

Y Chromosome Haplogroup R1b-V88: Biogeographical Evidence for an Iberian Origin

Michael R. Maglio

Abstract

The status on the origins of haplogroup R1b remains split between the Iberian Peninsula prior to the Last Glacial Maximum (LGM) and a West Asian origin coinciding with the Neolithic expansion. The majority of focus on subgroup R1b-V88 has concentrated on African populations and neglected European and West Asian populations. To get a complete understanding of the back to Africa migration, a holistic network approach is necessary. Biogeographical Multilateration (BGM) illustrates directional flow as well as chronological and physical origins at the haplogroup level. The resulting phylogenetic relationships for R1b-V88 support an Iberian origin, a Mediterranean expansion and a Europe to Africa back migration.

Introduction

When we talk about the origins of haplogroup R1b, what we are really referring to is the origin of SNP R-M343. There is a consensus that the parent of R1b, R1-M173, has West Asian roots. Previously published papers have used samples from SNP R1b-M269 and downstream subgroups to provide ample records. Genetic records from upstream SNPs (V88, M73 and PF6398) closest to the root are better suited for identifying geographic origins. Focus on only high-density populations prevents a holistic view of the entire population. There is a strong biogeographical case for the Iberian origins of the parent SNP R1b-L278 (Maglio 2014) and evidence that a Neolithic Expansion origin is unjustified (Busby et al 2011). This paper will detail the use of new analysis tools to show evidence for an Iberian origin of R1b-V88, reinforcing the evidence for its parent, R1b-L278.

Considering that the V88 subclade is small, the number of publically available

records are few. Haplotype Aggregation, using known haplotypes to expand the record set, was required to achieve a reasonable dataset. That dataset is doubled through phylogenetic common ancestor reconstruction. The entire set is run through Biogeographical Multilateration (BGM) to determine origins and migration patterns. The results show a definitive Iberian origin, a phylogenetic backbone along the European Mediterranean coast, an ancient Europe to Africa back migration and a return migration to Europe (Fig 1).

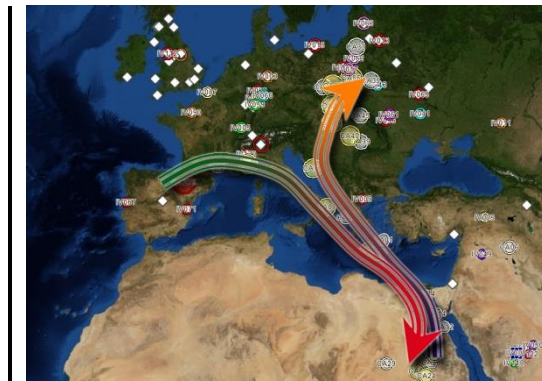


Fig. 1 R1b-V88 demonstrates an African back migration and subsequent remigration to Europe

Methods

As part of the data collection, initial Y-DNA record selection for the seed population is restricted to those that have at least 37 STR markers tested and detailed SNP testing showing positive for haplogroup R1b-V88. From a primary datasource (FTDNA), a dataset of 3,500 records returned 69 records meeting these criteria.



Fig. 2 Geographic locations of the amplified dataset. The next step is to plot common ancestors between highly related pairs.

Through Haplotype Aggregation, the haplotype of each seed record is used to identify a larger set of genetically related records from a secondary datasource (ySearch). Review of the resulting amplified dataset illuminates any errors in the seed population. The dataset has increased from 69 to 122 records with a high degree of genetic relation (Fig. 2).

Time to most recent common ancestor (TMRCA) is generated to a 95% confidence (Walsh 2001) using FTDNA derived mutation rates. This output is then used by the *Neighbor-joining* method, which is part of the PHYLIP package for inferring phylogenetic relationships. The high level subclade records form a haplogroup

backbone on the phylogenetic tree (Fig 3 & Fig 4).

