

# Biogeographical Origins and Y-chromosome Signature for the House of Normandy

Michael R. Maglio

## Abstract

---

A sixty-seven marker short tandem repeat (STR) genetic analysis of Y-chromosomes reveals an unnoted modal haplotype showing a significant association with surnames claimed to have descended from the Norman dynasties. This suggests that such phylogenetic prevalence is a biological record, supports the reliability of early genealogies for Rollo 'The Dane' and the House of Normandy. Current Y-chromosome haplogroup maps only cover the broadest-brush strokes of the highest-level origins. Existing methods only generalize geographic patterns based on large population genetic frequency and SNP. Biogeographical Multilateration (BGM) illustrates the paternal ancestor migration flow as well as Rollo's chronological and physical origins at the individual haplotype level.

Published online 29 April 2014

## Introduction

---

Based on a study of 3,800 y-DNA records with surnames historically associated with William the Conqueror, a genetic signature of the first Norman King of England was identified (Maglio 2013). The intent of this current study is to increase the resolution of the 37 STR marker haplotype to 67 markers. The resultant is to be used in conjunction with biogeographical multilateration (BGM) (Maglio 2014) to resolve the contested origins of Rollo, the first Duke of Normandy and founder of the House of Normandy.

The earliest documented origins for Rollo come from Dudo of Saint-Quentin in 1015 and William of Jumièges in 1060. Both 'histories' were commissioned by the House of Normandy and attribute a Danish origin to Rollo. The Norwegian Orkneyinga Saga, from the 13<sup>th</sup> century, gives Rollo a Norwegian origin. The debate continues.

To determine the extended Rollo Norman Modal Haplotype (RNMH), the final 152 records from the Conqueror study, spanning 21 allied surnames, were updated to 67

markers and SNP data collected, where available.

How do we get to Rollo's origins at the individual haplotype level? Y-chromosome testing can illuminate distant origins of about 10,000 years ago or greater. Heat maps give us distribution of high level SNPs within the haplogroup. These distributions are based on the current locations of test populations at a macro level (Underhill et al 2001). To get to the specific details for Rollo, the biogeographical data of his genetic cousins is required. Geographic analysis can add validity to the modal haplotype, detect the genetic cluster from the *stationary* phase and identify *nomadic* migration routes.

New tools are available to resolve the *stationary* and *nomadic* phases.

## Methods

---

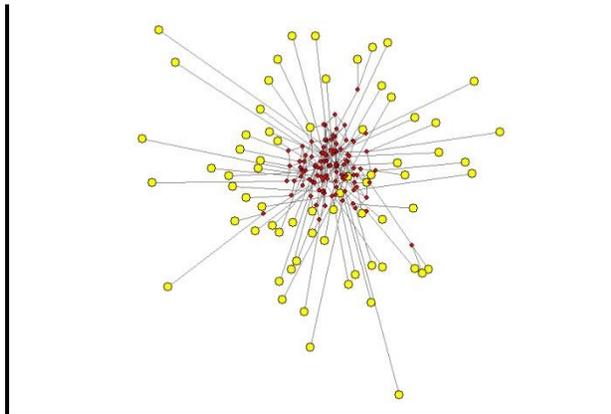
As part of the 67 STR marker modal haplotype stage, where possible, all records were updated from 37 to 67 markers and SNP data was added. An initial subset of records were eliminated due to lack of data. The remaining records were grouped by SNP and

the majority were retained. This final set consist of 25 records, all of which belong to SNP R-DF13. The resulting Rollo Norman Modal Haplotype (RNMH) is shown in Fig. 1.

