

# An experimental computer-generated Y-chromosomal phylogeny, leveraging public Geno 2.0 results and the current ISOGG tree

Chris Morley\*

26th July 2013†

## Abstract

The author of this report has created software to (i) facilitate the analysis of new Geno 2.0 data, and (ii) automate the preparation of a comprehensive draft Y-chromosomal phylogenetic tree incorporating publicly available Geno 2.0 Y-SNP data. The second item is presented in Part 2.

### Author's note

Thank you for your interest in this project!

This report was first made public on 16 July 2013. Since then it has gone very mildly viral. This is the second version of this report. I'd would still like to temper some of the excitement directed toward these initial versions of my experimental Y-tree. Please do not overlook the "experimental" nature of this tree. Some subclades proposed by my algorithm are known to be erroneous.

Stated another way, nobody should go out and get their linens embroidered with "M26→L707→YSC0000078" just because of what my experimental tree or predictor said.

The feedback I have received has been positive, encouraging and helpful. I will be the first to admit that further validation of this tree is needed. The computer-generated tree is too vast for one person to validate manually, so I am working on a separate set of tools that will identify – as a last step in the tree-generation process – a (hopefully) short list of inconsistencies in the experimental tree. Each inconsistency will need to be manually investigated in order to determine its cause. One way or another, I think I will be able to improve my algorithm so that the number of inconsistencies is reduced. Those wishing to devote extensive time to manually validating my tree should probably wait until the next release.

These is another class of inconsistencies – a class containing inconsistencies only detectable by haplogroup specialists. The size of this class remains to be seen.

So yes, I have some ideas on how to further refine this process and address some of the outstanding issues. But I will not be able to immediately act on my ideas and release an updated tree. And I have no doubt that further flaws in this tree will surface.

All in all, please regard this experimental phylogeny as informative rather than authoritative. As I wrote in my announcement, it's not perfect, but it's a start.

ISOGG volunteers and Y-DNA project administrators: if you are asked to explain differences between this tree and the ISOGG one, you can save yourselves some time by directing people to read [this FAQ item](#).

Chris Morley

## 1. The need for automation

### 1.1 Background

#### Genographic Project 2.0

The second phase of the Genographic Project [1] was announced in July 2012 [2, 3]. One facet of this project is wide-spectrum Y-SNP testing offered at an accessible price. The aggregation of these results will reinforce much of the existing Y-SNP phylogeny, while concurrently leading to the discovery of new branches and the slight repositioning of some existing branches.

The initial round of test results was released in December 2012 [4]. The Genographic Consortium is apparently preparing an updated Y-SNP phylogeny, but to the chagrin of genetic genealogy hobbyists it has not yet been released [5].

Manual analysis of this dataset is time-consuming. The vastness of the data is only one factor; there are several additional challenges to overcome before the data can be compared:

\*Morley Y-DNA Project Administrator; lead researcher, "Northwest England" cluster of R1a-L448. Contact: <http://www.MorleyDNA.com/contact>.

†An earlier version of this report, shaped as a proposal, was released for limited distribution on 5 July 2013.

- the dataset is growing every day;
- the phylogenetic nomenclature is in perpetual flux;
- the publicly available Geno 2.0 Y-SNP data is decentralised;
- many Y-SNPs go by several names.

Even then, after monitoring the above, there are many more challenges at the comparison stage:

- the main source for publicly available Geno 2.0 data (various Family Tree DNA haplogroup project Y-SNP reports, for example, [6]) does not distinguish between negative calls and no-calls;
- some chip-tested SNPs are proving to be globally erratic [7];
- other chip-tested SNPs may be locally erratic (inconsistent results for an SNP may indicate an inadequate choice of primer, or the presence of additional nearby mutation(s) in one or more derived branches);
- some kits have had additional à la carte testing performed for SNPs not offered on the Geno 2.0 chip;
- a few SNPs have in some branches reverted to the ancestral state, potentially resulting in a false-negative call;
- some SNPs are recurrent, and it can be challenging to separate the multiple instances;
- it is not straightforward to determine where newly-discovered SNPs fit in relation to existing SNPs;
- existing phylogenies, against which comparisons are made, may place some SNPs erroneously, complicating the comparison procedure.

The members of the genetic genealogy community – the majority of them part-time volunteers – would therefore benefit greatly from the introduction of a robust tool to automate, standardise and centralise much of this analysis.

Depending on the cause of the “official” Geno 2.0-based phylogenetic tree’s publication delay, the Genographic Consortium may also benefit from the availability of such automation.

In light of the above challenges, this task is more complicated than it would initially appear. This report showcases the results from an initial attempt to tackle this problem.

## Full-Y sequencing

Full Y-chromosomal sequencing will see growth over the next few years, soon supplanting chip-based wide-spectrum SNP testing as the test of choice for those seeking to extend the Y-chromosomal phylogenetic tree and determine their own position thereon. This technology will circumvent some of the chip-specific issues, but the corresponding increase in data will accelerate demand for automated analytical tools.

## 1.2 Prior Technology

Generally, the genetic genealogy volunteer community is quicker than the academic community to act on new phylogenetic discoveries. ISOGG’s Y-tree [8] can therefore be regarded as the most current phylogenetic authority. But as data volume increases, ISOGG’s volunteers will likely need help in order to keep pace with all the new developments. Many of these volunteers maintain separate, decentralised provisional trees for their own branches/haplogroups of interest. The level of detail varies between haplogroups.

The YCC phylogeny has not been updated since 2008 [9]. The Genographic Consortium’s awaited phylogeny is supposed to supersede this.

Thomas Krahn of Family Tree DNA maintains a draft Y tree [10]. It leads the ISOGG tree in some aspects, and lags behind in others. This too could probably benefit from increased automation.

FTDNA forum user Felix has just released *My Y-SNP Tree*, a free tool for comparing Y-SNPs against Thomas Krahn’s draft Y tree [11].

Ethio Helix has roughly classified some of the higher-level novel Geno 2.0 SNPs, using a non-fuzzy set theoretic approach [12].

Public raw Geno 2.0 data is scarce. Itaï Perez is analysing the Y-SNP facet of this raw Geno 2.0 data [13]. This will in particular be useful for external verification of FTDNA’s calling algorithm (applied during the Genographic-to-FTDNA transfer process).

The YFull service has recently classified 1000 Genomes project samples by terminal Y-SNP, using ISOGG tree 8.57 as the basis [14].

## 1.3 New Technology

Prior to preparing this report, the author developed a Y-SNP-based subclade predictor (now available at <http://ytree.MorleyDNA.com>), with the inputs being a person’s Y-SNP results. It suggests the person’s terminal subclade (or the closest parent subclade tested). This algorithm was able to correctly classify the Full Genomes Y-SNP demo results [15]. It also performed robustly when tested on Geno 2.0 data – either in the form of calls scored from a kit’s raw file, or in the form of calls reported in a public Family Tree DNA project after being transferred from the Genographic project.

The algorithm is most effective when using the ISOGG tree as a basis for classification.

Evidently, the Y-tree clade predictor performs better the more detailed the underlying phylogenetic tree is. And the refinement of a phylogenetic tree is greatly aided by the use of an accurate predictor and other analytical tools. The two pillars bootstrap each other.

To illustrate this, the author has used this Y-tree clade predictor (using the latest ISOGG tree as a basis for comparison) to classify over 1650 sets of publicly accessible Geno 2.0 Y-SNP calls. This information was then used as an input into another algorithm designed by the author – an algorithm developed to automate the construction of a phylogenetic Y-tree, while overcoming the challenges identified above. The technical details of this process will remain proprietary for the time being.

## 1.4 Data

ISOGG's latest Y-tree was used as a starting point. Off-tree SNPs identified by ISOGG as private or phylogenetically unreliable were not used as inputs. The algorithm used to develop these reports will include these SNPs in the output, if they show up in the available Geno data.

Over 1650 sets of Geno 2.0 positive calls (data transferred from the Genographic Project to Family Tree DNA) were used. These Geno 2.0 results were all collected from publicly accessible haplogroup and surname projects, including the haplogroup projects identified in the ISOGG wiki [16]. Coverage was skewed toward the most heavily-tested haplogroups.

User-provided earliest known ancestor information has been removed from this version of the phylogeny.

## 1.5 Results

The output of this automated tree-building process is presented in Section 2. See Table 1.1 for notation.

## 1.6 Discussion

Owing to the automated nature of the process, it is straightforward for the author to issue revised computer-generated draft phylogenies, as new information becomes available.

The draft phylogenies resulting from this process will not replace the existing ISOGG tree. But they should help reduce the workload of ISOGG project administrators. It will not make their roles obsolete. Their judgement is still needed to decide which SNPs are worthy of inclusion in the ISOGG tree. Moreover, the algorithm will identify some SNPs with inconsistent information, requiring manual examination in order to place on the tree. The reports developed by the author will serve as a *guideline* for further ISOGG tree expansion; ISOGG will retain editorial control over its Y-tree.

The Genographic Consortium phylogenetic tree – whenever it is published – should be superior to the tree presented here. The Genographic Consortium has a vastly larger and more diverse dataset to work with, and consequently one would expect the resultant paper to contain numerous branches not included here. Additionally, the Genographic Consortium, having access to all the raw data, will know which non-positive calls are negatives and which are in fact no-calls, thereby simplifying the tree construction process. The Genographic Consortium's paper will likely be treated to pre-publication peer review. This will delay the delivery but ensure the quality of the final result. It remains to be seen whether/how/how frequently the Genographic Consortium will update its phylogenetic tree after its initial publication.

