them. Notice that all three of them have 25c at the middle marker.
- **Group 5** So far, all of the Ewing men in Group 5 that have been tested for DYF399X are in Group 5, part 2, and for the most part do not know their

conventional genealogic connection with one another. Three of them (JM2, RC & AL) match one another exactly but have 26c where the men in Group 4 have 25c. A fourth (JL) also has 26c at the middle marker, but is off modal at the first marker, where he has 22t. The fifth Group 5 man (TG) also has 22t at the first marker, but does not have 26c at the middle marker like the others in Group 5, but rather has 25c like the men in Group 4.

- **Group 7** We have results on three of the men in Group 7. DL2 & SC exactly match the Group 4 pattern (or rather, match two of the three men in Group 4 exactly—the middle marker is the one of interest, here). I have an idea that this may be a rather exciting result for those in the Group 7 line, because they have not been able to convincingly connect their line to one of the immigrant ancestors, and if this finding holds up as we get more results, it could develop into some pretty strong evidence that they are in fact connected with the John of Carnashannagh line, as some of them have suspected. PA2 also has 25c at the middle marker, but he is the odd man out at the third marker, where he has 25.1t—a value shared by two of the men in Group 3. I do not think we can make anything of that, except to say that it illustrates the vagaries or working with such a rapidly mutating marker.
- **Group 3** The Group 3 DYF399X results are confusing and not very helpful so far, I'm afraid. DN and SR are seventh cousins once removed, both descended from Alexander b. 1693/4, the eldest son of James Ewing of Inch, and as such, are more closely related to one another than they are to any of the other men in the project. Each is genetic distance one from the Ewing 37-marker modal, but at different markers, so they are genetic distance two from one another, a finding entirely consistent with their conventional genealogies as outlined in the Group 3 Relationship Diagram. Their DYF399X results are quite different—one step at the first marker, three at the second, and two at the third—so not helpful at all in establishing a pattern characteristic of the Alexander b. 1693/4 branch. HW & GR are fourth cousins, both descended from the same grandson of James Ewing of Inch, Samuel Ewing b. c1741. They also do not match one another so well on the DYF399X markers, and as you can see GR and DN have results that are fairly close, notwithstanding that they are in different branches of this family. This is an example of the same problem we were trying to get around by using the DYF399X markers: sometimes more distantly related men have closer results than more closely related men. Distinguishing branches as we hope to do depends on the science, to be sure, but also on luck. And remember, even widely divergent results on DYF399X cannot be used to disprove relationships.
- **Group 1** You will recall that Ewing Group 1 consists of the men in the closely related group of Ewings who do not know their conventional genealogic connection with the others. Three more men have results in Group 1 (WE3, WR3 and RA).

- WE3 and WR3 are known third cousins of one another; thank heavens they match one another exactly. They also match WC3 and the Group 4 pattern (or rather, they match two of the three Group 4 men exactly, and all of them at the middle marker with 25c).
- Two men (CA & MT) exactly match three of the men in Group 5; two (DC2 & JC) also match these except that they have 26.1t instead of 27.1t at the third marker.
- Our third new result in Group 1 is on RA, who also has 26c at the middle marker, but is the only man in our project who has 23t at the first marker, and he matches only RD in Group 4 with 28.1t at the last marker. As it happens, RA and JC have very similar 37-marker haplotypes, matching one another at the unusual value of DYS 390 = 24, and differing only at the three most rapidly mutating markers in the 37-marker panel. If their DYF399X results had matched (particularly if they had matched at some off-modal values), this would have been strong evidence of a close relationship, but the fact that they differ at two of the DYF399X alleles does not rule this out. Remember, the fact that DYF399X mutates so rapidly requires us to treat it somewhat the opposite of the other markers: we can use it to support a hypothesized relationship, but we cannot use it to rule out a relationship.
- This is the case with FI. His result not only has me stumped, but also Thomas Krahn (skip to the bottom of this page to see some correspondence between us and my attempt to answer FI's daughter Kathy Sterk's questions about this). On first consideration, this finding suggests that WC3 should look for a connection with Group 4, while CA, MT, DC2, & JC might look for a connection with Group 5 (but see the next paragraph below). This is exactly the sort of increase in resolution that we were hoping to find in the project, and we consider it very promising.

Branch Specific Markers

So, how about branch specific markers? There is a clear modal for the first allele, but there is little diversity, and this seems to be distributed randomly among our groups. At first glance, one might want to argue that there could be a branch characterized by 22t for JL and TG, but the fact that they are at genetic distance six on the 37-marker haplotype pretty much rules this out. The situation is only somewhat more promising at the third allele, where there is a little more diversity. Still, though it looks like we might argue that 25.1t delineates a branch for DN and HW in Group 3, in fact conventional genealogy shows that it is SR and DN who are on a branch together.

