BigY demo (personalised report); 10th June 2014
https://www.clarifYDNA.com

An experimental computer-generated Y-chromosomal phylogeny, leveraging BigY results and the current ISOGG tree (Big Y[1] Demo)

clarifYDNA

10th June 2014

The contents of this report were produced by a computer algorithm. This report will be frequently re-generated as more information becomes available.

The *pilot-scale* implementation of this algorithm is able to process a dataset of over 4000 Big Y kits (over 400 real and 3600 simulated) in one run. clarifY DNA's automation capabilities analyse large Y-SNP datasets with great speed, great accuracy and great comprehensiveness. These facets are critical for: helping a testing company's customers make informed SNP-ordering decisions; uniting customers and/or research participants with their most meaningful patrilineal matches; and, overall, scientific progress, customer satisfaction and further growth.

All in all, clarifY DNA's software is the key to truly realising the "Y Tree" in "Family Tree".

The phylogenetic algorithm employed here was initially developed in June 2013 for Geno 2.0 data; see http://ytree.morleydna.com/experimental-phylogeny for similar reports (from an earlier version of the phylogenetic algorithm) leveraging public Geno 2.0 data. While this report represents a large advance over existing Y-DNA trees, please treat some aspects of this report as experimental and preliminary; some enhancements specific to next-generation sequencing have not been exhaustively tested, and there are several discrepancies over the definitions of high-level SNPs.

Terminal-level SNPs marked as recurrent (^{R}) or preceded by a question mark should be taken with some caution. Many of these could be spurious.

This report mostly follows the same conventions as the reports at http://ytree.morleydna.com/experimental-phylogeny. There are a few differences which have not yet been documented. Unnamed novel variants are prefixed with "C:". Novel variants which are suspiciously close to other novel variants are marked with ^{C}.

If you found this service useful, please help us spread awareness. Or consider making a donation to offset the months of research and development that went into it.

This report may not be used for commercial purposes.

Reference: Base tree

[1] International Society of Genetic Genealogy, *Y-DNA Haplogroup Tree 2014, Version: 9.60*. http://www.isogg.org/tree/, 9 May 2014 (accessed 10 May 2014).

References: SNP names

[1] International Society of Genetic Genealogy, *Y-DNA SNP Index, Version: 9.60*. http://isogg.org/tree/ISOGG_YDNA_SNP_Index.html, 9 May 2014 (accessed 10 May 2014).

[2] ISOGG YBrowse, http://ybrowse.isogg.org/cgi-bin/gb2/gbrowse/chrY/, accessed 2 May 2014.

[3] The genetic genealogy community, *Analysis of public Y chromosome "next gen" sequencing datasets* (Personal Genome Project and 1000 Genomes Project datasets), https://sites.google.com/site/yanalysis/ (accessed 4 March 2014).

[4] YFull team, *YFull.com Y-SNPs* (public spreadsheet of coordinates for Y- and YP- series SNPs), (accessed 2 May 2014).

[5] James Wilson, *BritainsDNA Y-SNPs (chromo2 YSNP alt names.xlsx)*. via *Your Genetic Genealogist*, http://www.yourgeneticgenealogist.com/2013/11/a-list-of-alternate-snp-names-for.html, 20 November 2013 (accessed 2 December 2013).

[6] James Wilson, *BritainsDNA Chromo2 positions (BritainsDNAchromo2positions.xlsx)*. via *Your Genetic Genealogist*, http://www.yourgeneticgenealogist.com/2014/03/dr-jim-wilson-and-scotlandsdna-release.html, 4 March 2014 (accessed 4 March 2014).

[7] Chuan-Chao Wang and Hui Li, *Discovery of Phylogenetic Relevant Y-chromosome Variants in 1000 Genomes Project Data*, 2013 (arXiv:1310.6590), http://comonca.org.cn/Y_tree/.

[1] Big Y is a product offered by Gene By Gene, Ltd. clarifY DNA has no affiliation with this company.