Outstanding issues for the experimental phylogeny:

- The *PAGE#* class of SNPs needs to be equated with the *PAGES0000#* class of SNPs.
- The downstream SNP coverage statistics for known backmutations need to be corrected.
- Future versions of the phylogenetic report will split the tree into sections.
- Non-Latin letters (from the user-submitted earliest known ancestor descriptions) typeset as question marks.
- The author will release a separate automatically-generated report identifying discrepancies/gaps in the ISOGG data, suggesting potentially unreliable SNPs and listing SNPs whose positions need to be confirmed manually.
- See [this thread](#) for further issues.

## 1.7 Summary

There is present demand for tools that will automate the production and maintenance of a high-resolution Y-SNP phylogenetic tree. Demand will rise further. This report showcases the results from an initial attempt at tackling this problem.

## 1.8 References

- [1] Eran Elhaik, Elliott Greenspan, Sean Staats, Thomas Krahn, Chris Tyler-Smith, Yali Xue, et al., *The GenoChip: A New Tool for Genetic Anthropology*. *Genome Biol Evol*. 2013; 5(5): 1021-1031, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3673633/>, 9 May 2013 (accessed 15 July 2013).
- [2] CeCe Moore, *National Geographic and Family Tree DNA Announce Geno 2.0*. Your Genetic Genealogist, <http://www.yourgeneticgenealogist.com/2012/07/national-geographic-and-family-tree-dna.html>, July 25 2012 (accessed 5 July 2013).
- [3] Roberta Estes, *National Geographic – Geno 2.0 Announcement – The Human Story*. DNAeXplained – Genetic Genealogy, <http://dna-explained.com/2012/07/25/national-geographic-geno-2-0-announcement-the-human-story/>, July 25 2012 (accessed 5 July 2013).
- [4] Roberta Estes, *Geno 2.0 Results – First Peek*. DNAeXplained – Genetic Genealogy, <http://dna-explained.com/2012/12/11/geno-2-0-results-first-peek/>, 11 December 2012 (accessed 5 July 2013).
- [5] eng.Molgen.org community, *Geno 2.0 Y-phylogeny*. A Genetic Genealogy Community: Y-DNA, Mt-DNA, Autosomal DNA (Molgen.org English Forums), <http://eng.molgen.org/viewtopic.php?f=4&t=1080>, 24 April 2013 (accessed 5 July 2013).
- [6] Łukasz Łapiński et al., *R1a1a and Subclades Y-DNA Project – Y-DNA SNP*. Family Tree DNA Y-Haplogroup Project, <http://www.familytreedna.com/public/r1a/default.aspx?section=ysnp>, 15 July 2013 (accessed 15 July 2013).
- [7] Itaï Perez, *Re: [DNA] It's been a month, let's start complaining about Geno 2.0 transfer errors*. GENEALOGY-DNA-L Mailing List, <http://archiver.rootsweb.ancestry.com/th/read/GENEALOGY-DNA/2013-01/1358179571>, 14 January 2013 (accessed 5 July 2013).
- [8] International Society of Genetic Genealogy, *Y-DNA Haplogroup Tree 2013, Version: 8.64*. <http://www.isogg.org/tree/>, 19 July 2013 (accessed 25 July 2013).

Symbol Class	Symbol	Description
SNPs	M343	Positions known to ISOGG [8]. If on a branch with available Geno results: then the available Geno 2.0 results are consistent with ISOGG's positioning [8].
	CTS3368	SNPs unplaced in ISOGG's tree – placement has been proposed by the phylogenetic algorithm.
	DF27	SNPs not Geno-tested (or excluded by FTDNA from the list of positive SNPs). Names and placement come from ISOGG's tree [8].
	L440	Expected by ISOGG to be at this location, but the Geno 2.0 data suggests otherwise. When observed high up in the Y-tree: FTDNA may have the ancestral and derived states backwards, leading to positive results being scored negative, and vice versa.
	CTS2526 <sup>R</sup>	The SNP is recurrent – it has been also placed elsewhere in this tree. Recurrent does not necessarily mean erratic.
	PAGE65 <sup>b</sup>	ISOGG [8] has identified this instance as a back-mutation.
	CTS109 <sup>F</sup>	The novel SNPs upstream of haplogroup <b>R1</b> have been cross-checked against the Full Genomes demo data (comprised of one kit from <b>R1a</b> and one kit from <b>R1b</b> ). Novel SNPs with a superscript <sup>F</sup> did not test positive in one or both of the Full Genomes demo kits. It remains to be seen which discrepancies were caused by flaws in the Geno 2.0 data and which were caused by flaws in the Full Genomes data. Discrepancies are most abundant at the haplogroup <b>CT</b> level.
Kits	N114522	Blue kit numbers link back to FTDNA project SNP reports; follow the link to see the kit's positive calls obtained from Geno testing, and the positive and negative results obtained from à la carte SNP testing. Surnames and user-provided descriptions of the earliest-known patrilineal ancestor have not been included in this version of the report.
	*	A kit with this superscript may belong to one of the non-Geno-tested sister clades. À la carte SNP testing of the non-Geno-tested sister clade's (or clades') defining SNP(s) will be necessary to definitively place this kit.
Clades	<b>A1</b>	A clade featuring Geno-tested SNPs.
	<b>A0a</b>	A clade that does <i>not</i> feature Geno-tested SNPs.
	$-n$	Represents a subclade new to ISOGG's tree, proposed by the author's phylogenetic algorithm. For example, <b>R1b1-1</b> is proposed to be downstream of <b>R1b1</b> (and upstream of <b>R1b1a</b> ). Some of these new subclades, if terminal, may have already been deemed "private" by ISOGG.
	$\rightarrow^?$	This branch does not feature any Geno-tested kits, and consequently its position in the new phylogeny is unconfirmed.
	$\triangle$	The position of a subclade with this superscript cannot be fully resolved, because not all sister clades were Geno-tested, and none of the kits in this subclade have had the required non-Geno-tested SNPs tested à la carte.
	<b>A1b1</b> <sup>5</sup>	This clade has a footnote. Footnotes usually describe automatically detected discrepancies in the tree.
Coverage statistics	$[x/y]$	$y$ is the number of kits supposedly downstream of this SNP, and $x$ is the number of these kits that are positive for the SNP. Some non-positive results could in fact be no-calls or back-mutations. If this SNP is situated on a branch untouched by Geno 2.0 testing then the figure (which would be 0/0) is omitted. It is also omitted for SNPs that were not Geno-tested and have no positive-testing downstream kits.

Table 1.1: Notation used in the novel draft Y phylogeny

- [9] The Y Chromosome Consortium, *Ytree (YCC2008)*. <http://ytree.ftdna.com/index.php?name=YCC2008&parent=root>, 2008 (accessed 5 July 2013).
- [10] Family Tree DNA (Thomas Krahn *et. al.*), *Ytree (Draft Phylogeny)*. <http://ytree.ftdna.com/>, 2013 (accessed 5 July 2013).
- [11] Felix, *My Y-SNP Tree*. Genetic Genealogy Tools: Open source software for genetic genealogy, <http://www.y-str.org/tools/my-y-snp-tree/>, 3 July 2013 (accessed 5 July 2013).
- [12] Ethio Helix, *Geno 2.0 YDNA SNP Pathways*. Ethio Helix Blog, <http://ethiohelix.blogspot.com/2013/03/geno-20-ydna-snp-pathways.html>, 24 April 2013 (accessed 5 July 2013).
- [13] Itai Perez, *Geno 2.0 Y-Chromosome Genome Comparison: Extending Y-DNA Haplogroup Knowledge via Collaboration*. <http://itai.perez.free.fr/GenoCompare/>, 5 June 2013 (accessed 5 July 2013).
- [14] YFull Y-Chr Sequence Interpretation Service, *YFull 1000 Genomes (Tree)*. <http://www.yfull.com/tree/>, 12 July 2013 (accessed 15 July 2013).
- [15] Full Genomes Corp., *Demo for product: "Comprehensive Y-Chromosome Sequencing"*. <https://fullgenomes.com/demo>, 2013 (accessed 5 July 2013).
- [16] International Society of Genetic Genealogy, *Y-DNA haplogroup projects*. ISOGG Wiki, [http://www.isogg.org/wiki/Y-DNA\\_haplogroup\\_projects](http://www.isogg.org/wiki/Y-DNA_haplogroup_projects), 26 June 2013 (accessed 5 July 2013).

