Report on First Results from the Van Tuyl Family DNA Project

Rory Van Tuyl
March 31, 2012

Abstract
Examination of y-chromosome short tandem repeat (Y-STR) data from four Van Tuyl males in the United States and The Netherlands confirms that three of them are related as expected and that they share the single nucleotide polymorphism (Y-SNP) pattern characteristic of haplogroup R1b1a2a1a1. A preliminary draft Y-STR haplotype for their most recent common ancestor (MRCA) in the 17th century was derived from the data. The fourth participant was found to have a haplotype consistent with the theory that his great-grandfather was adopted into the family in the 1850s. Evidence points to this ancestor having been fathered by a male from a specific Dutch-American family in New York.

Section 1: Background Information

Genome analysis services now available to the genealogy community offer the possibility of verifying family connections, establishing links where none had been possible with the records-based approach, and extending the time horizon for line of descent beyond historical times. For those studying the male line, the best tool is the y-chromosome DNA, which is passed exclusively from father to son. For those interested in the maternal line, mitochondrial DNA (mtDNA), which is passed from a mother to all her children, offers the best evidence. And for those studying human prehistory, both y-chromosome DNA and mtDNA provide amazing insights into human migration in prehistoric times.[1]

In the Van Tuyl family DNA project, we are concerned with two features of the y-chromosome: the single-nucleotide polymorphisms (Y-SNPs) and the short tandem repeats (Y-STRs). Neither of these characteristics have anything to do with the genes that encode proteins and are responsible for such characteristics as height, hair color and susceptibility to certain diseases. SNPs and STRs are completely inert markers, useless in life, but of great value to human historians. SNPs are extremely rare copying errors made during human reproduction. Once a SNP appears, it is likely to be inherited forever by all descendants of the first person to acquire it. New SNPs appear in a family on the order of once per thousands of years. Groups of people sharing a certain complement of SNPs are said to be of a certain *haplotype* and to belong to a certain SNP *haplogroup.* Some of these haplogroups have grown over time to contain millions of members.

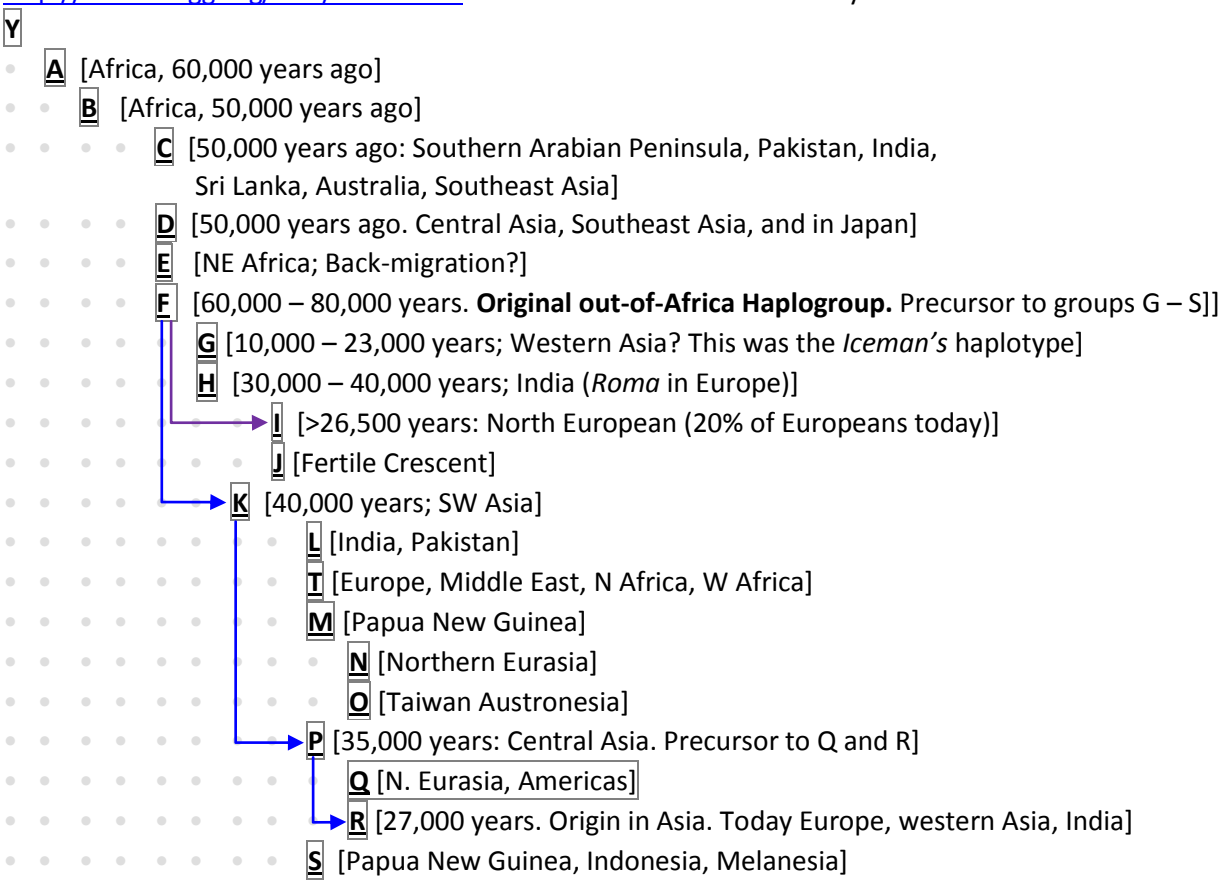
Table 1-1 presents a *phylogenic tree* of human SNPs, with the standard names for each haplogroup, along with the SNPs unique to that haplogroup. One of the 3 related Van Tuyls has been SNP-tested and found to belong to SNP haplogroup R1b1a2a1a1.[2] We can therefore assume all 3 Van Tuyls share this SNP haplotype.

Fig. 1-1 the major Y-SNP haplogroups of Europe arranged in a phylogenic tree alongside a timeline indicating when each of them emerged. For future reference: when presented as part of a phylogeny, as in Fig. 1-1, haplogroups are sometimes referred to as *clades* or *subclades*.
The two haplogroups (clades) of greatest interest to the Van Tuyl project are R1b and I1, which taken together account for most modern men in The Netherlands. Fig. 1-2 shows how the frequency of occurrence for these two haplogroups varies over Western Europe.

And finally, Fig. 1-3 shows the relative frequency of all the major haplotypes in
Europe. Clearly, being a member of haplogroup R1b has its advantages and disadvantages. On the plus side, there is plenty of data on R1b and it subclades from which to form an analysis. On the minus side, there is little specificity. Being R1b in Europe is sort of like having blue eyes: you and millions of others share the characteristic.

[1] For more information on human prehistory as determined from y-DNA and mtDNA, see: https://genographic.nationalgeographic.com/genographic/lan/en/atlas.html
[2] RLVT was tested in February, 2012 by 23andMe, Inc.

Table 1-1. Below are listed the principal Y-DNA SNP haplogroups with hypothesized origins and present-day areas of concentration. Each group is associated with one or more discovered SNPs of the group and its ancestral haplogroups. For example, a man in R will carry SNPs from F (28 known), K (4 known), P (24 known), as well as R (9 known). The **blue** and **purple** arrows show two lines of descent from out-of-Africa to Europe. Detailed description of these SNPs can be found at: http://www.isogg.org/tree/index.html. New SNPs are discovered each year.



Below are the SNPs of haplogroup R and its subclades. A man in R1b1a2a1a1a will carry 45 SNPs from haplogroup R in addition to 56 from its precursors. http://www.isogg.org/tree/ISOGG_HapgrpR.html

R M207/Page37/UTY2, P224, P227, P229, P232, P280, P285, S4, S9
R1 M173/P241/Page29, M306/S1, P225, P231, P233, P234, P236, P238, P242, P245, P286, P294
R1b M343
R1b1 L278, M415, P25_1, P25_2, P25_3
R1b1a P297, L320
R1b1a2 L265, M269, M520, S3, S10, S13, S17
R1b1a2a L23/S141, L49.1
R1b1a2a1 L150
R1b1a2a1a L51/M412/S167
R1b1a2a1a1 L11/S127, L52, L151, P310/S129, P311/S128
R1b1a2a1a1a M405/S21/U106