Report format and phylogenetic analysis copyright © 2013-2014 Chris Morley (o/a clarifY DNA)

BigY demo (personalised report); 10th June 2014
https://www.clarifYDNA.com

BigY demo (personalised report); 10th June 2014 https://www.clarifYDNA.com

Experimental Y tree (automated, using BigY results)

1 **R1a**[2] M420/L146/PF6229 [28/28], M449/L145/PF6175 [28/28], M511/L63/PF6203 [28/28], M513/L62/PF6200 [28/28] (between R1a and R1a1a1: CTS– [28/28], CTS– [27/28], CTS– [28/28], CTS– [28/28], CTS– [28/28], CTS– [28/28], CTS– [28/28], F– [27/27], F– [28/28], F– [28/28], FGC– [28/28], FGC– [28/28], L–/PF– [28/28], M–/CTS–/PF– [28/28], M–/CTS–/PF– [28/28], M–/F–/PF– [28/28], M–/CTS–/PF– [28/28], M–/F– [28/28], M–/F– [28/28], M–/PF– [28/28], M–/PF– [28/28], M–/F–/PF– [28/28], M– [28/28], M–/F– [28/28], M–/PF– [27/27], M– [28/28], M– [28/28], M–/PF– [28/28], M–/PF– [28/28], M–/CTS–/PF– [28/28], M–/F–/PF– [28/28], M–/CTS– [28/28], M–/F–/PF– [28/28], M–/CTS– [28/28], M–/CTS–/PF– [28/28], M–/CTS– [28/28], M–/CTS– [28/28], M–/CTS–/PF– [28/28], M–/CTS–/PF–^{R_2} [28/28], M–/CTS–/PF– [28/28], M–/CTS– [28/28], M–/CTS– [28/28], M–/CTS–/PF– [28/28], M–/CTS–/PF– [28/28], M–/F–/PF– [28/28], M–/F–/PF– [28/28], M–/F–/PF– [28/28], M–/CTS–/PF– [28/28], M–/CTS–/PF– [28/28], M–/CTS–/PF– [28/28], M–/CTS– [28/28], M–/F–/PF– [28/28], M–/CTS– [28/28], M–/F–/PF– [28/28], M–/CTS– [28/28], M–/CTS–/PF– [28/28], M–/CTS– [24/24], M–/CTS– [28/28], M–/F–/PF– [28/28], M–/F– [28/28], M–/CTS–/PF– [28/28], ?M–/F–/VL–^{R_2} [28/28], M–/PF– [28/28], M–/F– [28/28], M–/F– [28/28], M–/PF– [28/28], M–/PF– [28/28], M–/PF– [28/28], M–/PF– [28/28], M–/F– [28/28], M–/F– [28/28], M–/F– [28/28], M–/PF– [28/28], M–/PF– [28/28], M–/PF– [28/28], M–/Y– [28/28], M–/CTS– [28/28], M–/CTS–/PF– [28/28], M–/CTS– [26/26], M–/F–/PF– [28/28], M–/CTS– [28/28], M–/CTS–/PF– [28/28], M–/CTS– [28/28], M–/CTS–/PF– [28/28], M–/CTS–/PF– [28/28], M–/F–/PF– [28/28], M–/F–/PF– [28/28], M–/Y– [28/28], M–/Y– [28/28], M–/F–^{R_2} [28/28], PAGE–/PF–/SRY–/SRY–/PAGE–.– [28/28], PF–/F– [27/27], PF–/F– [28/28], PF– [28/28], PF–/CTS– [28/28], PF–/CTS– [28/28], PF– [28/28], PF–/CTS– [28/28], PF–/F– [28/28], PF–/F– [28/28], PF–/PAGE–/SRY–/SRY–/PAGE–.– [28/28], PF– [28/28], PF– [28/28], PF–/F– [28/28], PF– [26/26], C:– [28/28], C:– [28/28], C:– [28/28], C:– [28/28], C:– [28/28], C:– [28/28], C:–^{R_3} [27/27], C:– [25/27], C:– [28/28], ?S–^{R_2} [28/28], Y– [28/28], Y– [28/28], Y– [28/28], Y– [28/28], Y– [28/28], Y– [28/28])