## 2. Experimental computer-generated Y-SNP phylogeny

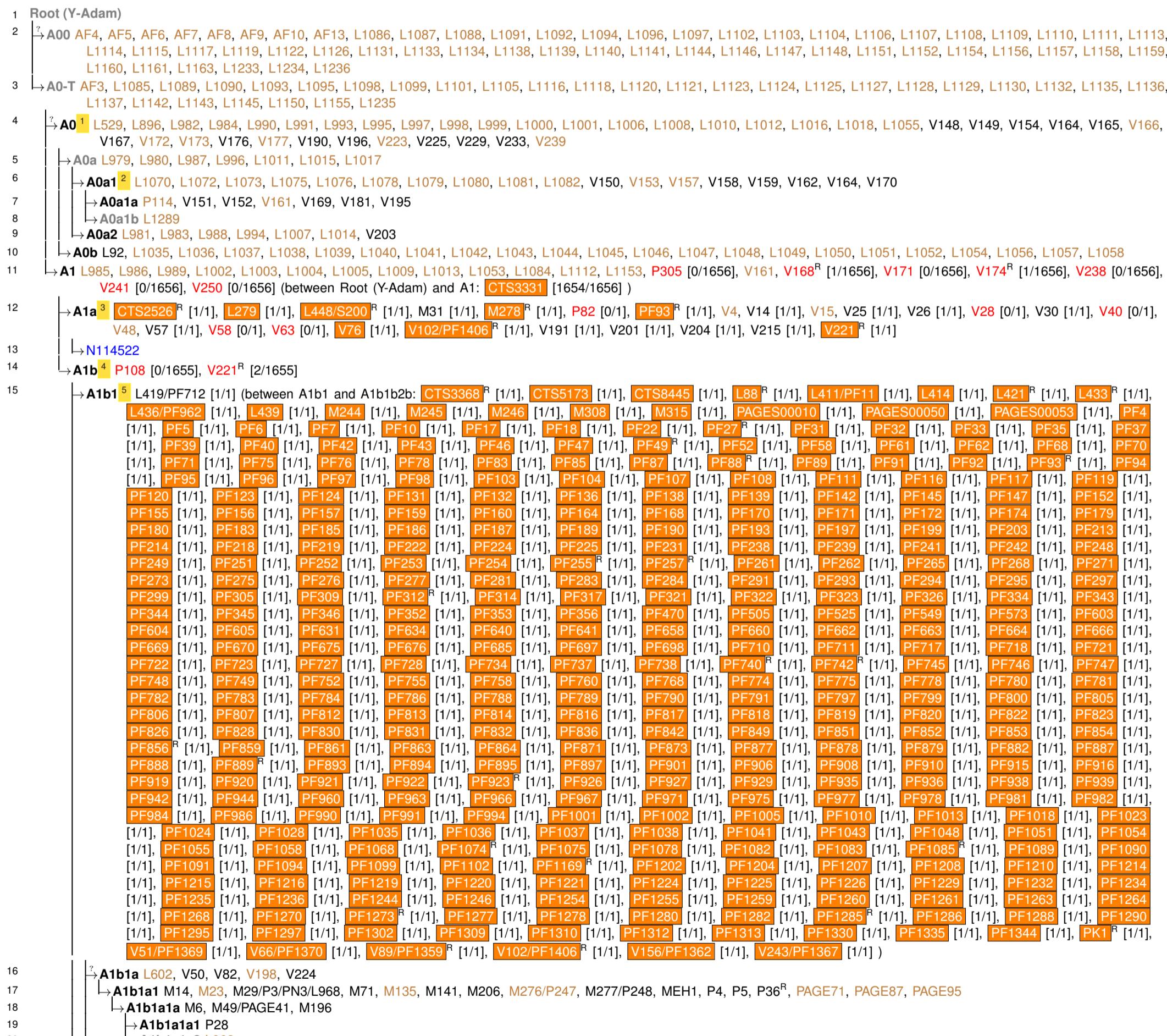
### 2.1 Experimental Y tree (automated)

#### Notice

This phylogeny uses ISOGG Y-DNA Haplogroup Tree version 8.64 (19 July 2013) [8] as a starting point.

This is an experimental tree. Please do not base major SNP-purchasing decisions on this, yet. It needs to be looked over by some experts to ensure that there are no systematic errors. ISOGG does not sponsor or endorse this experimental tree.

This tree is for personal, non-commercial use. It is to be distributed according to the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license, with the additional condition that it may not be used in Y-subclade prediction software without the author's prior consent.



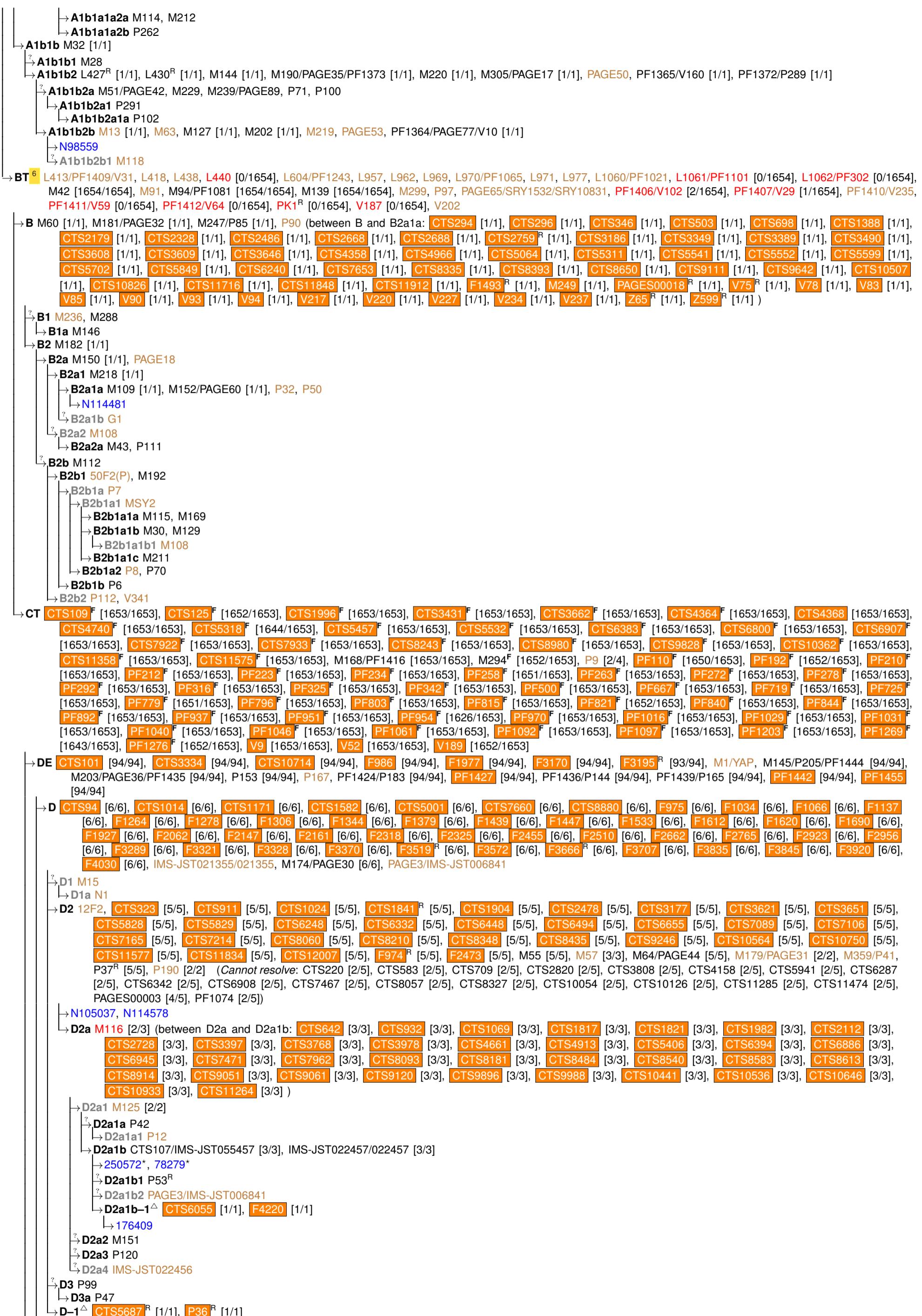
<sup>1</sup> V164 also found in nearby clade A0a1 (line 6 on page 6). One set of instances may be erroneous. Further investigation required.

<sup>2</sup> V164 also found in nearby clade A0 (line 4 on page 6). One set of instances may be erroneous. Further investigation required.

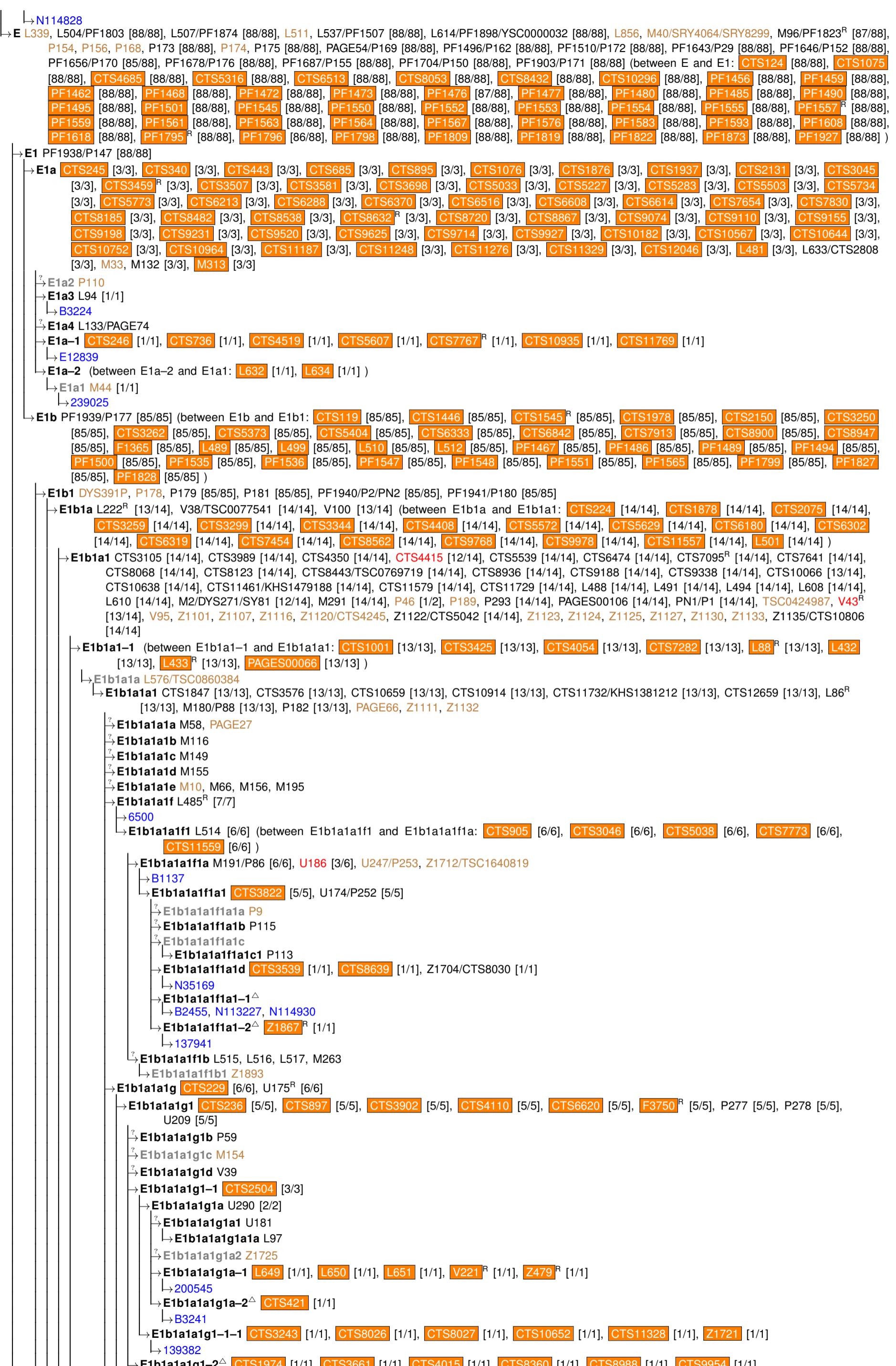
<sup>3</sup> PF93 and V102 also found in nearby clade A1b1 (line 15 on page 6). V221 also found in nearby clade A1b (line 14 on page 6). One set of instances may be erroneous. Further investigation required.