Return to text

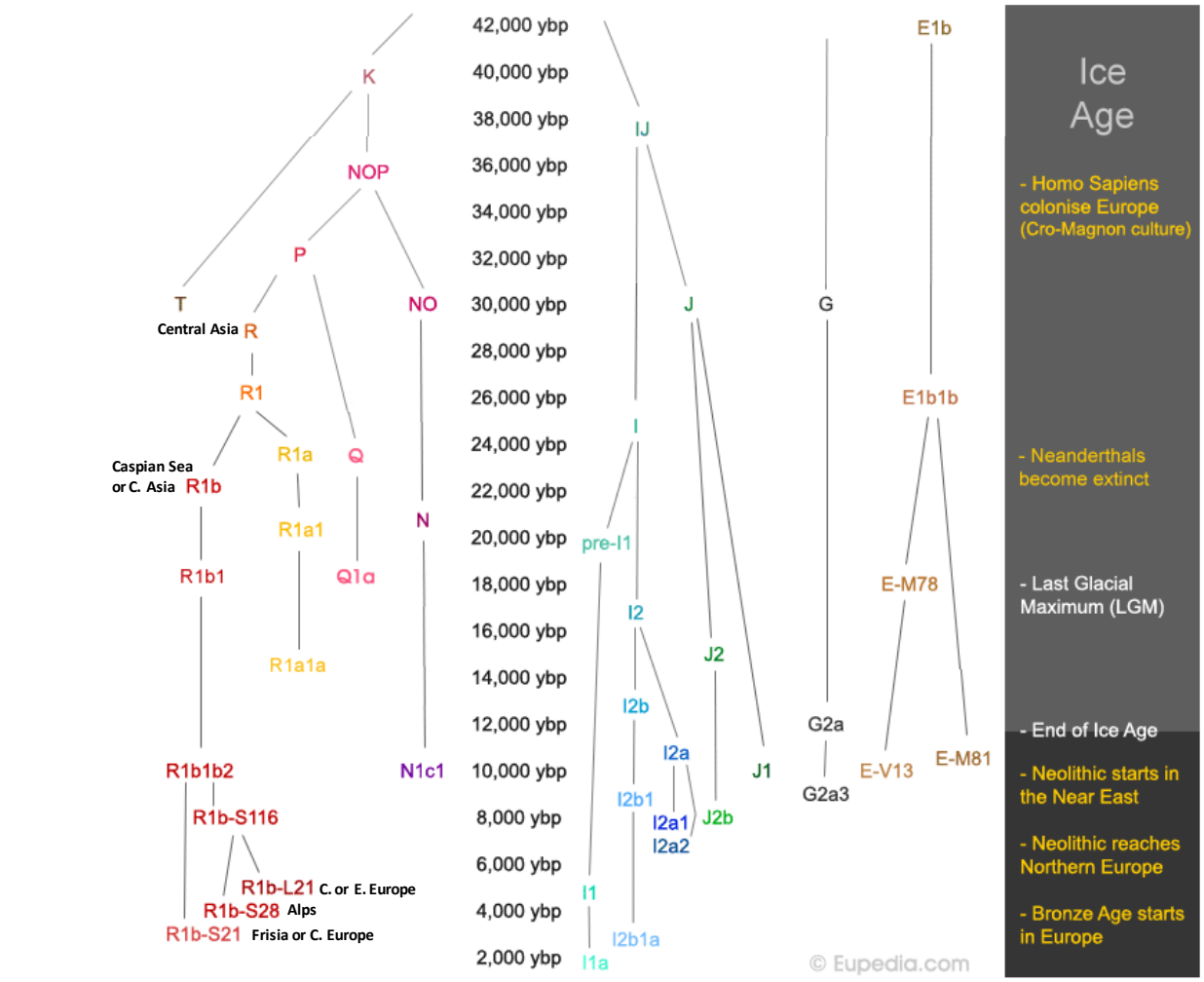


Fig. 1-1 Phylogeny and timeline for major Y-DNA Haplogroups. Each haplogroup is associated with a series of Y-chromosome SNPs acquired from ~40,000 years ago through ~3,000 years ago. The major European Haplogroup, R1b, traces its origin back some 29,000 years to central Asia. Today, R1b is dominant in Western Europe, and R1a in Eastern Europe. Present-day Europe is also home to haplogroup I1, associated with Scandinavia and northern Europe, which branched off from R1 more than 42,000 years ago. Reference: http://www.eupedia.com/europe/origins_haplogroups_europe.shtml

Return to text

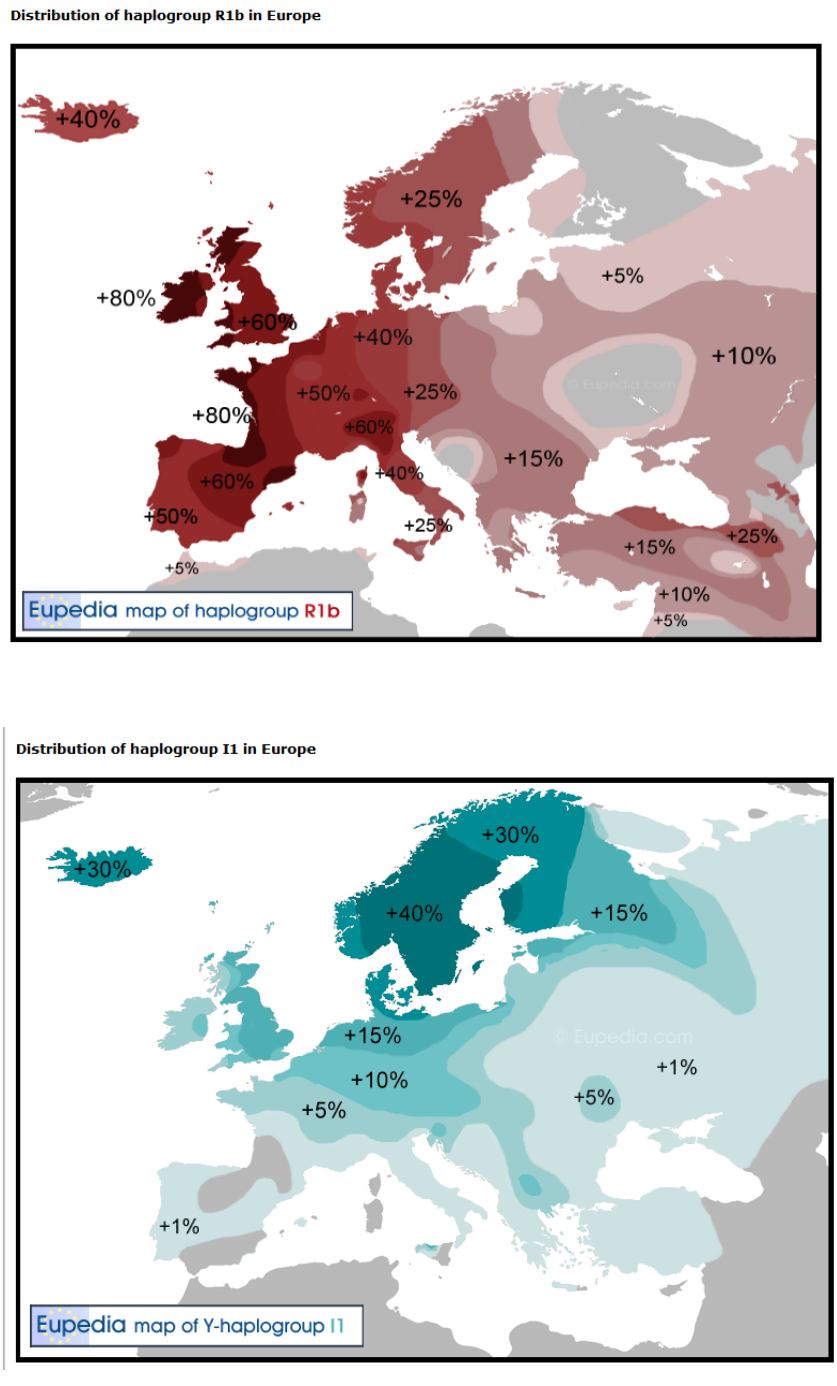


Fig. 1-2 Geographic distribution and prevalence of the two major Y-DNA haplogroups of Western Europe today: R1b and I1. Reference: http://www.eupedia.com/europe/origins_haplogroups_europe.shtml