2 ↳**R1a1** M448/L122/PF6237 [28/28], M459/PF6235 [28/28], M516/L120/PF6236 [28/28], SRY10831.2/PAGE65/PF6234/SRY1532/SRY10831/PAGE65.1 [28/28] (between R1a and R1a1a1:)

3 ↳**R1a1a** L168 [28/28], L449/PF6223 [28/28], M17 [0/28], M198/PF6238 [28/28], M512/PF6239 [28/28], M514/PF6240 [28/28], M515 [28/28]

4 ↳**R1a1a1** M417 [28/28], PAGE7/PAGES00007 [18/18] (*Cannot resolve*: CTS– [13/27], CTS– [20/20], CTS–/M– [19/19], CTS–/M– [17/17], CTS–/M– [21/21], FGC– [20/20], PF– [23/24], PF– [25/25], C:– [26/27], C:– [26/26], C:– [24/26], C:– [24/26], C:– [26/28], ?C:–^{R_3} [18/18], C:– [25/27], C:– [26/27], C:– [16/16], C:–^{R_2} [17/17], C:– [25/25], C:– [8/8], C:– [11/11], C:– [20/20], C:–^{C} [19/19], C:–^{C} [21/22], C:–^{C} [24/24], C:–^{C} [25/25], C:–^{C} [25/25], C:–^{C} [23/23], C:– [22/22], C:–^{R_2} [16/16], C:– [13/14], C:– [15/15], C:– [14/14], Y– [26/26], Y– [20/20])

5 ↳**R1a1a1b**[3] Z645/S224/PF6162 [21/21], Z647/S441/PF6158 [21/21] (between R1a1a1b and R1a1a1b1a: F–^{R_2} [21/21], M–/CTS–/PF–/Z– [21/21], M–/CTS–/PF–/Z– [21/21], M–/F–^{R_2} [21/21], C:–^{R_2} [21/21], Z–/CTS– [21/21], Z–/F– [21/21])

6 ↳**R1a1a1b1** Z283/PF6217/S339 [17/17]

7 ↳**R1a1a1b1a** Z282/S198 [21/21] (*Cannot resolve*: ?C:–^{R_3} [14/14], C:–^{R_2} [5/6], ?C:–^{R_2} [4/4], ?C:–^{R_2} [6/6], ?C:–^{R_2} [8/10])

8 ↳**R1a1a1b1a3** FGC– [11/11], Y– [11/11], Z284/S221 [11/11], Z– [11/11], Z289/S443

9 ↳**R1a1a1b1a3a** CTS1123 [9/9], CTS9808 [10/10], CTS11934 [10/10], L448/S200 [10/10], S4458 [10/10], S5153 [10/10], S5301 [10/10], S5302 [10/10] (*Cannot resolve*: FGC11881 [7/7], FGC11910 [4/4], C:25031104(A|G) [4/4], ?C:27894346(A|T)^{R_3} [4/4], ?C:28803952(G|T)^{R_3} [6/6], ?C:28805012(T|A)^{R_3} [3/3])

10 ↳**R1a1a1b1a3a1** CTS3438 [7/7], CTS4179 [8/8] (*Cannot resolve*: C:7450195(GAAGTCATTTTTGAAGACAC|G) [5/5], ?Y1967^{R_2} [1/1])

11 ↳**R1a1a1b1a3a1–1**[4] 21 “private” SNPs

12 ↳CLd1

13 ↳**R1a1a1b1a3a1–2** CTS– [1/1], M–/F–^{R_2} [1/1], M–^{R_3} [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:– [1/1], C:–^{R_3} [1/1], S– [1/1], S– [1/1], S–^{R_2} [1/1][5]

14 ↳CL00 [DEMO-BigY-Data-20140410]