We do not have a modal for the middle DYF399X allele, so we cannot say anything convincing about what the ancestral value may have been at this

marker. I inquired on the RootsWeb genealogy-DNA and R1b1c7 lists asking whether anyone has collected enough DYF399X haplotypes to calculate an R1b1c7 (now more properly called R1b1b2e, notwithstanding that the list has not yet changed its name) modal for these markers, but this has not been done,[1] and generally folks are skeptical of the potential value of this. Although I really have to strain my milk to make an argument that the middle allele can be used as a branch marker,[2] I am struck by the fact that all the Group 4 men and all of the Group 7 men have 25c, especially in view of the fact that some have raised the question of a relationship between these two groups on independent grounds. Group 3 is practically no help, but the fact that HW has 26c and none of the others has 25c weakly supports the idea that 26c is the ancestral value, as does the fact that the "26c's" out number the "25c's" in Group 1 by 5:3.

Let us suppose that the ancestral value at the middle marker is 26c. This means that a mutation occurred somewhere in the line leading to the common ancestor of Group 4 and Group 7 (and that these two groups are more closely related to one another than they are to the other Groups[3]), or there were two independent parallel mutations from 26c to 25c leading to each of these lines (and they are no more closely related to one another than to the other groups). This means that there is "nothing special" about having 26c, and no reason to suppose that DC2, JC, CA & MT and are especially closely related to the men in Group 5.

If we assume on the other hand that the ancestral value at the middle marker is 25c, then a mutation occurred somewhere in the line leading to the common ancestor of Group 5 and DC2, JC, CA & MT and the DYS 391 = 10 mutation occurred sometime after that, or there were two independent parallel mutations from 25c to 26c leading to each of these lines. We are doing way too many "what ifs" here, but if the first of these possibilities were what actually happened, we might expect that the Group 5 cluster, being younger, would have less diversity than the overall group, but this seems not to be the case.

Hooboy. This kind of speculation could go on and on and on—some of you may think I have already made rather too good a start on that. Also, I think trying to explain this reasoning without diagrams is just too confusing, but we don't have a chalk board, and drawing diagrams on the computer is just too much trouble.

In balance, I think these results are at least somewhat promising. To establish with any certainty how helpful the DYF399X markers will end up being will require us to have more men in the closely related group tested for these markers. I am hopeful that others will be ordering this relatively inexpensive test as time goes along. (Let me emphasize that presently we think that this test is potentially helpful *only* to men in the closely related group, though it might

---

[1] In fact, I think that the Ewing project has more of this data than anyone, and certainly it has more that has been correlated with conventional genealogy than anyone.

[2] Remember that DYS 391 = 10 is the branch-specific marker for Group 5, because DYS 391 = 11 is the ancestral value for all the Ewings in the closely related group—including those in Group 5, but Group 5's common ancestor had a mutation to DYS 391 = 10.

[3] Note that if John of Carnashannagh turns out to be the progenitor of Group 7, this is just a special case of what I have just said.

become useful in other closely related subgroups in the project, such as Group 6, perhaps.)

To Order DYF399X

If you are a member of the Ewing project and would like to add this test, please go to your FtDNA personal page, click on ***Order Tests & Upgrades*** in the menu on the left-hand side of the page, then click on ***Advanced Orders***, scroll down to ***Y-STR DNA-FP Panel 5 Palindromic Pack*** but mark only the check box for ***DYF399X - $16.00***. (There will also be an additional "one time transfer fee" of $9.50 if this is the first time you have used the Advanced Order System.) Then click ***Continue*** and fill out the billing information. It will not be necessary to submit another specimen, because FtDNA already has the one you submitted previously in storage. If you have trouble ordering the test or are not sure whether it will be an informative test for you, please send an email to david_ewing@clanewing.org and he will help you.

References

From: Thomas Krahn <thomas@familytreedna.com>
To: David Ewing <davidewing93@gmail.com>,
Nov 29, 2007
Re: DYF399X

Dear David,

DYF399 was discovered by the Manfred Kayser group in Leipzig (Max Plank Institut / Germany) [1] by a software based tandem repeat search of the shortly before published Y chromosome sequence [2]. Only a handful of samples were tested by then. The forensic community disregarded this marker because of its complicated and error prone typing process. The usually 3 allelic pattern and [the fact that] one allele carry[s] an extra insertion of a single base made it difficult for inexperienced researchers to properly identify the haplotype. I still agree that nobody should ever be convicted by the pattern of a DYF399 profile.