Return to Text

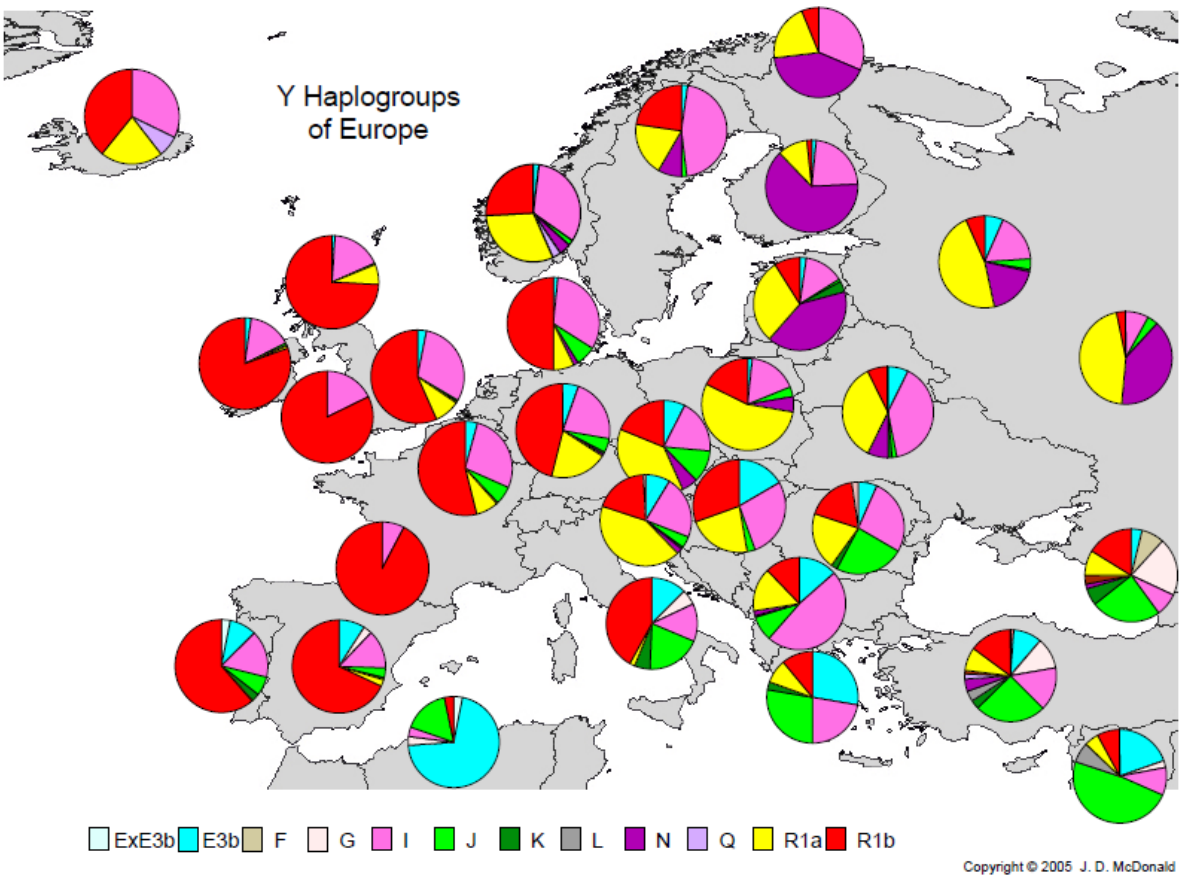


Fig. 1-3. Distribution and relative prevalence of Y-DNA haplotypes in Europe. R1b dominates in Western Europe, and R1a is strongest in eastern Europe. Haplotye I is the most prevalent in western Scandinavia and is the second most frequently-occurring in Western Europe and the British Isles.
Ref: J. D. McDonald: http://www.scs.illinois.edu/~mcdonald/WorldHaplogroupsMaps.

Return to Text

STR Gazing

As useful as they are for marking forks in the road of human migration, SNPs cannot provide detailed genealogical information for a simple reason: they occur too infrequently. The major SNPs portrayed in Table 1 and Fig. 1-1 have occurred, on average, once every 2500 years in the R clade. Lesser-known SNPs have occurred more recently, and some have been discovered to be useful in further narrowing the field of view. For example, a SNP numbered M222 was found in 2005 to occur at high incidence among men in Northern Ireland and Scotland. This SNP is now the definitive marker for haplotype R1b1a2a1a1b4b, representing perhaps 2 million men living today.[3] But to which clan does an M222 man belong? Which family? SNPs won't tell.

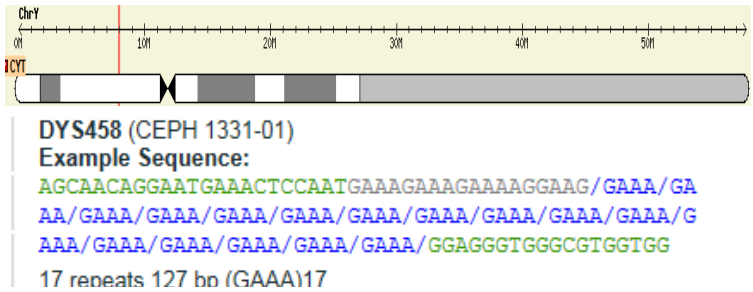
Whereas a SNP is like a needle in a haystack – does he or doesn't he have it? - Y-STRs are present in every man on the planet. The question is not "Does he have it?" but "How many repeats does he have?" As a result, testing is more definitive, and the results are able to prove or disprove consanguinity to a high degree of certainty. Y-STRs are typically four nucleotide sequences repeated 10 – 25 times. They are located in non-coding areas of the chromosome so have no known practical value in the genome. Fig. 1-4 shows two Y-STRs, one fast-mutating and the other slow-mutating. Note that over time, the variance of the probability distribution for Y-STRs increases. In both cases, the variance for worldwide data is about twice as large as for Dutch men in the R1b1a2a1a1a clade. This is because the common ancestor for the Dutch men was between 3,000 and 4,000 years ago, whereas for the worldwide data set the time to most recent common ancestor (TMRCA), though unknown, is much larger.

But who was this common ancestor? Guess what: it's the guy with the SNP that defines the clade or subclade of the population in question. When the SNP first appeared, the man born with it, the man destined to father millions of sons in future generations, had a particular pattern of Y-STRs. There are over 100 Y-STRs known and available today for commercial genetic testing. This ancestor had them all, and he had them in a particular numerical combination. In the absence of mutations, all his descendants would have exactly the same combination of Y-STRs. This combination is called the Y-STR haplotype, and people with this haplotype, or something close to it, are said to belong to a particular haplogroup. But mutations of a Y-STR length do occur, and they occur as frequently as once in 100 generations or so. So the statistical distribution of occurrences for a particular Y-STR locus in today's descendants will have diffused into something like a normal distribution centered about the number of short tandem repeats the ancestor had for that particular locus. It is this diffusion that creates the identifying characteristics of a family that are distinct from the rest of the haplogroup.

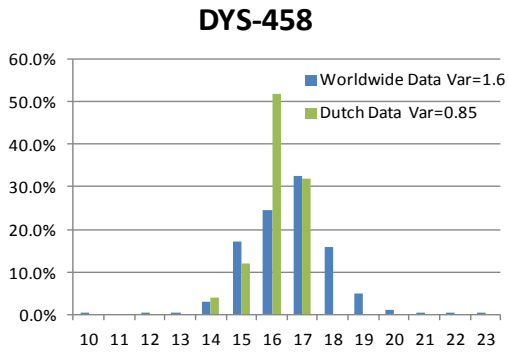
The process of Y-STR diffusion and segregation into Y-SNP haplogroups is diagrammed in Fig.1-5 .

[3] http://www.familytreedna.com/public/R1b1c7

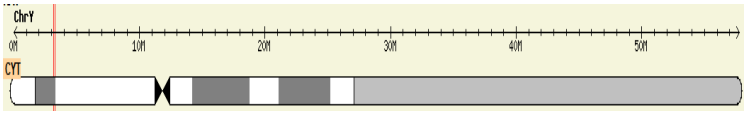
A



DYS458 (CEPH 1331-01)
Example Sequence:
AGCAACAGGAATGAAACTCCAATGAAAGAAAGAAAAGGAAG/GAAA/GA
AA/GAAA/GAAA/GAAA/GAAA/GAAA/GAAA/GAAA/GAAA/G
AAA/GAAA/GAAA/GAAA/GAAA/GAAA/GGAGGGTGGGCGTGGTGG
17 repeats 127 bp (GAAA)17



B



DYS393 (CEPH 1347-01)
Example Sequence:
GTGGTCTTCTACTTGTGTCAATAC/AGAT/AGAT/AGAT/AGAT/AGAT/
AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/
ATGTATGTCTTTTCTATGAGACATACCTCATTTTTTGGACTTGAGTTC
15 repeats 132 bp (AGAT)15

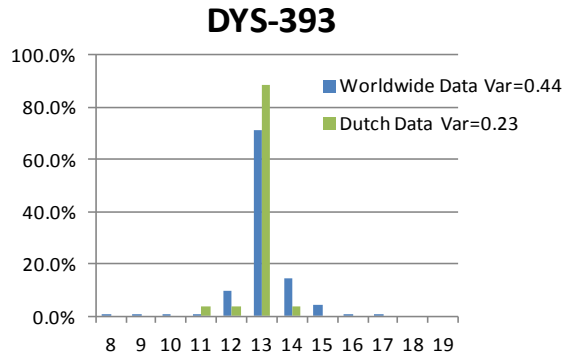


Fig. 1-4. Examples of two Y-STRs, one fast-mutating, one slow. Both are located on the P arm of the Y chromosome and feature multiple repeats of 4-base segments. The precise location of each is called its *locus* (plural=*loci*), with each locus being given an identifying number (e.g. DYS391). Distributions are from >33,000 worldwide samples [http://www.smgf.org/] and 25 Dutch R1b1a2a1a1a individuals. (A) DYS-458: mutation rate ~7.8E-03 per generation; worldwide variance = 1.6; Dutch variance=0.85; (B) DYS-393: mutation rate ~0.9E-03 per generation; worldwide variance = 0.44, Dutch variance = 0.23.

