Biogeographical Evidence for the Iberian Origins of R1b-L278 via Haplotype Aggregation

Michael R. Maglio

Abstract

The status on the roots of haplogroup R1b remains split between an Iberian origin prior to the Last Glacial Maximum (LGM) and a West Asian origin coinciding with the Neolithic expansion. Existing methods generalize geographic patterns based on large population genetic frequency and diversity. Haplotype Aggregation delivers a coherent genetic record selection and Biogeographical Multilateration (BGM) illustrates directional flow as well as chronological and physical origins at the haplogroup level. The resulting phylogenetic relationships across multiple high level branches of R1b support an Iberian origin and a rapid Western Atlantic 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 presented R1b origins as Iberian, West Asian, Near East and South Asian (Semino et al 2000, Myres et al 2010, Klyosov 2012). This is far from a consensus. This paper will detail the use of new analysis tools to show evidence for an Iberian origin of R1b-L278, currently one branch below R1b-M343.

The process of genetic record sampling, collection and selection is critical. A poor collection strategy can return a record set with incoherent results. Biases in record collection, adaptive or otherwise, can lead to exactly the results that you expected. I use a semi-random seed method for record collection, which minimizes any bias that could be introduced. While there is geographic accuracy in using selected populations for genetic record collection, there is an inherent bias toward those populations. Evidence of existence is not evidence of origin.

These data sets are typically converted to a graphical presentation as a spatial-frequency distribution map, displaying the density of high level SNPs within the haplogroups. These distributions are based on the *current* locations of 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. How do we get to the historic origins and migration patterns at the haplogroup level?

Within haplogroup R1b, there are extremely large subclades, R-U106 and R-P312. While these subclades are important to the overall picture, their size leads to noise in the analysis of an R1b origin. It is the minority branches of R1b (R-L278*, R-V88, R-M73*, R-YSC0000072/PF6426 and R-L23-) that provide the resolution required. (While the data from R-V88 supports an Iberian origin, it will be handled in a

separate paper due to its complexity.) Considering that these subclades are small, the number of publically available records are few. Seed records were selected from these subclades. The seed pool is expanded using Haplotype Aggregation. The resulting set is doubled through phylogenetic common ancestor reconstruction. The entire dataset is run through Biogeographical Multilateration (BGM) to determine origins and migration patterns. The results show a definitive Iberian origin, a phylogenetic backbone along the Western Atlantic coast and ancient eastward migration patterns for the minority subclades.

Methods

As part of the data collection, initial record selection for the seed population is restricted to those that have at least 37 STR markers tested, detailed SNP testing showing positive for high level haplogroup R1b branches and negative for lower level branches. From a primary datasource (FTDNA), a dataset of 3,500 records returned 111 records meeting these criteria. Using publically available genetic data sources introduces an inherent bias in the seed record collection and an uncertainty in their provenance. To remove the bias and uncertainty, the records are put through an amplification process.

Fig. 1 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 111 to 253 records with a high degree of genetic relation (Fig. 1). The Haplotype Aggregation process will return records that are not SNP tested or have SNPs outside the seed population. This is expected, as nearly identical haplotypes can exist within multiple related SNP branches. This is a by-product of having a common paternal ancestor.

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 2).

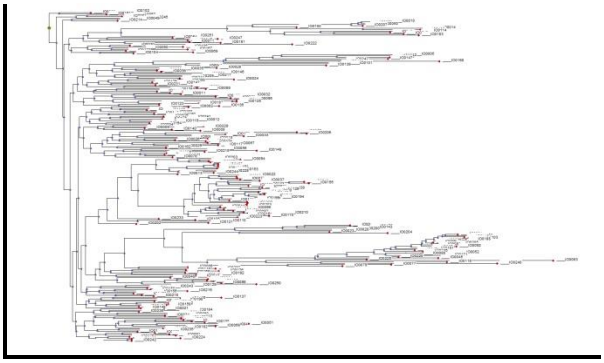

Fig. 2 Phylogenetic tree for high level R1b branches.

The PHYLIP data is processed through Biogeographical Multilateration (BGM) (Maglio 2014), doubling the number of records to 508 by reconstructing the haplotypes for the common ancestors on the tree. The combination of the self-reported ancestral locations from the amplified

dataset and the estimated locations from the common paternal ancestors allows the haplogroup backbone to be plotted. Figure 3 shows the phylogenetic backbone along the Western Atlantic coast, with R-L278 origins south of the Pyrenees.

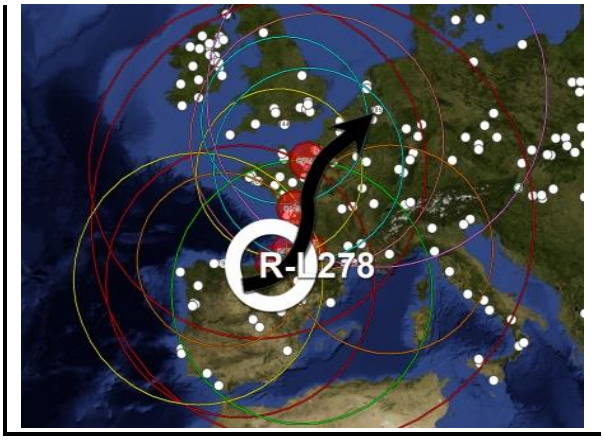

Fig. 3 The BGM analysis resolves the ancestral locations based on the relationship within the phylogenetic tree.