<sup>4</sup> V221 also found in nearby clade A1a (line 12 on page 6). One set of instances may be erroneous. Further investigation required.

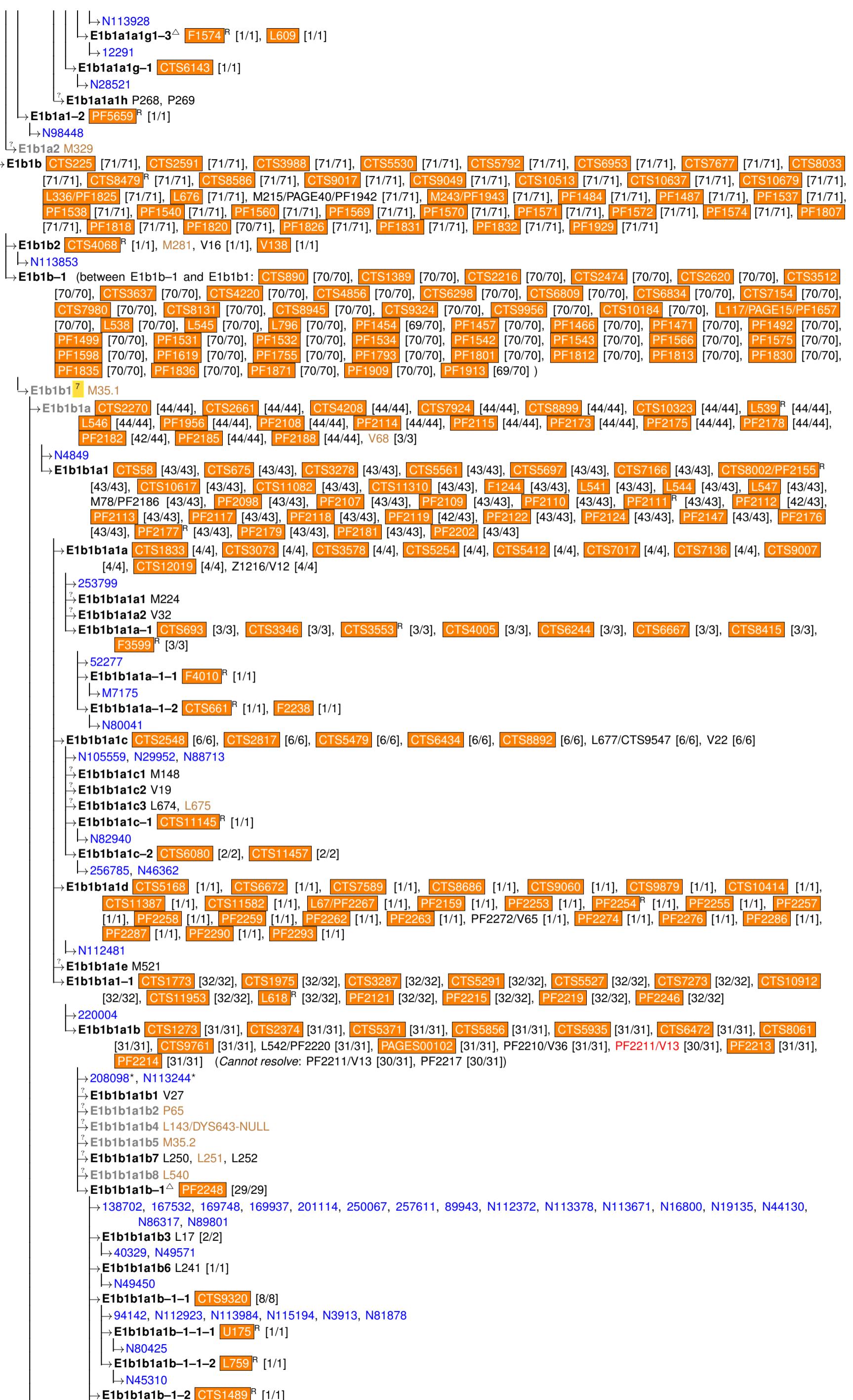
<sup>5</sup> PF93 and V102 also found in nearby clade A1a (line 12 on page 6). PK1 also found in nearby clade BT (line 32 on page 7). One set of instances may be erroneous. Further investigation required.



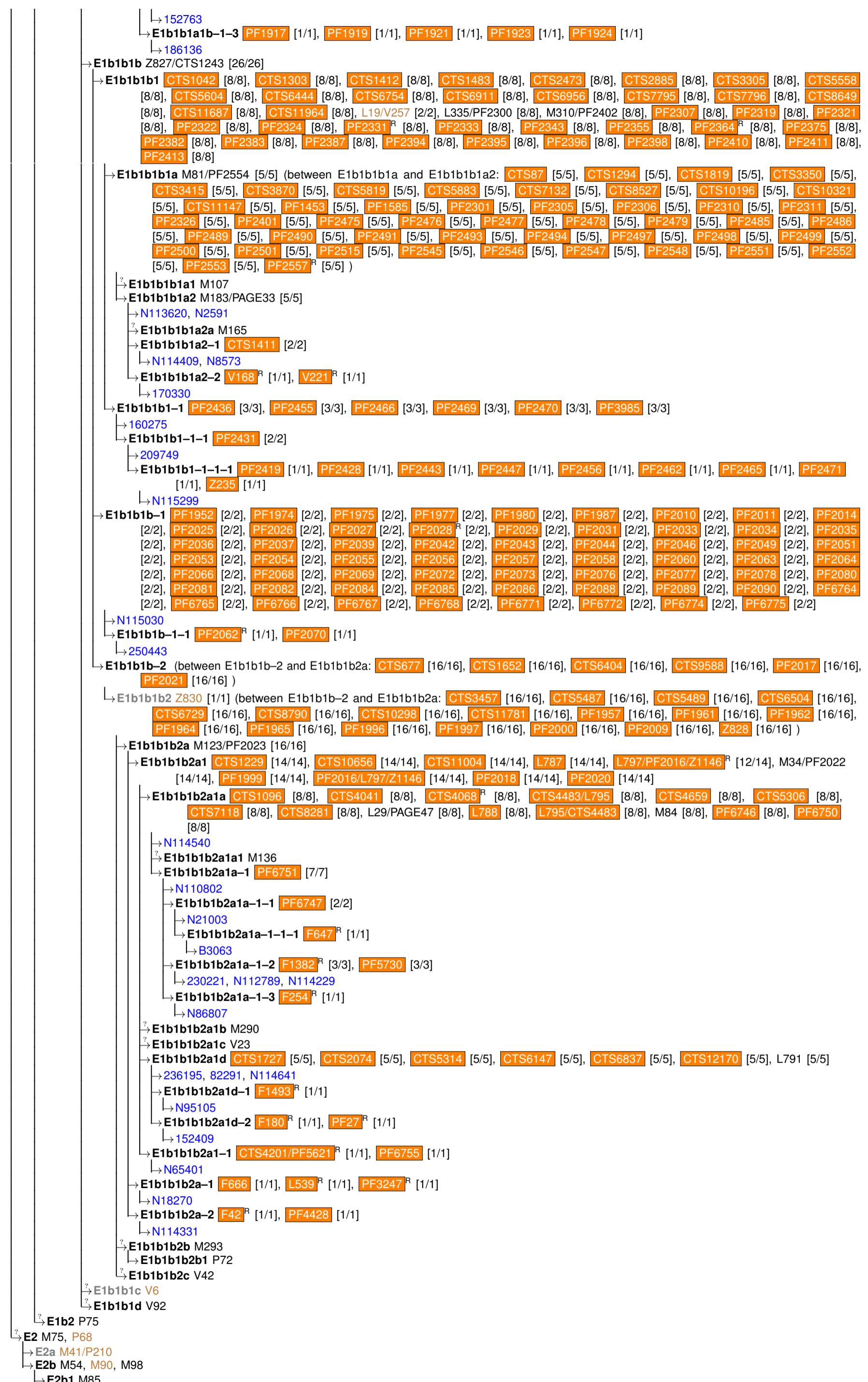
<sup>6</sup> PK1 also found in nearby clade A1b1 (line 15 on page 6). One set of instances may be erroneous. Further investigation required.