Return to Text

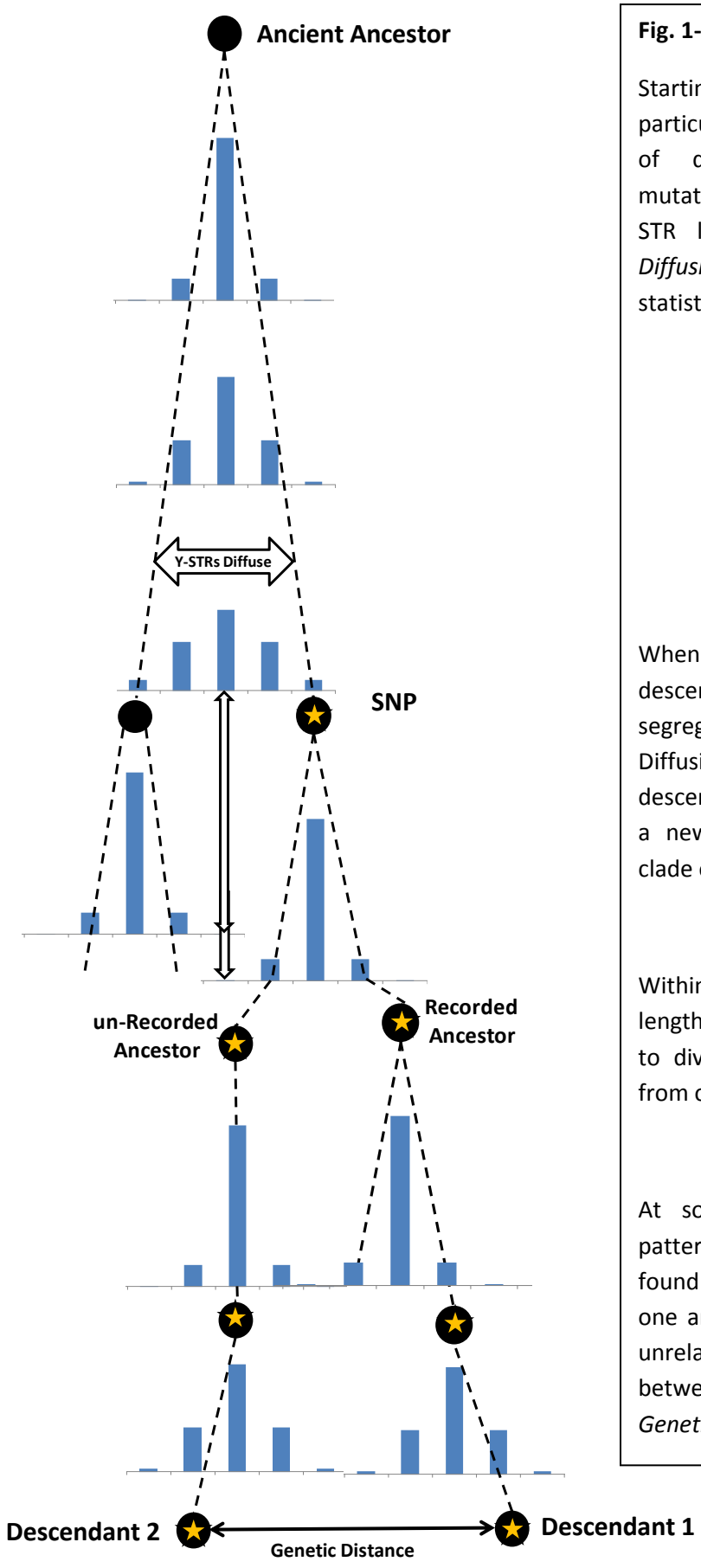


Fig. 1-5

Starting with an ancient ancestor with a particular Y-STR haplotype, generations of descendants, through random mutation, gain a wider distribution of STR lengths. This process is called *Diffusion*, and can be described with statistical math.

When a SNP appears at random, the descendants of this SNP ancestor segregate into a new haplotype. Diffusion of STRs continues, and all descendants with the new SNP comprise a new haplogroup, and form a new clade of the phylogeny.

Within the clade, diffusion of STR lengths continues, causing descendants to diverge more and more over time from one another.

At some point in time, the Y-STR patterns for any two descendants will be found to be somewhat different from one another, but more similar than for unrelated males. The difference between them can be thought of as a *Genetic Distance*.

Return to Text

Section 2: Initial Results of the Van Tuyl Y-STR Analysis

Y-STR data from 67 loci was obtained for three descendants of Ott Jansz van Tuyl (b. ~1606), two of whom are American descendants of Abraham Van Tuyl (b. 1681). Per genealogical research published in *A Van Tuyl Chronicle*[4], these three men, RDVT, CDVT, and RLVT are related as follows:

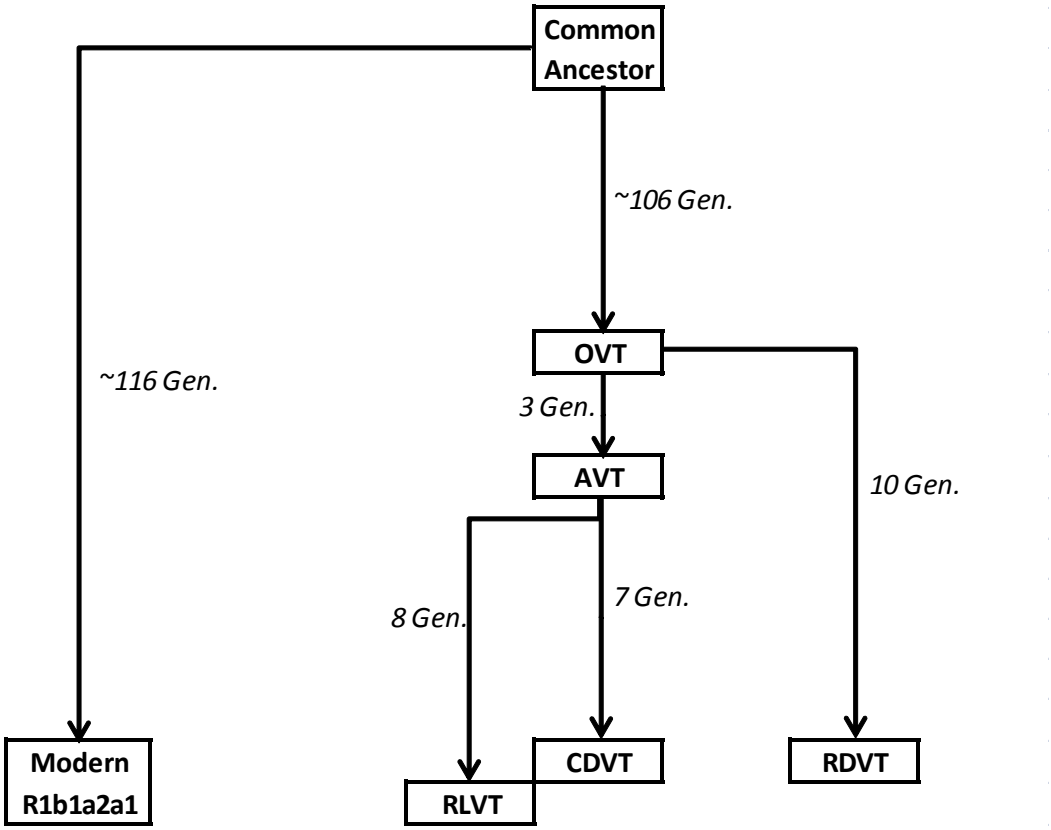


Fig. 2-1: Phylogenic tree for three distant Van Tuyl cousins, RLVT, CDVT, and RDVT in relation to their ancestors. CDVT is separated by 15 generations from RLVT and by 20 generations from RDVT. The greatest separation is 21 generations, between RLVT and RDVT. OVT [Ott Jansz van Tuyl (b. ~1606)] is the Most Recent Common Ancestor (MRCA) for all three. AVT is Abraham Van Tuyl (b. 1681), the MRCA for the American Cousins RLVT and CDVT. All modern members of the R1b1a2a1a1 haplogroup are estimated to be about 116 generations (~3500 years) removed from the R1b1a2a1 ancestor.

Results of the 67-locus Y-STR tests are summarized in figures 2-2 and 2-3. Fig. 2-2 presents the raw data for RLVT, CDVT and RDVT, along with the best estimate of what the AVT Y-STR haplotype would have been. Fig. 2-3 shows which Y-STR loci have mutated in each branch of the above phylogenic tree. Fig. 2-4 shows the actual data and statistical expectations for genetic distance between the Van Tuyls and their R1b1a2a1a1 non-relatives.