DYS 393	DYS 390	DYS 19/39A	DYS 391	DYS 385a	DYS 385b	DYS 426	DYS 388	DYS 439	DYS 389-1
13	24	14	11	11	14	12	12	12	13
DYS 392	DYS 389-2	DYS 458	DYS 459a	DYS 459b	DYS 455	DYS 454	DYS 447	DYS 437	DYS 448
13	29	17	9	10	11	11	25	15	19
DYS 449	DYS 464a	DYS 464b	DYS 464c	DYS 464d	DYS 460	GATA H4	YCA IIa	YCA IIb	DYS 456
29	15	15	17	17	11	11	19	23	15
DYS 607	DYS 576	DYS 570	CDY a	CDY b	DYS 442	DYS 438	DYS 531	DYS 578	395S1a
15	17	17	36	37	12	12	11	9	15
DYS 395S1b	DYS 590	DYS 537	DYS 641	DYS 472	DYS 406S1	DYS 511	DYS 425	DYS 413a	DYS 413b
16	8	10	10	8	10	10	12	23	23
DYS 557	DYS 594	DYS 436	DYS 490	DYS 534	DYS 450	DYS 444	DYS 481	DYS 520	DYS 446
16	10	12	12	16	8	12	22	20	13
DYS 617	DYS 568	DYS 487	DYS 572	DYS 640	DYS 492	DYS 565			
12	11	13	11	11	12	12			

**Fig. 1** Rollo Norman Modal Haplotype (RNMH)

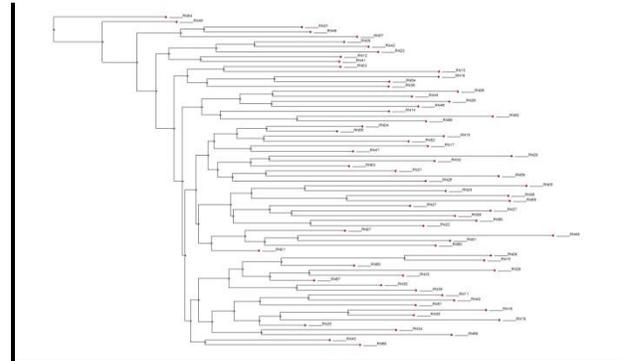
During the second stage of data collection, the newly determined RNMH is used to identify genetic cousins. Record selection is restricted to those containing the most distant known paternal ancestor (MDKPA) and self-reported origin.



**Fig. 2** MJ Network of 67-STR marker y-DNA haplotypes collected based on Rollo Norman Modal (n = 66). Each circle represents a unique haplotype. Analysis shows a coherent cluster. Haplogroup R-P312

The second stage dataset consists of 65 records from the RNMH 67 STR marker haplotype and upstream SNPs within a slightly broader R-P312 group (YCC 2002). This allows the analysis to calculate distant paternal ancestors. Time to most recent common ancestor (TMRCA) is generated to a 95% confidence (Walsh 2001) using FTDNA

derived mutation rates. Fluxus phylogenetic data was analyzed to determine whether a coherent cluster existed. Fig. 2 shows a single distinct cluster, indicating a common R1b SNP branch (R-P312) as expected.



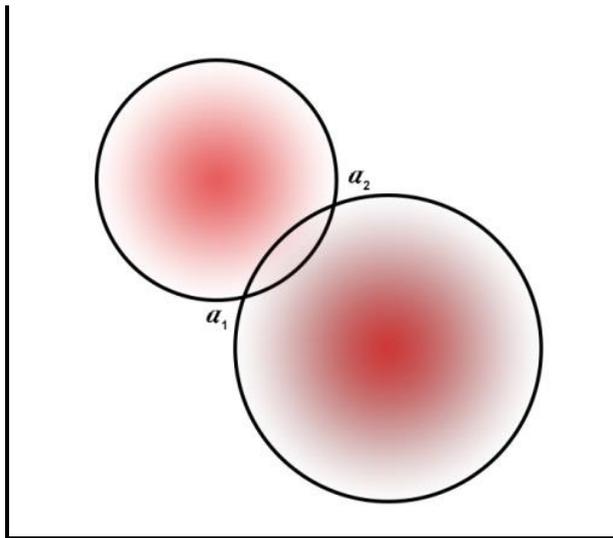
**Fig. 3** Phylogenetic tree (n=66) haplogroup R-P312

After this validation step, the output is then used by the *Neighbor-joining* method, which is part of the PHYLIP package for inferring phylogenetic relationships. A phylogenetic tree, Fig. 3, and chronological distances are produced for the data set. Data points for each genetic cousin are mapped using self-reported genealogical origin, Fig. 4.