While the forensic community dropped that marker right from the beginning, I had a fruitful discussion with Gareth Henson about possible recombination events in the P1/P2 palindromic region of the Y chromosome. This was inspired by an article from the medical community in infertility science where the transfer of coding SNP mutations from one palindromic arm to another was observed [3]. By following the inheritance patterns of DYS464 alleles in deep rooting pedigrees in genealogical DNA projects we already predicted a mechanism of recombination in palindromic STR markers which is today known as recLOH event. We needed a way to prove that systematic insertion and deletion events are happening based on recombination mechanisms in the palindromic arms. Gareth proposed a closer observation of the DYF399 marker [3] because of its asymmetric position on the 3 green (g) segments in the AZFc region. If a deletion

occurred, one peak should disappear which changes an asymmetric profile to a symmetric profile in the electropherogram. In most cases this should give us an answer on the position of the deleted segment.

Soon after the discussion with Gareth I started developing primers for DYF399 and tested them on regular- and outstanding DYS464 patterned male DNA samples in various haplogroups. I also worked out a nomenclature [5] that was usable in practice and understandable to the (educated) project administrators while it still accounted for the complexity of the STR units. The new marker showed an extremely high variance in allelic patterns and an extraordinary high mutation frequency. The allele with the 1bp insertion was found stable in most haplogroups but it sometimes disappeared in whole branches like haplogroup I and J. Haplogroup G showed a very distinct X.2 allele in addition to the still inherited X.1 allele. Also I found 5 allelic patterns for DYF399 in most duplicated haplotypes and 2 allelic patterns for all deleted haplotypes in the AZFc region. A rare duplication with 4 DYF399 alleles was observed in a R1a family and a secondary duplication in J2 even yielded 7 alleles at this marker.

Driven from the success of my DYS464X multiplex I also developed new DYF399X primers that were specific for distinct SNP mutations in the flanking region of DYF399 which allowed us to distinguish the position of each separate allele in most haplotypes. The position of these alleles are noted in my palindromic map [6] and give us a better understanding of recombination mechanisms in AZFc by comparing deleted haplotypes with related regular haplotypes and by observing the flip over of SNP indicators to the other palindromic arm in case of recLOH events.

In your project the members will experience a much higher diversity in their results than you are used to with regular Y-STR markers. Note that this marker has a much higher mutation frequency and mismatches are not an indicator of un-relatedness. Make sure that you compare the X.1 alleles only with other X.1 alleles and the C-types with C-types and the T-types with T-types only. The alleles may mutate forward and back again. Consider DYF399 always as an unstable construction that may introduce some white noise to your otherwise cleanly matching haplotypes. However it may drastically increase the resolution in between closely related individuals.

I hope this helps for now. Feel free to ask, if you have questions.

Thomas [Krahn]

Literature:

[1] Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyddin 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) http://www.journals.uchicago.edu/AJHG/journal/issues/v74n6/41028/41028.html?erFrom=-8827500027679023125Guest

[2] Skaletsky H. et al. (2003): The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes; Nature 423, 825-837 http://www.nature.com/nature/journal/v423/n6942/full/nature01722.html

[3] Kuroda-Kawaguchi T. et al (2001): The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men; Nature Genetics 29, 279 - 286 http://www.nature.com/ng/journal/v29/n3/full/ng757.html

[4] Henson G. (2005) DYF399S: A unique three-copy short tandem repeat on the human Y chromosome, Journal of Genetic Genealogy; Volume 1, Number 1, Page 8 (http://www.jogg.info/11/henson.htm)