[4] R. L. Van Tuyl and J.N.A. Groenendijk, "A Van Tuyl Chronicle, 650 Years in the History of a Dutch-American Family," (1996). http://books.google.com/books/about/A_Van_Tuyl_chronicle.html?id=K49YAAAAMAAJ

Kit Number	Abbreviated Name	Paternal Ancestor Name	Country	Haplogroup	
222095	RLVT	Heer Ghijsbrecht van Tuyl (bef. 1345 - 1383)	USA	R1b1a2a1a1	SNP Tested
227732	CDVT	Heer Ghijsbrecht van Tuyl (bef. 1345 - 1383)	USA	R1b1a2	Inferred
222096	RDVT	Heer Ghijsbrecht van Tuyl (bef. 1345 - 1383)	Netherlands	R1b1a2	Inferred
Calculated	VTM	Mode of above individuals			
Inferred	OVT	Ott van Tuyl (~1606 - bef. 1666)	Netherlands		
Databases	AHT	Ancestor Haplotype (~116 generations b.p.)	Europe	R1b1a2a1a1	

	1	2	3	4	5	6	7	8	9	10	11	12
Name	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389i	DYS392	DYS389b
RLVT	12	23	14	11	11	14	12	12	11	14	13	17
CDVT	12	23	14	11	11	14	12	12	11	13	13	17
RDVT	12	24	14	10	11	14	12	12	11	13	13	16
VTM	12	23	14	11	11	14	12	12	11	13	13	17
OVT	**12**	**24**	**14**	**11**	**11**	**14**	**12**	**12**	**11**	**13**	**13**	**17**
AHT	13	24	14	11	11	14	12	12	12	13	13	16

	13	14	15	16	17	18	19	20	21	22	23	24	25
Name	DYS458	DYS459	DYS459	DYS455	DYS454	DYS447	DYS437	DYS448	DYS449	DYS464a	DYS464b	DYS464c	DYS464d
RLVT	17	9	10	11	11	25	15	20	28	14	15	17	17
CDVT	17	9	10	11	11	25	15	20	28	14	15	17	17
RDVT	17	9	10	11	11	25	15	20	29	14	15	17	17
VTM	17	9	10	11	11	25	15	20	28	14	15	17	17
OVT	**17**	**9**	**10**	**11**	**11**	**25**	**15**	**20**	**29**	**14**	**15**	**17**	**17**
AHT	17	9	10	11	11	25	15	19	29	15	15	17	17

	26	27	28	29	30	31	32	33	34	35	36	37
Name	DYS460	Y-GATA-H4	YCA-IIa	YCA-IIb	DYS456	DYS607	DYS576	DYS570	CDY_1	CDY_2	DYS442	DYS438
RLVT	11	11	19	23	15	15	19	18	38	38	12	13
CDVT	11	11	19	23	15	15	18	18	38	39	12	13
RDVT	11	11	19	23	15	15	18	18	38	38	12	13
VTM	11	11	19	23	15	15	18	18	38	38	12	13
OVT	**11**	**11**	**19**	**23**	**15**	**15**	**18**	**18**	**38**	**38**	**12**	**13**
AHT	11	11	19	23	15	15	18	17	37	37	12	12

	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
Name	DYS531	DYS578	DYF395S1	DYF395S2	DYS590	DYS537	DYS641	DYS472	DYF406S1	DYS511	DYS425	DYS413a	DYS413b	DYS557	DYS594
RLVT	11	9	15	16	8	10	10	8	10	10	12	22	23	15	10
CDVT	11	9	15	16	8	10	10	8	10	10	12	22	23	15	10
RDVT	11	9	15	16	8	10	10	8	10	10	12	22	23	15	10
VTM	11	9	15	16	8	10	10	8	10	10	12	22	23	15	10
OVT	**11**	**9**	**15**	**16**	**8**	**10**	**10**	**8**	**10**	**10**	**12**	**22**	**23**	**15**	**10**
AHT	11	9	15	16	8	10	10	8	10	10	12	23	23	16	10

	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67
Name	DYS436	DYS490	DYS534	DYS450	DYS444	DYS481	DYS520	DYS446	DYS617	DYS568	DYS487	DYS572	DYS640	DYS492	DYS565
RLVT	12	12	16	8	12	23	20	13	12	11	13	11	11	12	12
CDVT	12	12	16	8	12	23	20	13	12	11	13	11	11	12	12
RDVT	12	12	17	8	12	23	20	13	12	11	13	11	11	12	12
VTM	12	12	16	8	12	23	20	13	12	11	13	11	11	12	12
OVT	**12**	**12**	**16**	**8**	**12**	**23**	**20**	**13**	**12**	**11**	**13**	**11**	**11**	**12**	**12**
AHT	12	12	15	8	12	22	20	13	12	11	13	11	11	12	12

Fig. 2-2: Table of 67 Y-STR values (haplotype) for three Van Tuyl project members, along with their mode (majority of 3). OVT is the inferred haplotype of their most recent common ancestor [MRCA] Ott van Tuyl; AHT is the inferred Ancestral Haplotype of their ancient common ancestor (~2000-1000 BC). Red entries indicate difference from the ancestral haplotype, and _fill_ indicates mutation from the OVT haplotype. The table has been altered from FTDNA format for purposes of analysis.

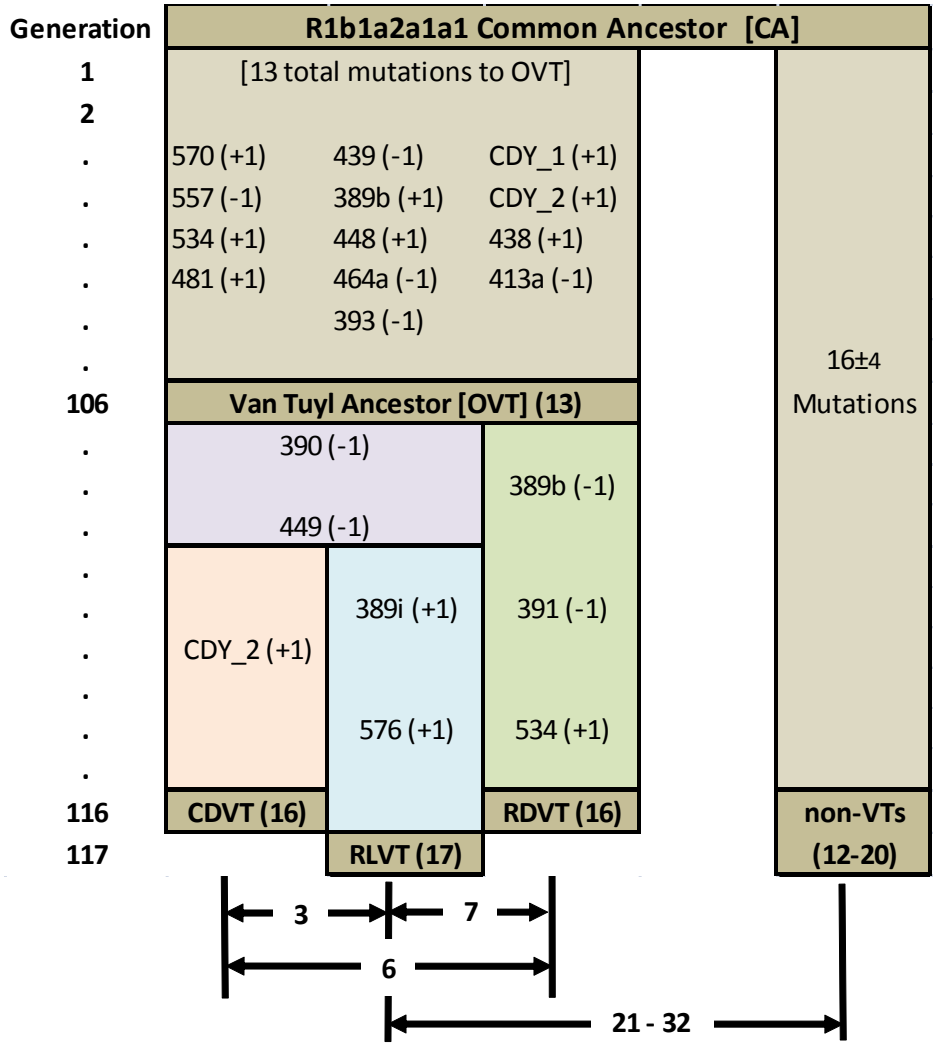


Fig. 2-3: Detailed Phylogenic Tree of the Van Tuyls and others in their SNP haplogroup. The ancient common ancestor [CA] was a man living ~116 generations (~3500 years) ago. He and his descendants are of the SNP haplotype R1b1a2a1a1. The three modern-day Van Tuyls are also believed to be of SNP haplotype R1b1a2a1a1, based on SNP test results from RLVT. Data on R1b1a2a1a1 descendants in the ysearch database indicates they have between 12 and 20 Y-STR mutations from the ancestral Y-STR haplotype. The 3 Van Tuyls have 16-17 mutations from the ancestral Y-STR haplotype, as would be expected. The Most Recent Common Van Tuyl Ancestor, OVT, was separated by ~106 generations and 13 mutations from the ancient ancestor [CA]. The three modern Van Tuyls have between 3 (CDVT, RDVT) and 4 (RLVT) mutations from their common ancestor OVT. The table shows which Y-STR loci mutated when, and by how much. (We assume that the locus 389b has mutated twice: gaining one repeat between CA and OVT, and losing one repeat between OVT and RDVT.)

Genetic distance between the 3 Van Tuyls is shown to be between 3 and 7 out of 67 loci. Genetic distance between the Van Tuyls and non-Van Tuyl members of the R1b1a2a1a1 haplogroup is between 21 and 32 out of 67 loci.