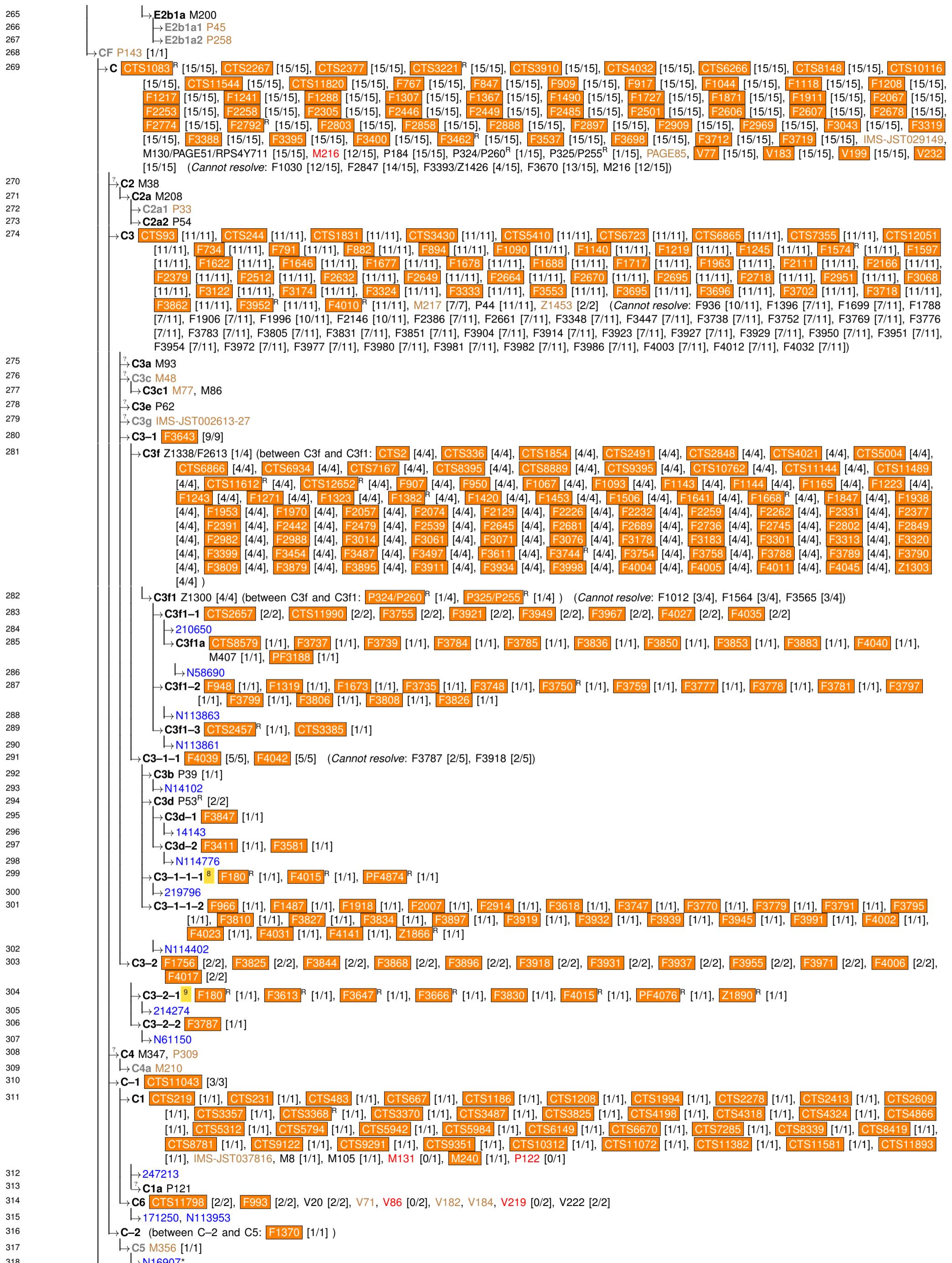


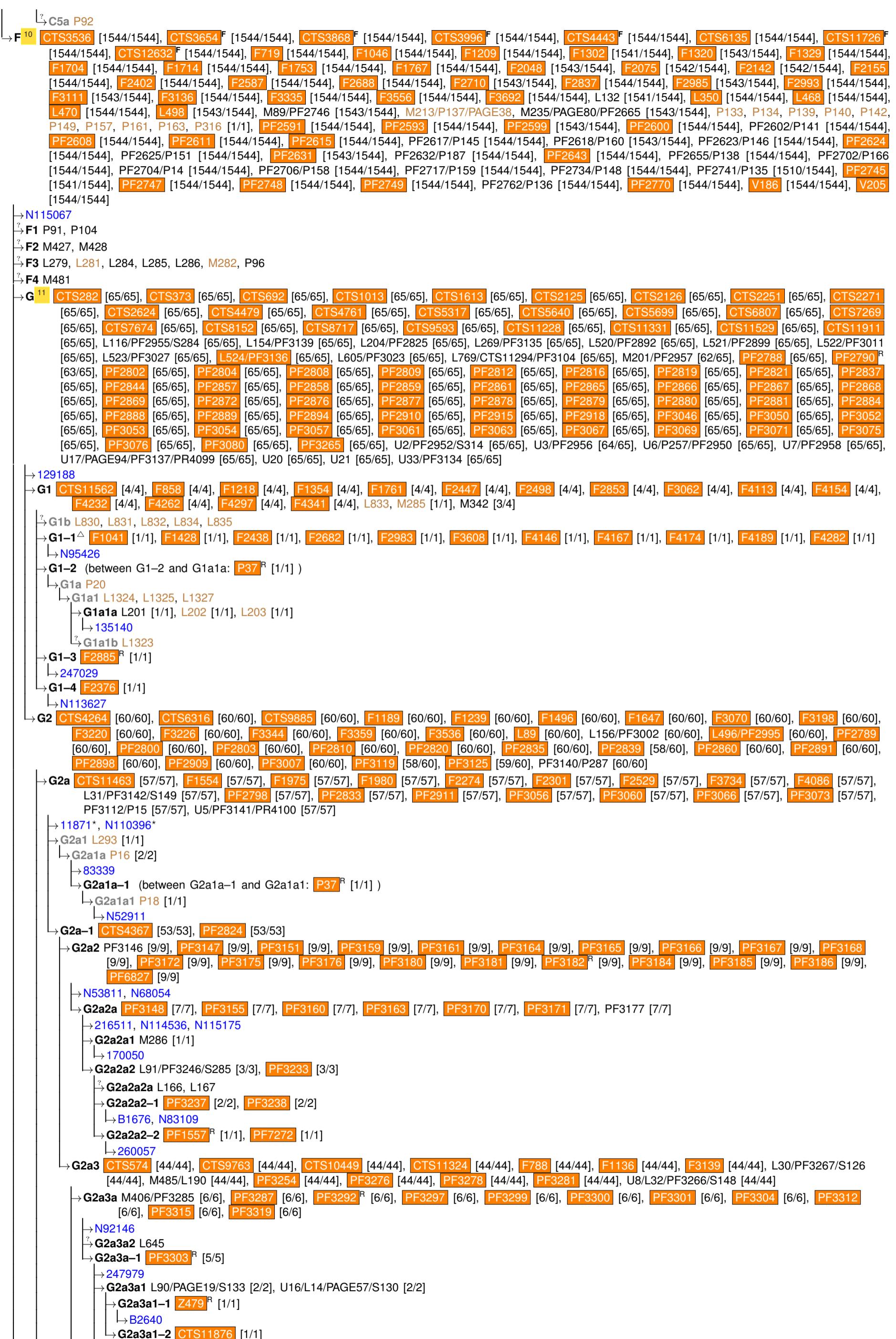
Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