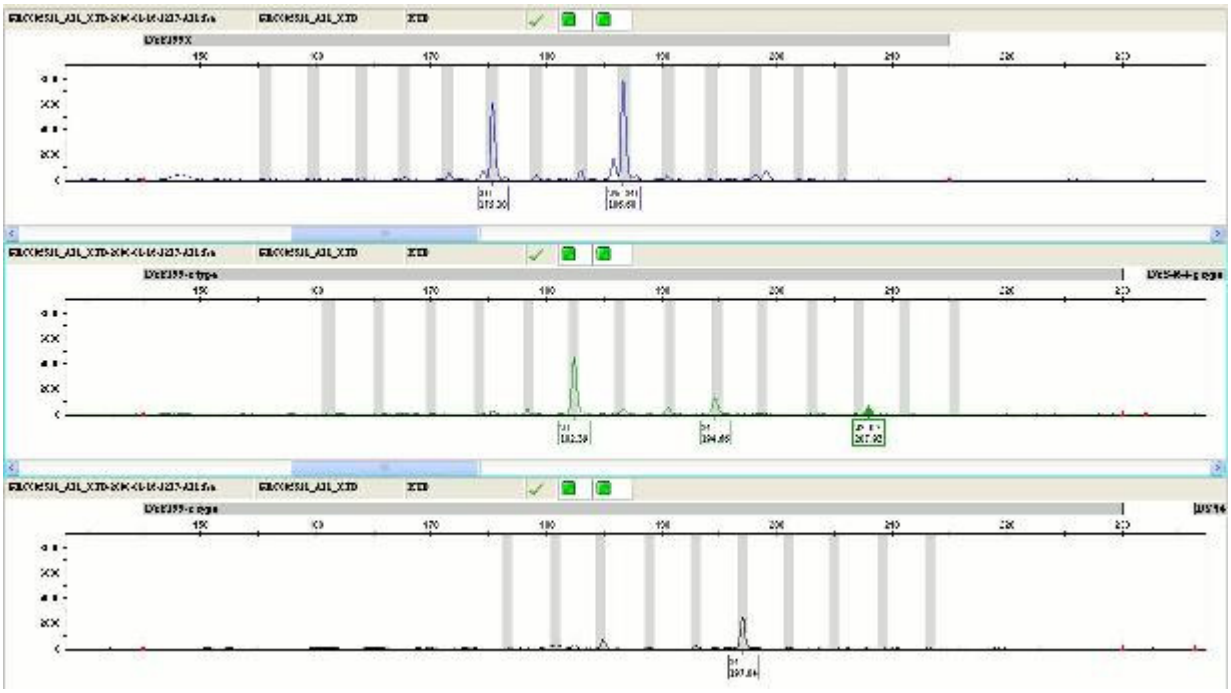
[5] Krahn T. (2005) Note on DYF399 Nomenclature; http://www.dna-fingerprint.com/modules.php?op=modload&name=Sections&file=index&req=viewarticle&artid=8

[6] Krahn T. (2005, Version2: 2006) Map of the Yq11 Palindromic Region; http://www.dna-fingerprint.com/modules.php?op=modload&name=Downloads&file=index&req=getit&lid=57

Feb 1, 2008
Hi, Thomas.
We have more resullts on our DYF399X series that include one result I cannot understand. Have a look at the table below. The result I do not understand is the fourth one down, for FI. To remind you of the Ewing Group structure, all but the last man in the group have rather similar 37-marker panels and are thought to be closely related. We have conventional genealogy linking the men in Group 3 with one another and the men in Group 4 with one another. The men in Group 1 and Group 5 are like them, but we do not know the conventional genealogic connection. The men in Group 5 differ from the other Ewings by having DYS 391 = 10 instead of 11. The last man on the list is John McEwan, who you may know. His 37-marker panel is unlike ours and he has other results that establish he has had a deletion event some time or another. The only result I am confused about is that for FI, specifically, why he should have 24t at the third allele instead of 2x.1t.
David

Feb 2, 2008
Dear David,
I have also no idea on how to explain the result for FI. I have verified the electropherogram [DNE: see below—this is the actual output of the machine that analyzes the DNA] and I agree with Astrid's interpretation. Although

there is a little tiny peak at 27.1t as indicated with a question mark in the green trace. It looks like this person had a major "bang" in his palindromic region which changed DYF399 drastically. Some few cells may still carry the ancestral 27t which shows up in the small peak. I've never seen a similar thing before. Thomas



Feb 2, 2008
Hi David,
What would cause such a sudden change? Could it have happened in dad's lifetime?
Kathy

Feb 2, 2008
Any mutation we find would have to have occurred as a specific sperm cell is being produced, so I guess we would have to know exactly what you mean by "in dad's lifetime." If his lifetime includes the moment in his father's testis that the single sperm that ended up fertilizing his mother's egg, then this could have happened "in his lifetime." But once he was born, once he was an embryo even, nothing could have happened that could have changed all of the cells in his body. Mutations only happen in one cell and they happen only when it is dividing. As a result, the only mutations that we ever find are those that take place in germ cells (sperm or egg), or in cancer cells--where one cell goes bad, then makes a large number of copies of itself.

To know when this happened, we would have to have samples from several of his relatives. If his brother is not the same as he is, then the mutation occurred when your grandfather was getting ready to father your dad. If your dad's

brothers are the same as him, but their first cousins are not, then the mutation took place in your great grandfather's testis when he was preparing to father your grandfather. If the mutation is also found in your dad's first cousins, but not in his second cousins, then the mutation took place in your 2nd great grand father's testis as he was preparing to father your great grand father.

As to what could have caused this, I think the only sensible answer is "luck." Any kind of mutation is rare. This kind of mutation is even rarer. But rare things happen, or we would be out of business. I think there is no evidence that such things as radiation exposure have any effect on the sort of mutations that we are interested in. Some people have argued that they are more likely when very old men father children, but I don't think even this is well established.

Warm Regards,

David