The Case of the 1850s Adoptee

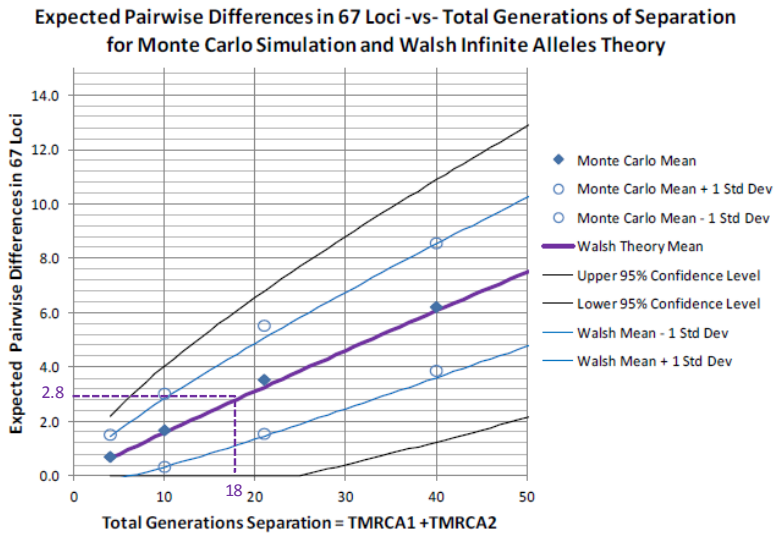
I have corresponded over the years with a woman – let me call her Annie Van Tuyl – who suspected, based on a search of family and public records, that her great-grandfather Van Tuyl had been adopted into the family. A scrap of paper obtained from an uncle indicated he was born in 1852 under the name of McCreary, and adopted by Annie's ancestors, a childless New York couple named Van Tuyl, sometime in that decade. At my urging, she submitted a sample from her brother to FTDNA for Y-STR analysis as part of the family project. The results were unequivocal: her brother, AWVT, was from haplogroup I1, not R1b as for RLVT, RDVT and CDVT. So clearly, the adoption story was probably true.

But where, Annie wondered, was her great-grandpa from? As previously discussed, the chance of nailing down a person's region of origin based on Y-STR haplotyping is pretty poor. In this case, despite much analysis by I1 haplogroup members, the data offered us very little. The chances of pinpointing the origin were nil. We were stuck with "somewhere in Europe or the British Isles".

But FTDNA maintains a private database of its clients, and supports a public database called ySearch [www.ysearch.org], which incorporates over 100,000 entries of Y-STR data. And this database is searchable. Unfortunately, the AWVT data was very close to the modal haplotype for the I1 group, so searches with fewer than 37 loci produced a lot of spurious responses. For 12 loci, more than 300 spurs reared their ugly heads, at 37 loci, there was still one interloper, but as soon as all 67 loci were in place, a group of 5 men popped out, and they were all named *Terwilleger*. Fig. 2-4 compares the Y-STR data for AWVT to that for one of the Terwilliger cousins. The relationship between them is clearly demonstrated.

The Terwilligers immigrated to New Amsterdam in 1663 and eventually settled north of the town at the settlement of *New Paltz*, on the Hudson River. Over the years, their descendants spread all around the area so that, by the 1850s, there were plenty of Terwilligers near where the childless Van Tuyl couple resided [See: http://lystykds.0catch.com/tfafiles/homepage.htm].

So, based on these Y-STR findings, a new hypothesis has emerged: a girl named McCreary was impregnated by a man named Terwilliger, and due to circumstances was forced to give up her child. The local Van Tuyl couple took the boy. These details are amenable to fact-checking using standard genealogical techniques, and Annie will be pursuing this new, promising lead - probably for years. In the meantime, she has now not one, but two, New Netherland families in her male line! Without the DNA evidence, her investigation would surely have remained stalled. This case demonstrates one of the biggest advantages of genetic genealogy: finding completely unsuspected relationships.



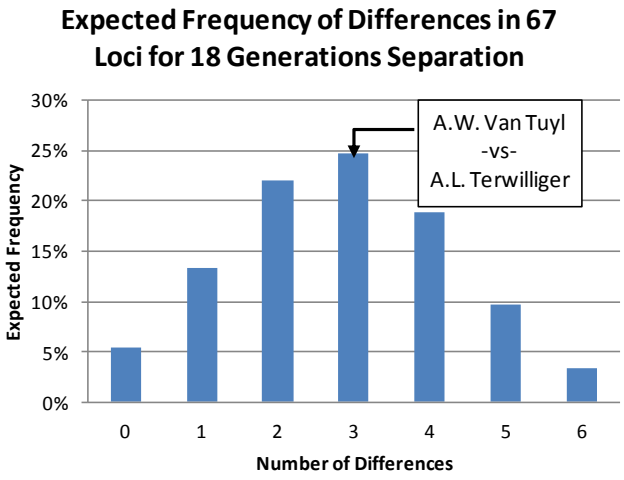


Fig. 2-4. Using the curve presented in Fig. 6b of the term paper, we see that for a total separation of 18 generations, we would expect 2.8±1.6 differences between related individuals. This agrees precisely with the observed difference between the Van Tuyl of the line in question and the one Terwilliger descendant who has tested and posted 67 Y-STR loci. Although spurious responses were common in 12 and 25 loci, by 67 loci, there were none.

Return_to_text

Section 3: Technical Critique and Discussion of the Results

How sure are we that the results are valid? This section presents some detailed technical information for those interested in the calculations and possible sources of error.

The statistical expectation for the genetic distance between two men separated by TMRCA1 + TMRCA2 from their most recent common ancestor [MRCA] was determined using both Monte Carlo simulation and the *Infinite Alleles Theory*.[5] The results are shown here:

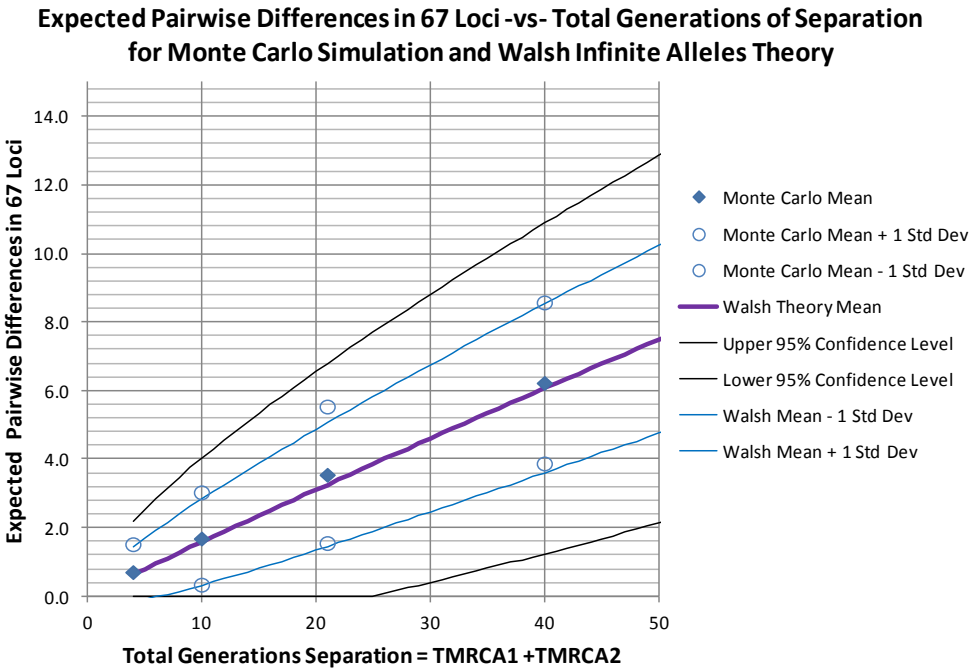


Fig. 3-1. Expected number of pairwise differences between two men of any haplogroup when 67 loci are summed. Monte Carlo and Walsh "Infinite Alleles" theory are in very good agreement. For example, two men, each separated by 20 generations from MRCA, would expect to have between 4 and 8 differences out of 67 loci (68% confidence). If separated from MRCA by 5 generations each, two related men could expect as many as 4 differences out of 67 loci with 95% confidence. For this suite of 67 loci, it takes about 6 generations of total separation to accumulate 1 pairwise difference, on average.

[5] B. Walsh, "Estimating the time to the MRCA for the Y chromosome or mtDNA for a pair of individuals," *Genetics* 158: 897—912 (2001).

The process of diffusion, as diagrammed in Fig. 1-5, can be visualized in the following graph, which plots the probability versus time for several cases of generational separation.