**Fig. 4** Self-reported genealogical origins

A radius is drawn on a Mercator projection for each mapped record. The radius is calculated by multiplying the TMRCA by the upper value of the Neolithic migration rate of 30 km per 25 years or 1.2km/yr (Cavalli-Sforza 2002, Hazelwood et al 2004). This gives us the distance to most recent common ancestor (DMRCA) for each node in the tree network. The resulting intersection between pairs, Fig. 5, illustrates the approximate location of a common paternal ancestor (PA).



**Fig. 5** Bilateration analysis

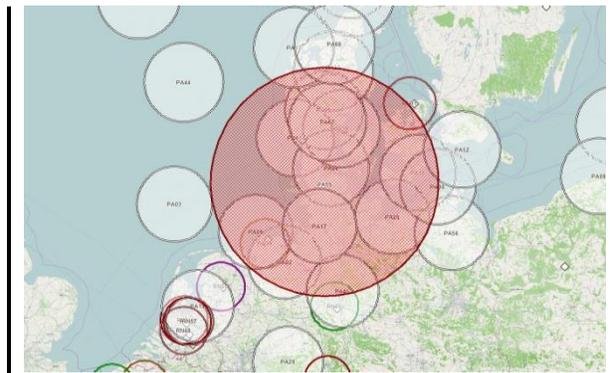
Bilateration analysis of pair data proceeds through the tree network of phylogenetic data defining haplotype directional flow as well as chronologic and geographic origins.

Haplotype data from this analysis is available in [Table 1](#) and [Table 2](#). DMRCA data is available in [Table 3](#).

## Discussion

Observations can be made about the raw mapped data. It is not surprising to see a number of points along the coast of Normandy. There is also a large number of points along the Rhine River. A scattering of points are found in the Baltic and Scandinavian regions. Only one point is in Denmark. Raw

mapping alone is misleading. You will only see the geographic correlations. SNP R-L21 distribution gradients (Myres 2011) will give you a similar plot. Heat maps can display the distribution of SNPs based on the current day location of samples, but they can only infer the location of a founder or common ancestor. Every pair of genetic cousins mapped will have a common ancestor. Phylogenetic analysis will determine best pairing within the set of data. Those best genetic pairs may have no geographic relationship. They will describe the distance and location of their common ancestor. The genetic cousin data points and their derived paternal ancestor data points define nodes of a phylogenetic tree network and a geographic network. Figure 6 shows the final simulation of the Rollo network, consisting of 66 cousin points and 66 paternal ancestor points. Where each of the cousin nodes may have an arbitrary geographic relationship, each of the 132 total tree nodes satisfies the constraints of geography, distance and relationship within the network.

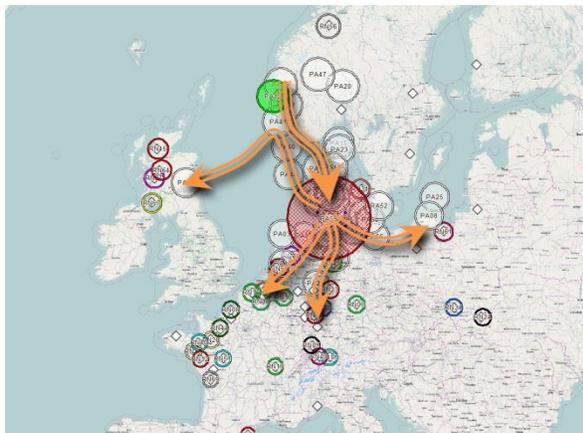


**Fig. 6** Multilateration analysis of paternal ancestors, back-calculating the location of PA59.

In the case of Rollo, we are trying to determine his origin. As expected, in Figure 4, there is no marker for Rollo. In the tree network, Rollo does have a common ancestor labelled node PA59, which is constrained by its peer nodes to be geographically located in Denmark. Based on the full analysis, Rollo's origins are within 226 km of PA59,