15 ↳**R1a1a1b1a3a1–3**[6] 23 “private” SNPs

16 ↳CLd2

17 ↳**R1a1a1b1a3a1–4**[7] 23 “private” SNPs

18 ↳CLd3

19 ↳**R1a1a1b1a3a1–5** (between R1a1a1b1a3a1–5 and R1a1a1b1a3a1a: FGC– [4/4], FGC– [4/4], FGC– [4/4], FGC– [4/4], FGC– [4/4], FGC– [4/4], C:– [4/4], YP– [4/4])

20 ↳**R1a1a1b1a3a1a**[8] FGC11888 [4/4], L176/S179^{R_2} [3/4] (*Cannot resolve*: ?CTS–/M–/PF–^{R_3} [1/1], ?FGC–^{R_3} [3/3], L–/S–/YSC–^{R_2} [2/2], C:– [2/2], ?C:–^{R_3} [1/1], Y– [3/3], ?Y–^{R_3} [1/1])

21 ↳**R1a1a1b1a3a1a–1**[9] FGC– [3/3], FGC– [3/3], FGC– [3/3], FGC– [3/3] (*Cannot resolve*: C:–^{R_3} [2/3])

22 ↳**R1a1a1b1a3a1a–1–1**[10] FGC– [2/2], K– [2/2], C:– [2/2], C:–^{R_2} [2/2], C:–^{C} [2/2], C:–^{C} [2/2], C:–^{C} [2/2], C:–^{C} [2/2], C:–^{C} [2/2], C:–^{C} [2/2], C:– [2/2]

23 ↳**R1a1a1b1a3a1a–1–1–1**[11] 22 “private” SNPs

24 ↳CLd4

25 ↳**R1a1a1b1a3a1a–1–1–2**[12] 12 “private” SNPs

26 ↳CLd5 (Recent unexpected negative call: L– [3 levels above, at R1a1a1b1a3a1a].)

27 ↳**R1a1a1b1a3a1a–1–2** 15 “private” SNPs

28 ↳CLd6

29 ↳**R1a1a1b1a3a1a–2** 16 “private” SNPs

30 ↳CLd7

31 ↳**R1a1a1b1a3a–1** YP– [2/2]

32 ↳**R1a1a1b1a3a–1–1**[13] 23 “private” SNPs

33 ↳CLd8

34 ↳**R1a1a1b1a3a–1–2**[14] 29 “private” SNPs

35 ↳CLd9

---

[2] M– also found in nearby clade R2 . One set of instances may be erroneous. Further investigation required.
PF– may have a wider scope than this subclade.

[3] F–, M– and C:– also found in nearby clade R1a1a1a–2 . One set of instances may be erroneous. Further investigation required.

[4] Neighbouring kit CL00 lacks calls for: C:–, C:–, C:–.
Neighbouring kit CLd2 lacks calls for: C:–, C:–, C:–, C:–.
Neighbouring kit CLd3 lacks calls for: C:–, C:–, C:–.

[5] These markers are (for the moment) specific to the report subject. They would not be blanked out in non-demo versions of this report.

[6] Neighbouring kit CLd1 lacks calls for: FD–, M–, C:–.
Neighbouring kit CL00 lacks calls for: FD–, M–, C:–, C:–, C:–.
Neighbouring kit CLd3 lacks calls for: FD–, M–, C:–, C:–.

[7] Neighbouring kit CLd2 lacks calls for: C:–.

[8] FGC11888 added to base tree because L176.1 is not consistently reported in the VCF files.

[9] Neighbouring kit CLd7 lacks calls for: FGC–.

[10] Neighbouring kit CLd6 lacks calls for: C:–.

[11] Neighbouring kit CLd5 lacks calls for: C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–.

[12] Neighbouring kit CLd4 lacks calls for: C:–, C:–, C:–, C:–, C:–, C:–, C:–.

[13] M– also found in nearby clade R1a1a1b1a2b3–1 . C:– also found in nearby clade R1a1a1b1a1a . One set of instances may be erroneous. Further investigation required.
Neighbouring kit CLd9 lacks calls for: CTS–, M–, C:–, C:–, C:–, C:–, C:–, C:–, C:–, C:–.

[14] Neighbouring kit CLd8 lacks calls for: C:–, C:–.