Haplotype data from this analysis is available in Table 1 and Table 2.

Discussion

We know that modern humans survived and flourished in the Iberian refugium during the Last Glacial Maximum (LGM) (Pereira et al 2005) based on mitochondrial DNA studies. The tribes in western Europe had a 1,000 to 2,500 year head start over the tribes in central and eastern Europe on repopulating the continent, as the ice sheets melted and retreated earlier on the west coast (Hoffecker 2006). This gave the inhabitants of the Iberian refugium an anthropological / geographical first-mover advantage – an advantage gained by the initial ("first-moving") significant occupant of a geographical area. This advantage may stem from the fact that the first entrant can gain control of resources that followers may not be able to match. First-movers can gain a land-monopoly status, a defensive and offensive high ground advantage (Lieberman et al 1988 – adapted from business strategy, applied to population science).

A tribe with a first-mover advantage and over a 1,000 year head start should be hard to displace. Evidence supports that haplogroup R1b is that original tribe. The alternative would be to suggest that R1b completely displaced an earlier Paleolithic tribe from Iberia to Scandinavia. There is no evidence to support that suggestion.

Haplogroup R1b had a recipe for success – separation, time and geography. With each generation, the population dispersed to exploit available resources (Hazelwood et al 2004)

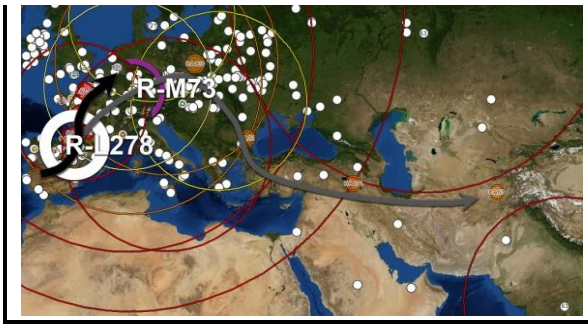

Fig. 4 Origins and migration for R-M73. The TMRCA for the data analyzed has R-M73 branching from the backbone at 4,200±900 ybp.

There is genetic evidence of small R1b subclades in West Asia, the Near East and Africa. These records are not examples of origin. They are examples of migration.

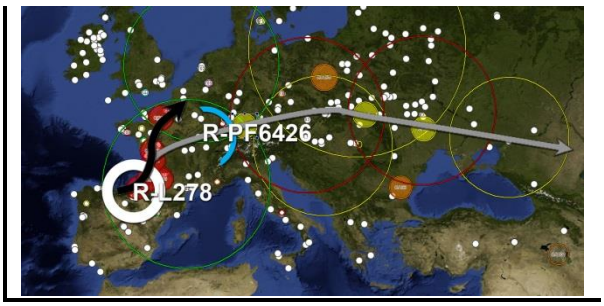

Fig. 5 Origins and migration for R-PF6426. The TMRCA for the data analyzed has R-PF6426 branching from the backbone at 2,900±700 ybp.

The BGM analysis (Fig 4 & 5) demonstrates a clear path of migration and relationship from the backbone R1b haplogroup in western Europe to these isolated populations. In some of these locations, there is a high percentage of R1b in the population. This is an example of the founder effect.

Conclusions

A small sample of 111 records was used as the seed population. This was amplified to 253 records. This population gives a TMRCA of 6,600 ± 1,400 ybp. This assumes a constant mutation rate over time. The size of the R1b population has the potential to accelerate the mutation rate (Elena et al 2007, Jiang et al 2010). Any increase in the documented mutation rates would result in a shorter TMRCA. A larger dataset should provide a more accurate period. This age calculation does not represent the age of R1b in Iberia. The calculation is a better representation of the time at which the smaller subclades divided from the main branch.

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

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.

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.

Elena SF, Wilke CO, Ofria C, & Lenski RE (2007) Effects of population size and mutation rate on the evolution of mutational robustness. Evolution, 61(3), 666-674

Hazelwood L, Steele J (2004) Spatial dynamics of human dispersals - Constraints on modelling and archaeological validation. J ARCHAEOL SCI , 31 (6) 669 – 679

Hoffecker J (2006) A Prehistory of the North: Human Settlements of the Higher Latitudes. Rutgers University Press: New Jersey

Jiang X, Mu B, Huang Z, Zhang M, Wang X, & Tao S (2010) Impacts of mutation effects and population size on mutation rate in asexual populations: a simulation study. BMC evolutionary biology, 10(1), 298

Klyosov AA (2012) Ancient history of the Arbins, bearers of haplogroup R1b, from Central Asia to Europe, 16,000 to 1500 years before present. Advances in Anthropology, 2(02), 87

Lieberman MB, & Montgomery DB (1988) First-mover advantages. Strategic management journal, 9(S1), 41-58

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

Pereira L, Richards M, Goios A, Alonso A, Albarrán C, Garcia O, ... & Amorim, A (2005) High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome research, 15(1), 19-24

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