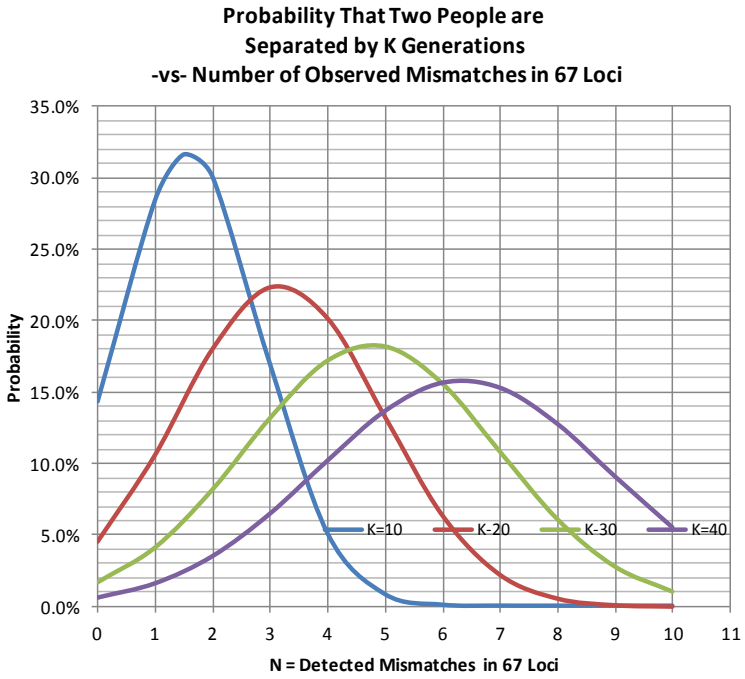


Fig. 3-2. Probability of occurrence –vs- the number of pairwise mismatches over 67 loci (the *Genetic Distance*). As the generational separation increases, the most probable genetic distance increases, along with its statistical uncertainty. For example, N=3 differences shows it is most likely (P=22%) that the generational separation, K, is 20. However, the separation could also be between 10 (P=17%) and 30 (P=13%). So it is clearly not possible to determine TMRCA with any confidence by this method.

But the resolution of this technique is sufficient to tell the differences between people who are related and those who are not, as illustrated here:

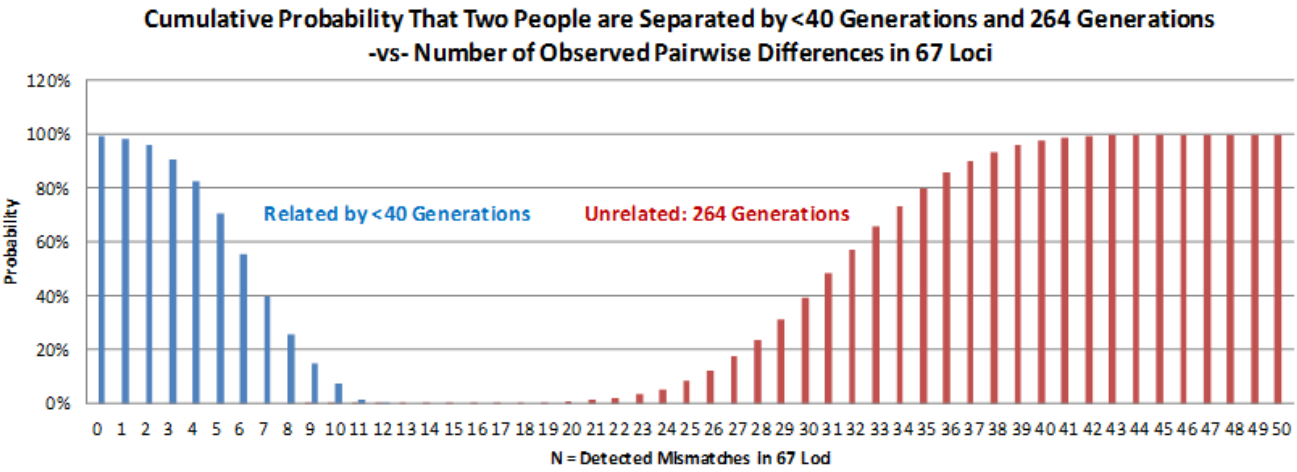


Fig. 3-3. Cumulative probability –vs- Genetic Distance, showing that a very clear distinction can be drawn between relationships up to K=40 (the maximum reliable extent of most records-based genealogies, ~600 years) and unrelated men from the R1b1a2a1a1a haplogroup (K=2x132 generations). With a confidence of 95%, 10 differences or fewer between two men would signal a common ancestor up to 20 generations back, whereas 25 or more differences would signal no common ancestor until 132 generations (~3000-4000 years) in the past, at the beginning of the Y-SNP clade.

Since Y-SNP testing of RLVT revealed that the Van Tuyls are members of haplogroup R1b1a2a1a1, and since data for unrelated men in that clade was available through ysearch, we were able to fit the data from these unrelated men to a normal statistical distribution and plot that distribution on the same scale as the expected and actual Genetic Distances between the three Van Tuyl subjects. Fig. 3-4 shows these plots. Clearly, the Van Tuyls are related to a degree consistent with theoretical expectation. By taking the ratio of expected probability for each match to expected probability for unrelated members of the haplogroup, we can determine the chances for an unrelated man to "masquerade" as a Van Tuyl by exhibiting – purely by chance – a Genetic Difference typical of related men. The worst case is 0.44% chance of falsely appearing to be related to 7 differences out of 67 loci. However, the chance that RLVT, who exhibits a Genetic Distance of 7/67 from RDVT, is this particular random man is miniscule: one chance out of the millions of European-descended men in the world. So we can be sure he is related to the Van Tuyl group. But if a non-Van Tuyl did appear with a Genetic Distance of 7 or less, additional analysis would be required to exclude him from – or include him in - the family.

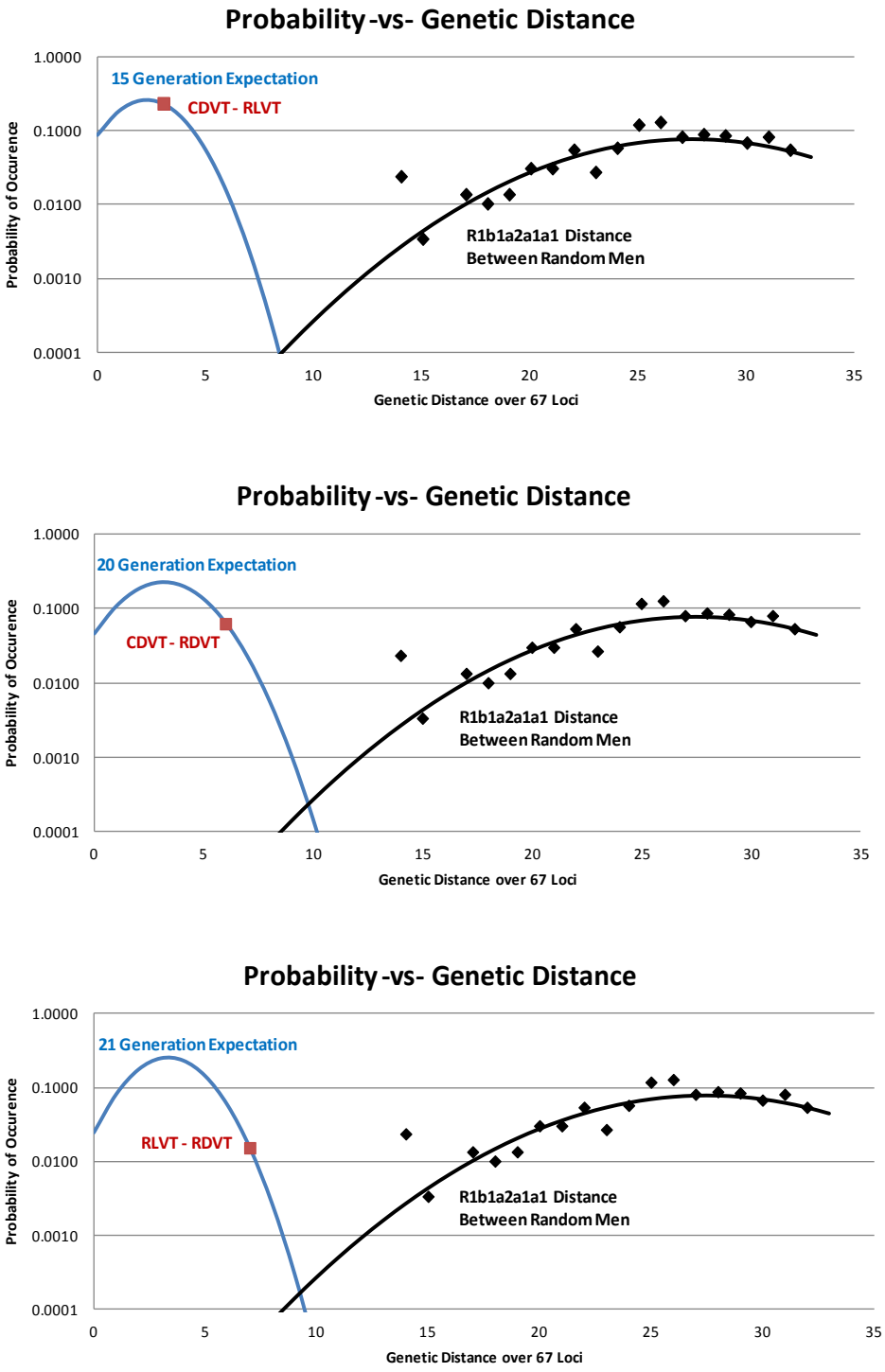


Fig. 3-4: Genetic Distance probability expectations for RLVT, CDVT and RDVT and for randomly-selected men from the R1b1a2a1a1 SNP haplogroup. Theoretical probability that a non-relative would match the genetic distance between RLVT and RDVT is =0.0044; between RLVT and CDVT it is =0.00051; between CDVT and RDVT it is =0.000012. The probability that any of the Van Tuyls would actually be unrelated yet match by pure chance is infinitesimal.