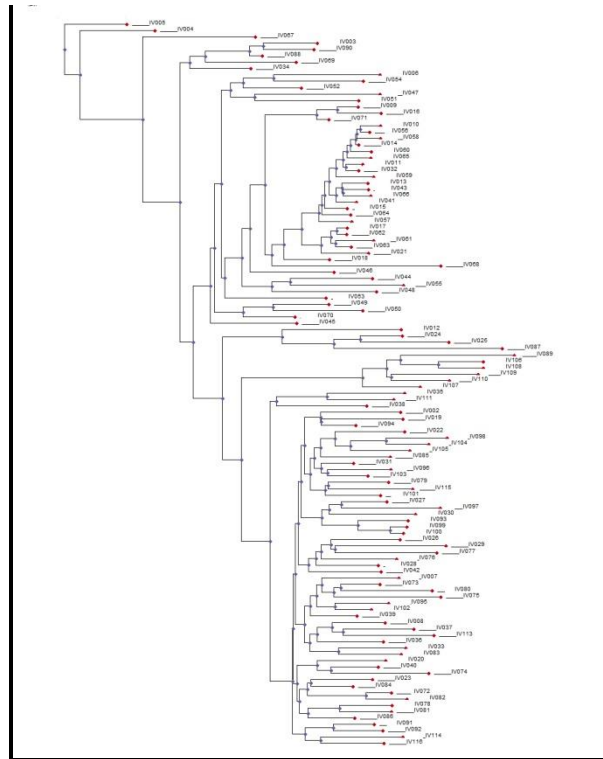


Fig. 3 Phylogenetic tree for R1b-V88 and R1b-M73.

The first pass of Haplotype Aggregation returned both additional V88 records and close genetic SNP M73. The phylogenetic tree places M73 downstream of V88. This is not an indicator that M73 is descended from V88. The tree is accurately organizing the records by genetic age. M73 is a relatively younger mutation. While the V88 and M73 STR markers were closely aligned, the significant mutation appeared as DYS464a&b. This is not a marker that is typically used to differentiate due to its high mutation rate. Values for V88 are 12,12 and values for M73 are 15,15. M73 records were removed from the dataset reducing it to 61 records with sufficient self-reported paternal ancestor geographic information.

The PHYLIP data is processed through Biogeographical Multilateration (BGM) (Maglio 2014), doubling the number of

records to 119 by reconstructing the haplotypes for the common ancestors on the

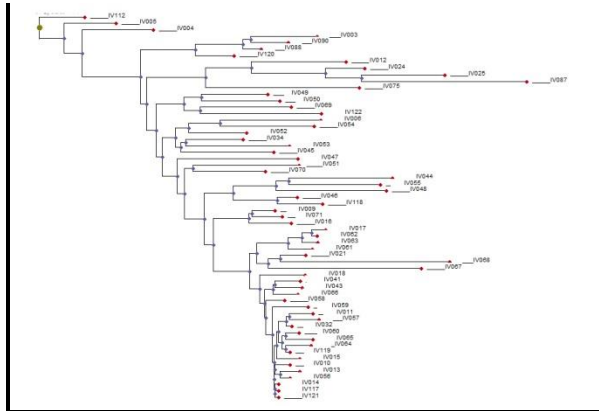


Fig. 4 Phylogenetic tree for R1b-V88.

tree. The combination of the self-reported ancestral locations from the amplified dataset and the estimated locations from the common paternal ancestors allows the V88 SNP backbone to be plotted geographically.

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

Discussion

An excellent case can be made for the spread of V88 with the spread of Chadic languages (Cruciani, 2010). However, this is an example of only sampling one region. Another example disregards genetic records outside of the Middle East and Africa (Myres, 2011). Both papers suggest a back to Africa migration for V88 from Asia without examining all the V88 records holistically. Locations or spatial-frequency distributions are based on the *current day* test populations (Underhill et al 2001). We must be careful not to misinterpret the genetic gradient of an organic process, as the direction of movement underlying a cline can be ambiguous (Chikhi et al 2002, Edmonds et al 2004). Current population densities have no correlation to historic migrations. The dataset, analyzed as a

network with phylogenetic relationships, has a vector quality that can illustrate migration patterns.