represented by the red circle in Fig. 6. While this does not indicate an exact city of origin, it does cover the majority of Denmark (Jutland).

The phylogenetic tree network is used to connect the nodes and show migration flow, Fig. 7. There is a distinct migration from Western Norway into Denmark. Multiple generations then cluster in Denmark, sending out unique branches into mainland Europe. The data indicates a back migration from Denmark (Jutland) to Norway and then on to the British Isles.



**Fig. 7** Simplified migration flow based on phylogenetic tree network.

## Conclusions

A small sample of 65 records was used in this analysis for simplification. Much larger data sets are recommended and would be required to determine the genetic flow in a greater geographic and chronologic view. Additional records within the same SNP grouping could result in a more accurate origin for Rollo. Records that are upstream from the SNP and STR group used in the study will indicate nomadic migration prior to the Western Norway settlement.

The results show that Rollo's origin was within 226 km of the center of Denmark and validates the earliest biographical accounts that he was Danish. In addition, Rollo's ancestors had Norwegian origins. These Norwegian ancestors are too distant to

account for Rollo's appearance in the Orkneyinga Saga.

The phylogenetic tree root distance can give us an estimated age of each common ancestor. PA12 is the oldest at 2,500 years ago ( $\pm 400$ ). Rollo's immediate common ancestor PA59 and the adjacent nodes cluster together in the next age range of 1,300 years ago ( $\pm 150$ ).

This study is specific to the RNMH and associated SNPs and may or may not apply to the larger R-P312 group.

This is an instance where the modal haplotype (RNMH) data does not have self-reported origins. The use of biogeographical multilateration (BGM) was able to narrow the geographic range and the analysis was able to solve for an unknown location.

BGM can be a major tool in developing genetic migration patterns at the individual haplotype level to bridge the gap between the modern era and the maps of our Neolithic origins.

## Conflict of Interest

The author declares no conflict of interest.

## Acknowledgements

I thank all of the DNA donors who have made their results publically accessible for review. Special thanks to Dean McGee for making his DNA analysis website available.

## Web Resources

Y-Utility: Y-DNA Comparison Utility, [http://www.mymcgee.com/tools/yutility.html?mode=ftdna\\_mode](http://www.mymcgee.com/tools/yutility.html?mode=ftdna_mode)

## References

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48

- Cavalli-Sforza LL (2002). Demic diffusion as the basic process of human expansions. Examining the farming/language dispersal hypothesis. Cambridge: McDonald Institute for Archaeological Research, 79-88.
- Hazelwood L, Steele J (2004) Spatial dynamics of human dispersals - Constraints on modelling and archaeological validation. *J ARCHAEOLOGICAL SCI*, 31 (6) 669 – 679
- Maglio MR (2013) A Y-Chromosome Signature of Polygyny in Norman England. *OriginsDNA*
- Maglio MR (2014) Y-Chromosome Haplotype Origins via Biogeographical Multilateration. *OriginsDNA*
- Myres, NM, Rootsi, S, Lin, AA, Järve, M, King, RJ, Kutuev, I & Underhill, PA (2011). A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *European Journal of Human Genetics*, 19(1), 95-101.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ & Cavalli-Sforza LL (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of human genetics*, 65(1), 43-62.
- Walsh B (2001). Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics*, 158(2), 897-912.
- Y-Chromosome-Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348