Return to Text

Despite these arguments, the fact remains that the Genetic Distance observed between RLVT and RDVT, and between CDVT and RDVT, is larger than expected. Why is this? Fig. 2-1 showed Genetic Distance increasing smoothly with increasing generational separation. In fact, this is not what really happens. Fig. 3-5 shows the results of a Monte Carlo simulation for Genetic Distance –vs- generational separation:

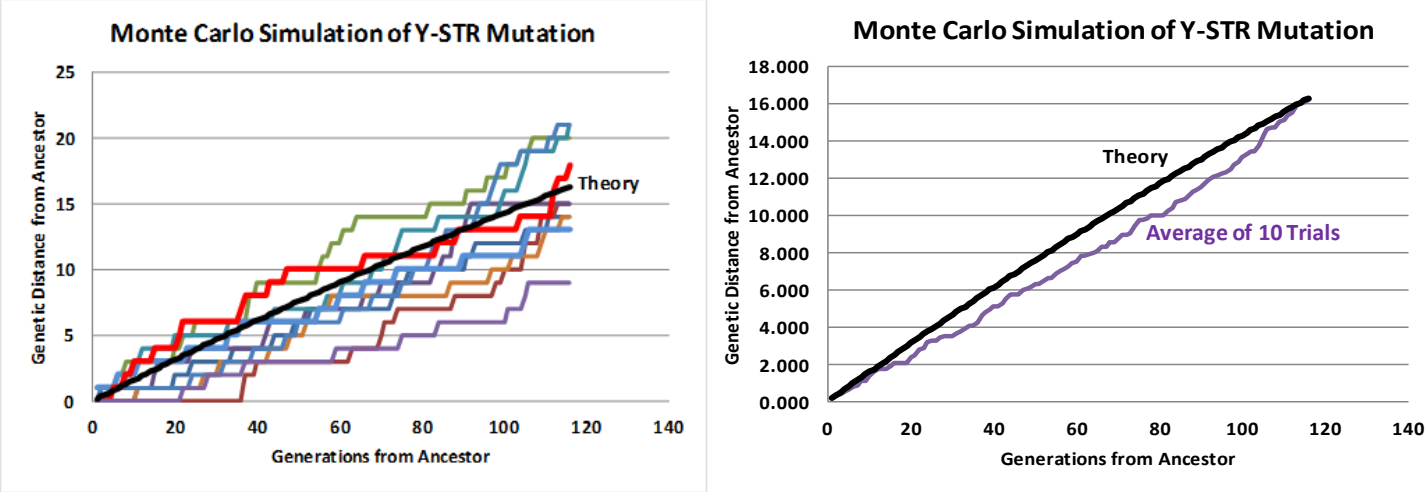


Fig. 3-5. Monte Carlo simulation of the mutation history over 116 generations. Ten different trials [left] give 10 different results for Genetic Distance at the end of 116 generations. The average of these 10 trials [right] results in a genetic distance of 16 after 116 generations with a standard deviation of 4. Notice that the trial plotted in Red shows a sharply increased mutation rate over that final 10 generations.

The increased genetic distance observed in the Van Tuyl data can be explained by recognizing the likelihood of an increase in mutation rate near the end of the 116 generations:

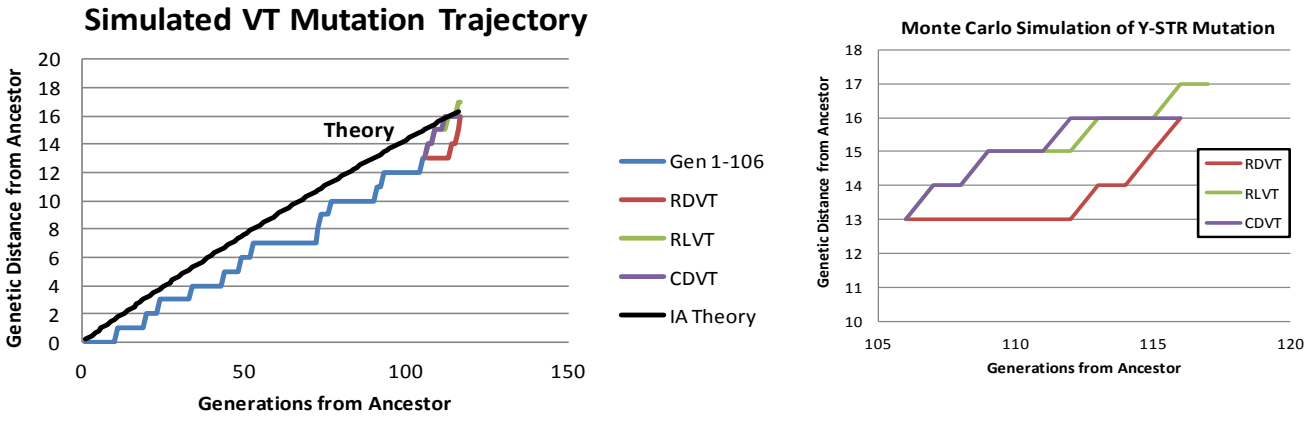


Fig. 3-6. Monte Carlo simulation resulting in one possible trajectory for the mutation record observed among the three Van Tuyl subjects. Fewer than 1000 iterations were required to produce this record. A complete trajectory is shown on the left, the final 11 generations from the MRCA AVT on the right.

The Van Tuyl "Signature" STRs

Y-STR analysis typically focuses on what is *different* between individuals. But it can be particularly revealing to observe what a family group has *in common* with one another. In fact, the Van Tuyls have mutated significantly from what is called the R1b1a2a1a1 haplotype. When we compare just 8 loci, we find that the Van Tuyl haplotype, OVT, varies in three out of the eight loci from the R1b1a2a1a1 ancestral haplotype.

			AHT	OVT	RVT/CDVT	RDVT
1	DYS	393	13	12	12	12
2	DYS	390	24	24	23	24
3	DYS	19	14	14	14	14
4	DYS	391	11	11	11	10
5	DYS	388	12	12	12	12
6	DYS	392	13	13	13	13
7	DYS	448	19	20	20	20
8	DYS	438	12	13	13	13
Non-Van Tuyl Occurences in ysearch database			**>1000**	**0**	0	1
Non-Van Tuyl Occurences in Sorenson database			2898	0	0	1

Whereas there are thousands of matches to the 8-locus ancestral haplotype in the ysearch database[6] and Sorenson database[7], for the OVT ancestral haplotype, changing these three loci by just one repeat causes the spurious matches to disappear. The three key loci characterizing the Van Tuyl family are seen to be: DYS393=12; DYS448=20; DYS438=13. These three numbers alone are highly likely to identify a person as a Van Tuyl.

[6] www.ysearch.org
[7] http://www.smgf.org/

Section 4: Conclusion and Suggestions for Further Research: How to join the Project

Results to Date:

1. We have demonstrated the genetic relationship between 3 Van Tuyl Cousins
2. We have derived the ancestral haplotype for their 17th century common ancestor
3. We have shown that 3 key numbers: DYS393=12; DYS438=13; DYS448=20, all available with a 37-locus test, are sufficient to identify a person as a descendant of the 17th century Dutch farmer Ott Jansz van Tuyl of Gameren
4. We have proven that A.W. Van Tuyl is in fact not a genetic Van Tuyl, but is rather descended from a member of the Terwilliger family of New York

Future Work:

The Van Tuyl family project Y-STR data are now publicly accessible at http://www.familytreedna.com/public/VanTuyl

We hope to attract more people worldwide to the project. Of particular interest are the following:

1. Dutch and American Van Tuyls unclear on their actual relationship to the family
2. Members of the family Van Tuyll van Serooskerken[8]. Proving or disproving their relationship to the Van Tuyl family is of particular historical interest
3. Members of the family in the US and The Netherlands who are listed in the genealogy of *A Van Tuyl Chronicle*, and who would like to help us refine and improve the conclusions in this report
4. People with the name Van Tuyl who suspect they are from an adopted line, and who would like to prove or disprove their genetic relationship to the family.

If you, or someone you know, has interest in joining this project, please contact Rory Van Tuyl at: roryvantuyl@gmail.com for details.

Or just order the recommended FTDNA 37-locus kit, which costs $149, at: http://www.familytreedna.com/group-join.aspx?&group=VAN_TUYL&vGroup=VanTuyl

For questions about this report, contact Rory Van Tuyl at: roryvantuyl@gmail.com

[8] http://en.wikipedia.org/wiki/Tuyll