The V88 records from Cameroon, Chad, Nigeria and Sudan have an affinity for genetic cousins from Bahrain, Saudi Arabia and Yemen. BGM analysis (Fig 5) shows a coalescence of common ancestors along the Nile River Valley. The vector for each node on the phylogenetic tree is calculated for

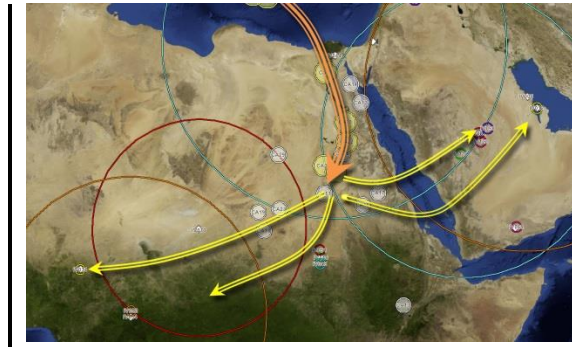


Fig. 5 Origins and migration for R1b-V88 within Africa. The TMRCA for the data analyzed has V88 entering the Nile River Valley at $5,500 \pm 1000$ ybp.

both location and distance to the next node. Walking the tree backwards takes us out of Africa and back to Europe. Multiple simulations were run. The island of Crete consistently acted as a stepping-stone between Adriatic common ancestor locations and Nile River locations. The



Fig. 6 Origins and migration for R1b-V88. The TMRCA for the data analyzed has R1b-V88 branching from the L278 backbone at $7,700 \pm 1,600$ ybp.

BGM analysis traces the genetic connection back westward along the Northern Mediterranean coast to align with its R1b-L278 predecessor origins in Iberia (Fig 6).

The BGM analysis (Fig 5 & 4) demonstrates a clear path of migration and relationship from the backbone R1b haplogroup in Iberia back to Africa and then



Fig. 7 R1b-V88 migration back to Europe from Africa. The TMRCA for the data analyzed has V88 returning to Europe at $3,200 \pm 1000$ ybp.

a subsequent dispersal east and west of the Nile River Valley.

The BGM analysis also demonstrates a back to Europe migration (Fig 7). There is a genetic flow back across the Mediterranean to Peloponnese and the Eastern Adriatic. Again, Crete plays a role as a stepping-stone. It is uncertain whether this re-migration is an example of colonization. There is then dispersal into Central and Eastern Europe. In some of these Eastern Europe populations, there is a high degree of relationship. This is an example of the founder effect.

Conclusions

A small sample of 69 records was used as the seed population. This was amplified to 119 records. This population gives an out of Iberia TMRCA of $7,700 \pm 1,600$ ybp, an into Africa TMRCA of $5,500 \pm 1,000$ ybp

and a re-migration to Europe TMRCA of $3,200 \pm 1,000$ ybp. While TMRCA calculations are notoriously inaccurate, these calculations give a relative chronological magnitude. Genetic data is too often treated as discrete units having no interaction. SNP populations tend to be analyzed in a vacuum. Population genetics needs to be viewed as a network analysis.

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

References

- Busby, GB, Brisighelli, F, Sánchez-Diz, P, Ramos-Luis, E, Martinez-Cadenas, C, Thomas, MG, ... & Capelli, C (2012). The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proceedings of the Royal Society B: Biological Sciences*, 279(1730), 884-892.
- Chikhi L, Nichols RA, Barbujani G, & Beaumont MA (2002). Y genetic data support the Neolithic demic diffusion model. *Proceedings of the National*

- Academy of Sciences, 99(17), 11008-11013.
- Cruciani, F, Trombetta, B, Sellitto, D, Massaia, A, Destro-Bisol, G, Watson, E, ... & Scozzari, R (2010). Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *European Journal of Human Genetics*, 18(7), 800-807.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci USA* 101: 975–979.
- 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 (2014) Biogeographical Evidence for the Iberian Origins of R1b-L278 via Haplotype Aggregation. *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 (2010) A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *European Journal of Human Genetics*, 19(1), 95-101
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A (2000). "The genetic legacy of paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective". *Science* 290 (5494): 1155–59
- 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