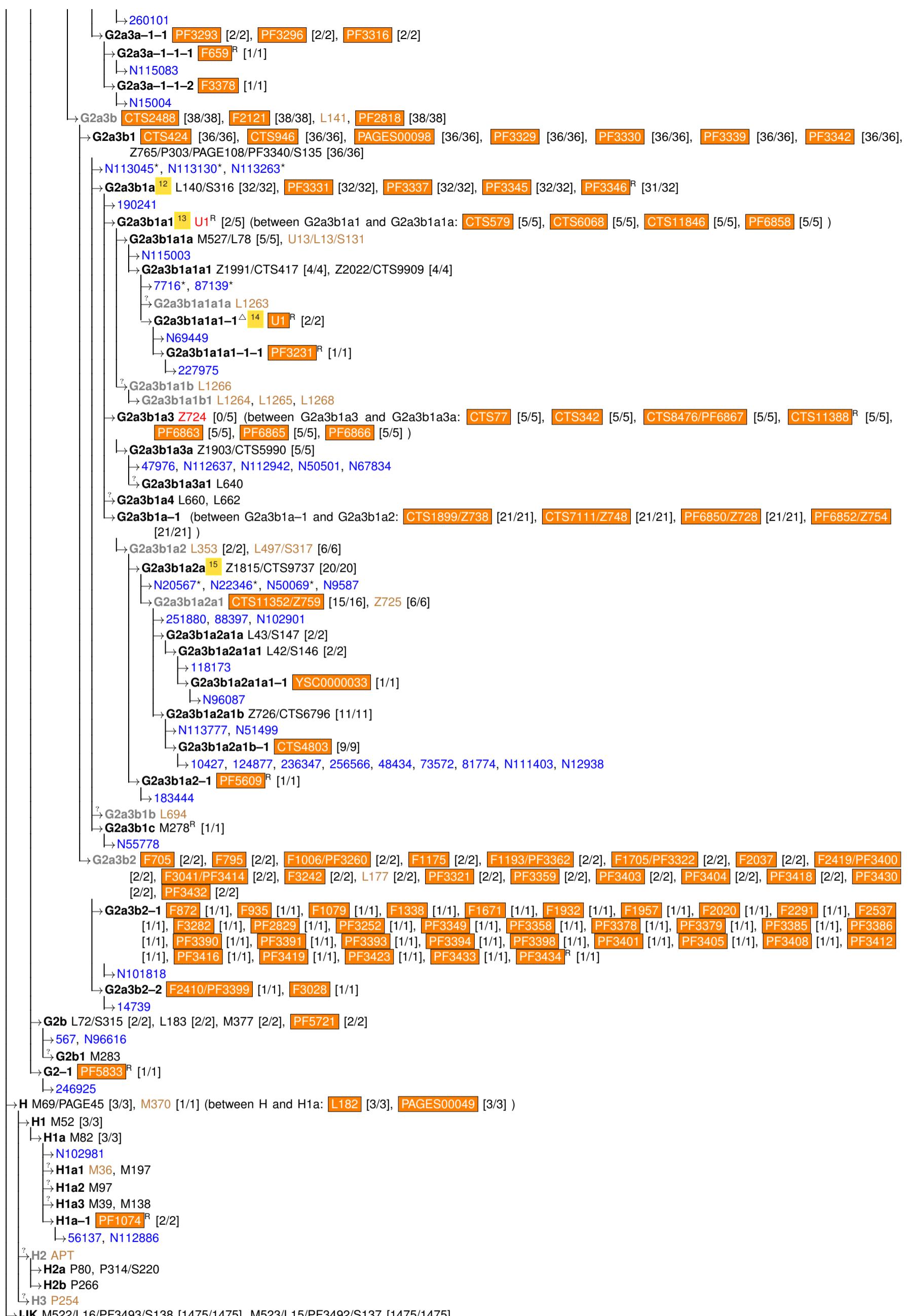


<sup>7</sup> An earlier version of this algorithm did not distinguish between M35.1 and M35.2, leading to errors in the phylogeny downstream of E1b1b1.

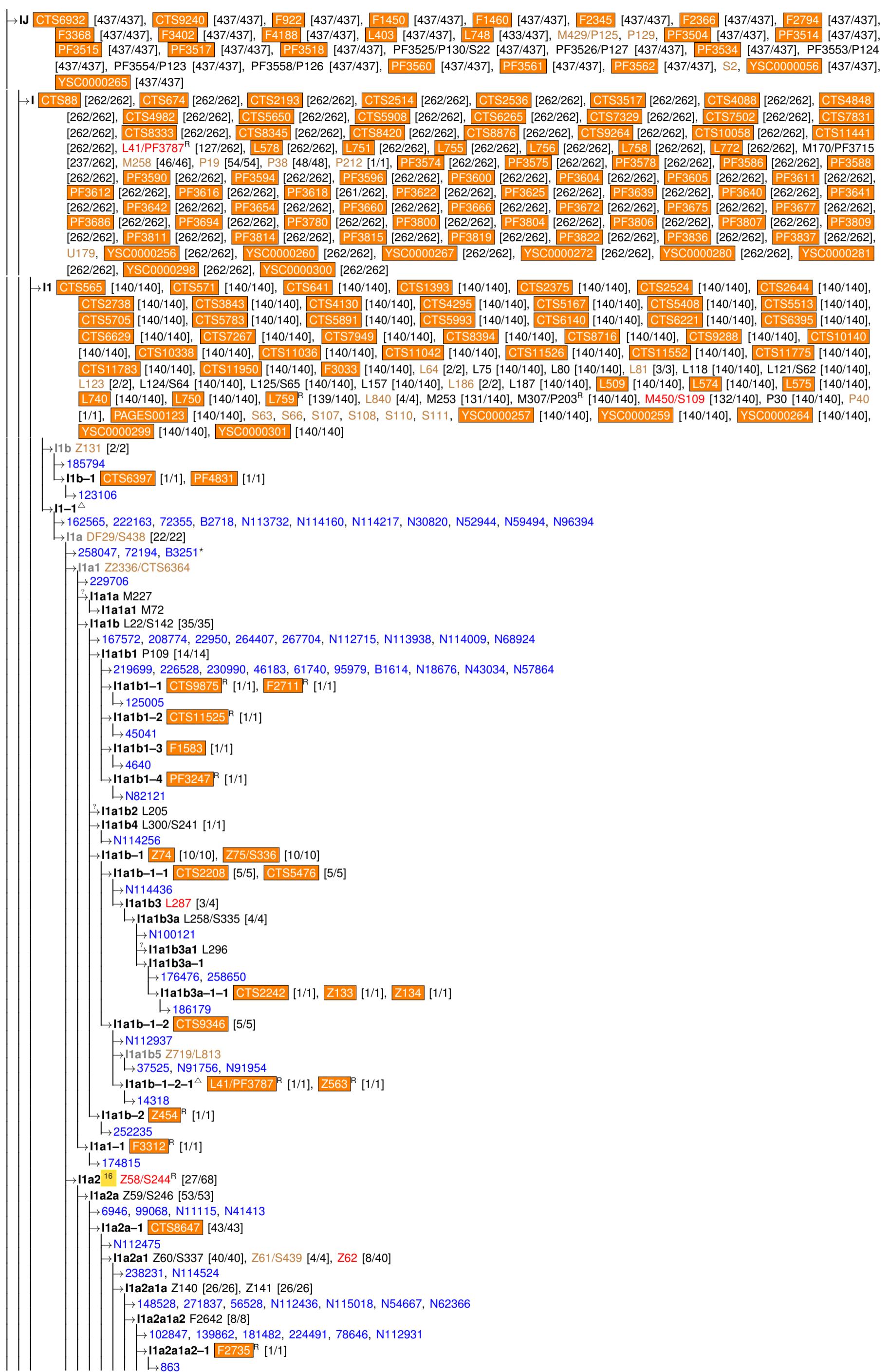
Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.<sup>8</sup> F180 and F4015 also found in nearby clade C3-2-1 (line 304 on page 11). One set of instances may be erroneous. Further investigation required.<sup>9</sup> F180 and F4015 also found in nearby clade C3-1-1-1 (line 299 on page 11). One set of instances may be erroneous. Further investigation required.

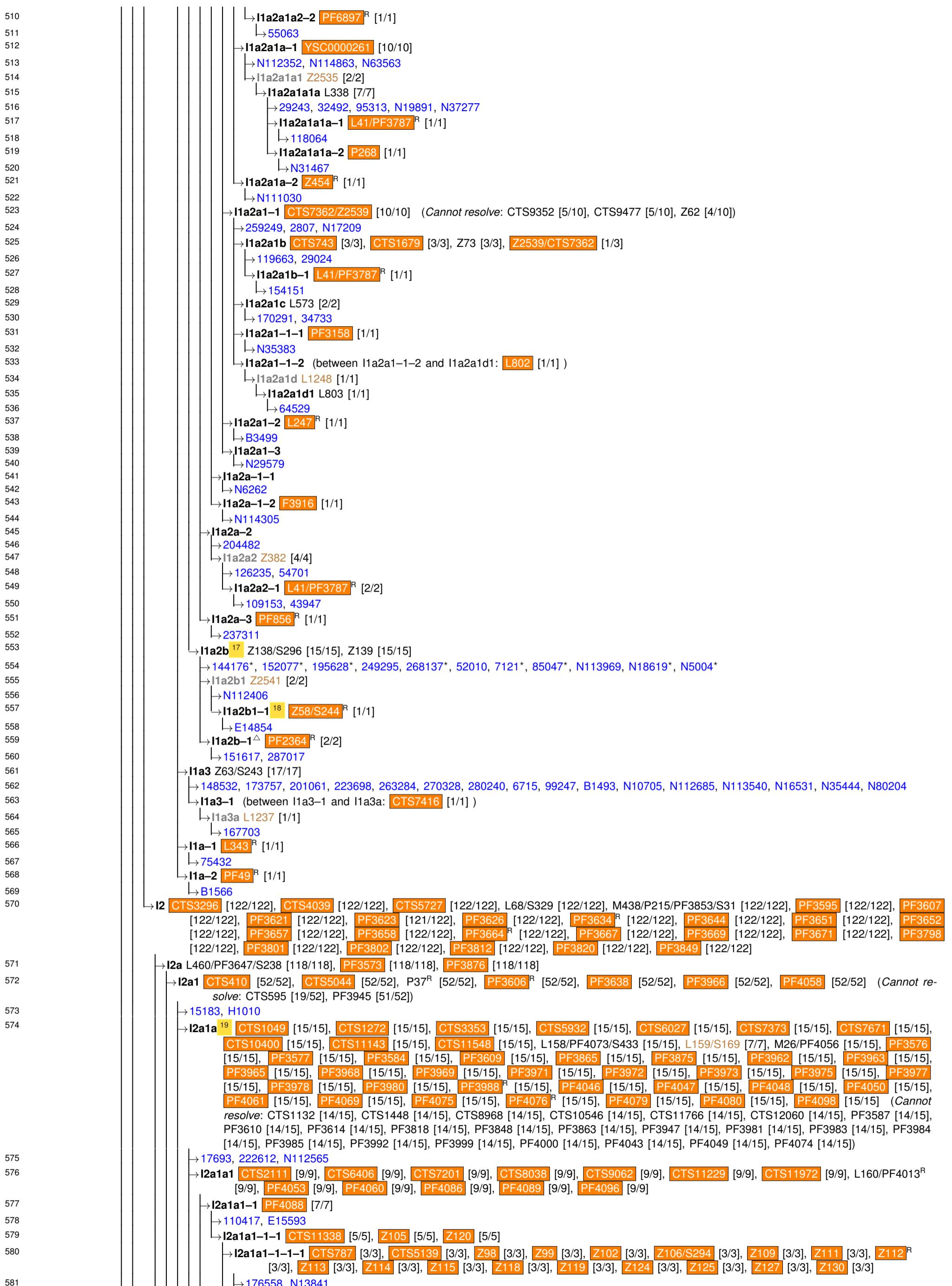
Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.<sup>10</sup> Kit N115067 has 1 positive SNP from the P level (line 1138 on page 24). Further investigation is required.<sup>11</sup> Kit 129188 has 15 SNPs proposed by this tree to be at the G1 level (line 328 on page 12), but this kit lacks positive results for any ISOGG-recognised SNPs at this level.

<sup>12</sup> Kit 190241 has 1 SNP proposed by this tree to be at the G2a3b1a3 level (line 397 on page 13), but this kit lacks positive results for any ISOGG-recognised SNPs at this level.<sup>13</sup> U1 also found in nearby clade G2a3b1a1a-1 (line 391 on page 13). One set of instances may be erroneous. Further investigation required.<sup>14</sup> U1 also found in nearby clade G2a3b1a1 (line 385 on page 13). One set of instances may be erroneous. Further investigation required.<sup>15</sup> These kits have one or more SNPs proposed to be at the G2a3b1a2a1 level (line 406 on page 13), but they all lack positive calls for any ISOGG-recognised SNPs at this level: N22346 (1), N50069 (1) and N9587 (1).

Version: 26th July 2013; see <http://ytree.MorleyDNA.com> for latest.

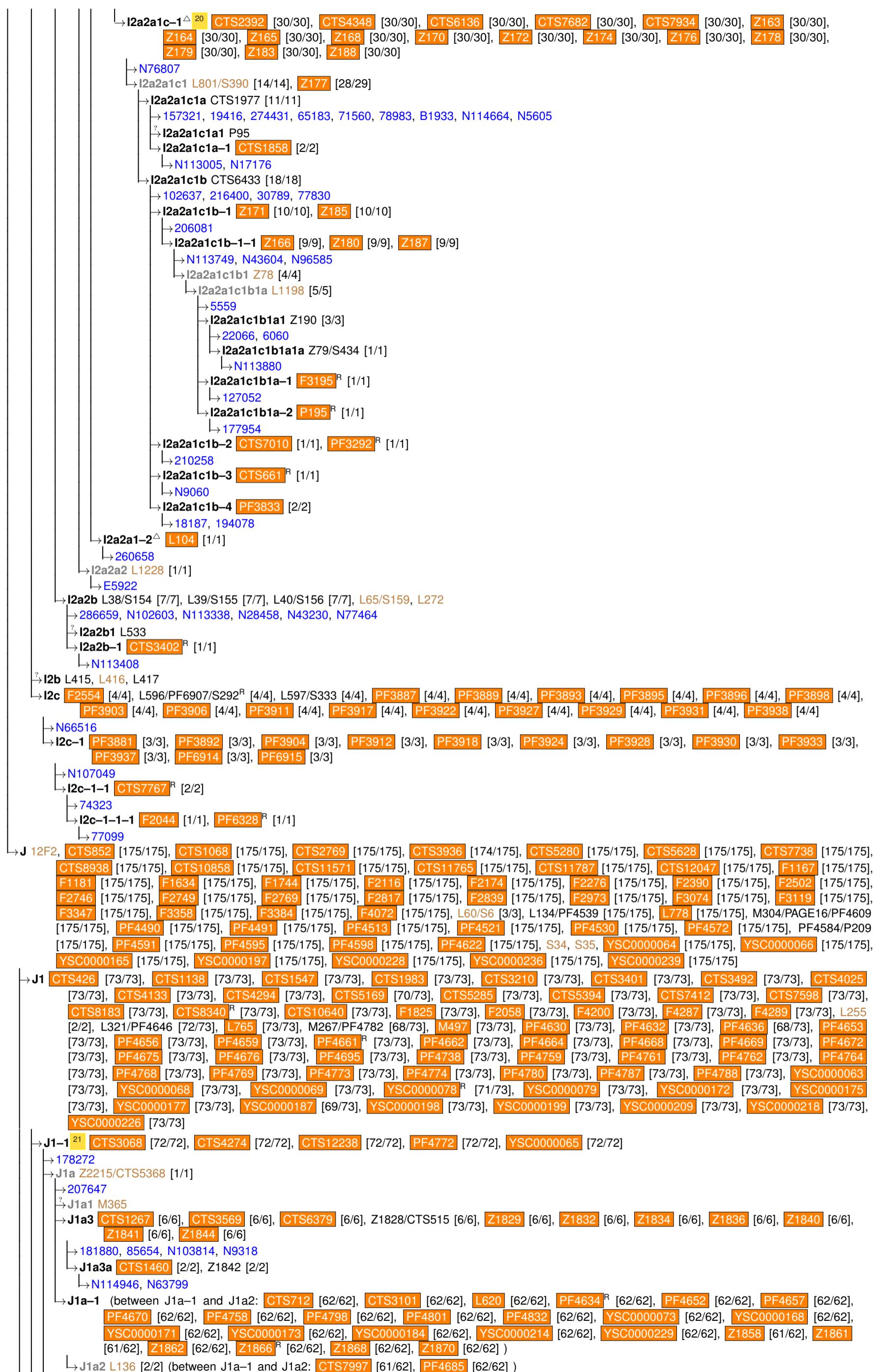


<sup>16</sup> Z58 also found in nearby clade I1a2b1-1 (line 557 on page 15). One set of instances may be erroneous. Further investigation required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.<sup>17</sup> These kits have one or more positive SNPs known or proposed to be at the I1a2b1-1 level (line 557 on page 15): 144176 (1), 7121 (1), 85047 (1) and N5004 (1). Further investigation is required.<sup>18</sup> Z58 also found in nearby clade I1a2 (line 497 on page 14). One set of instances may be erroneous. Further investigation required.<sup>19</sup> Results for M707 have been discarded.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.



Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.<sup>20</sup> Kit N76807 has 1 SNP proposed by this tree to be at the I2a2a1c1 level (line 665 on page 17), but this kit lacks positive results for any ISOOGG-recognised SNPs at this level.<sup>21</sup> Kit 178272 has 13 positive SNPs from the J1a-1 level (line 723 on page 17). Further investigation is required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.



<sup>22</sup> PF4678, Z1865, Z1885, Z1889 and Z1892 also found in nearby clade J1a2b-2 (line 754 on page 18). One set of instances may be erroneous. Further investigation required.

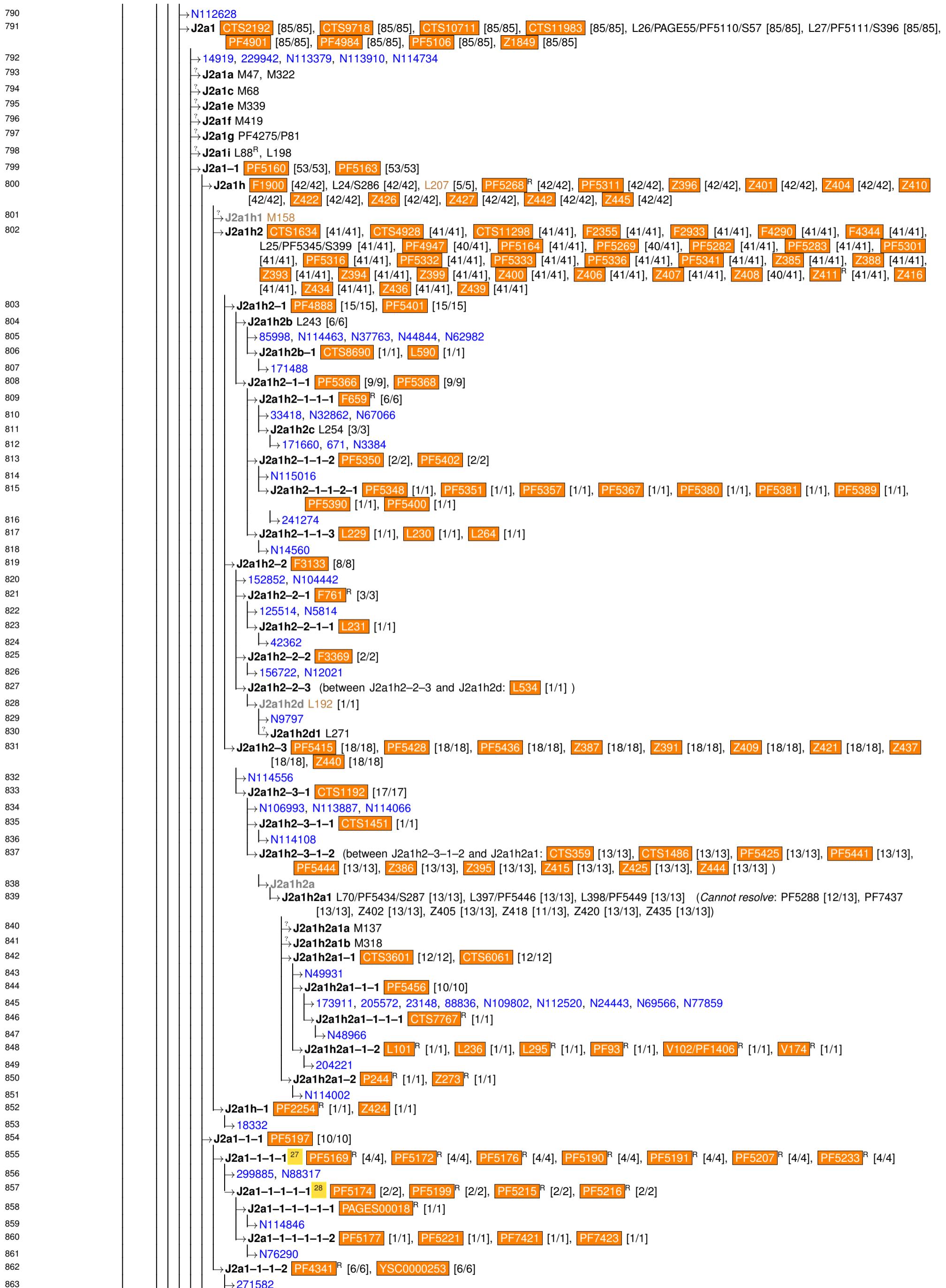
<sup>23</sup> CTS5269, PF4635, PF4838, YSC0000169, Z1853, Z1854, Z1855, Z1856, Z1867, Z1871, Z1875, Z1886 and Z1887 also found in nearby clade J1a2b-2 (line 754 on page 18). One set of instances may be erroneous. Further investigation required.

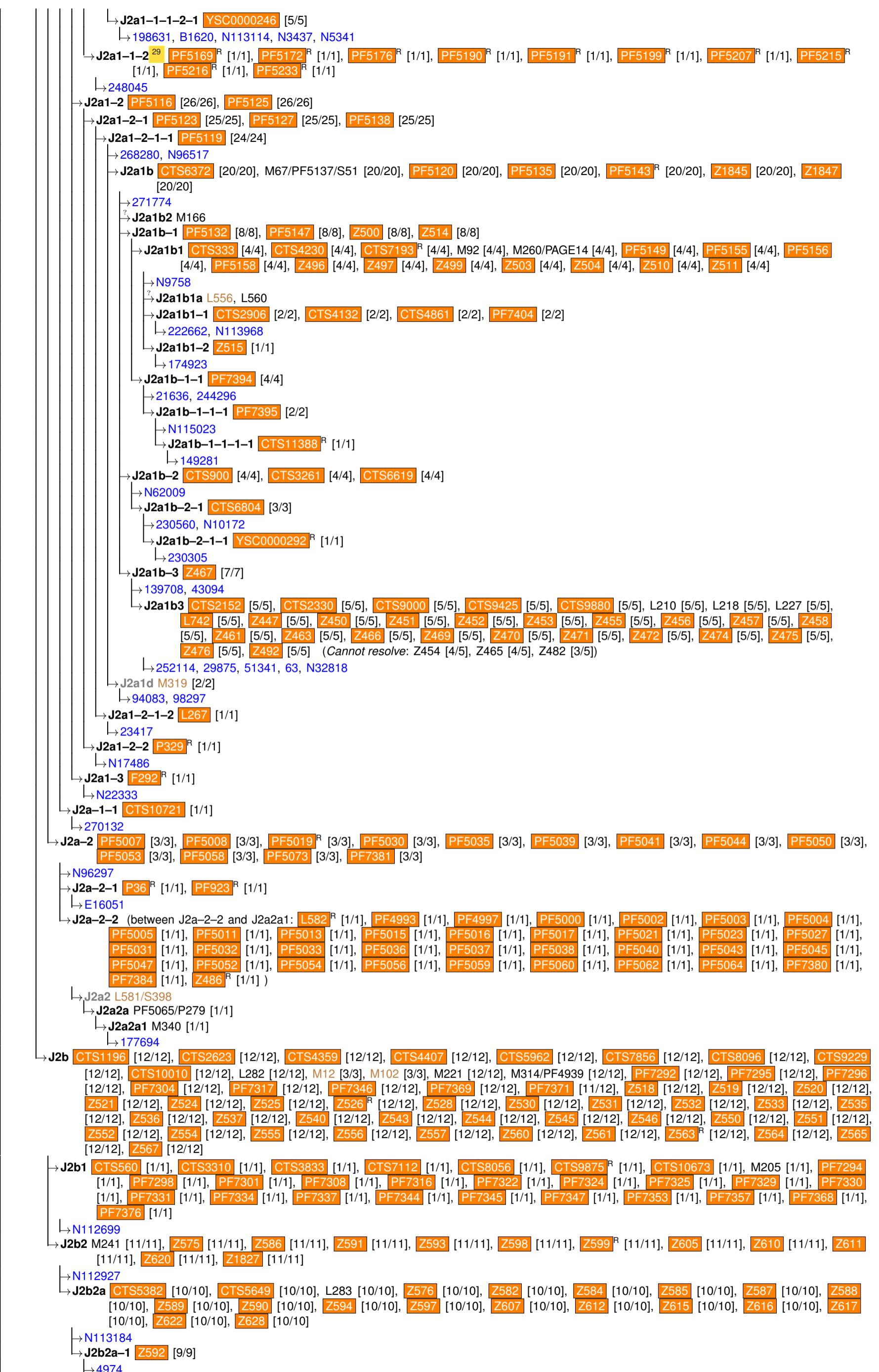
<sup>24</sup> PF4843 and Z1877 also found in nearby clade J1a2b1-1-1-2 (line 743 on page 18). One set of instances may be erroneous. Further investigation required.

**PF4843** and **Z1877** also found in nearby clade J1a2b-1-1-1-1-2 (line 745 on page 18). One set of instances may be erroneous. Further investigation required.

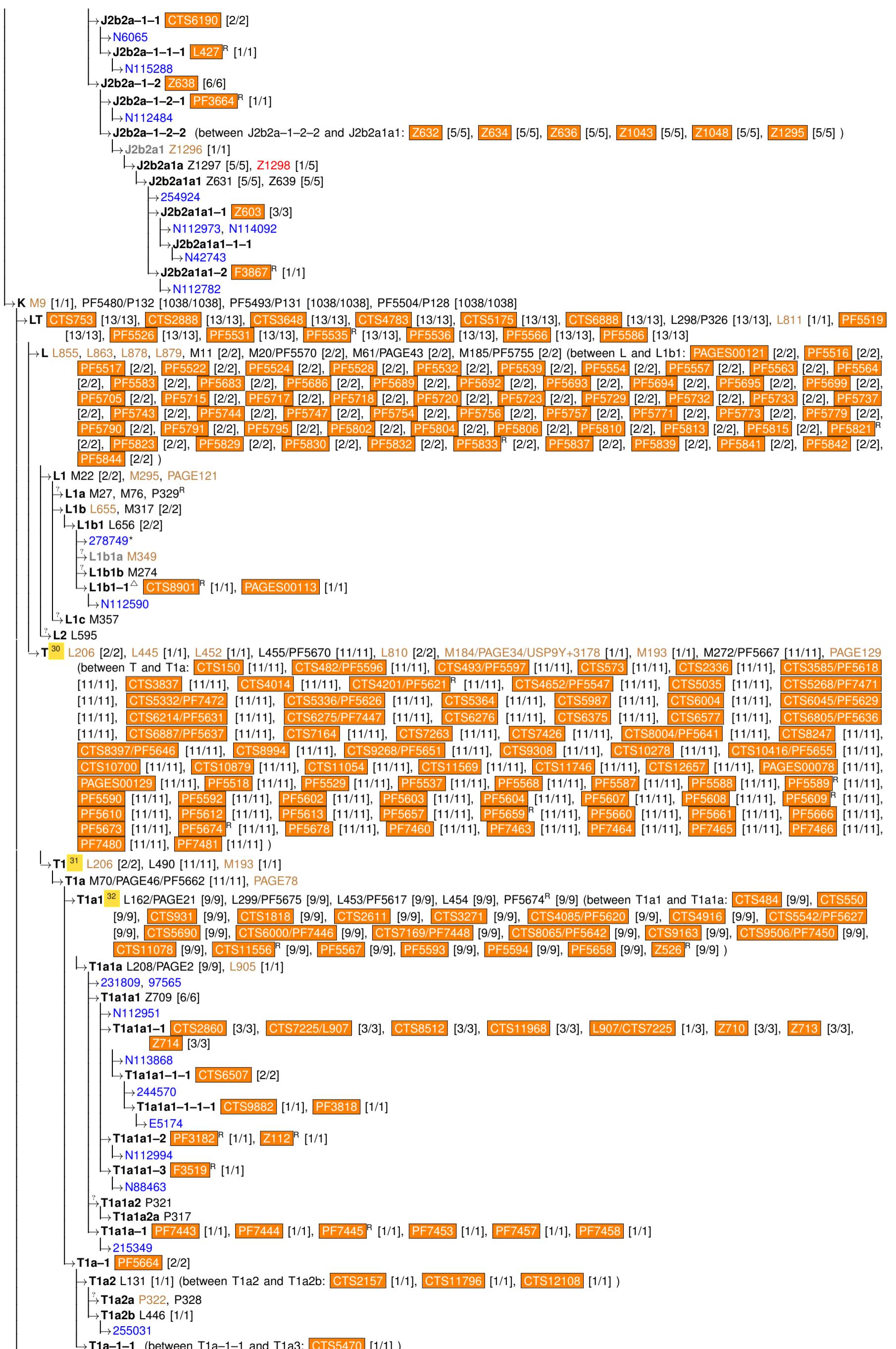
<sup>26</sup> CTSS5269, PF4635, PF4838, YSC00000071, YSC00000169, Z1853, Z1854, Z1855, Z1856, Z1867, Z1871, Z1875, Z1886 and Z1887 also found in nearby clade J1a2b1-1-1-1 (line 733 on page 18). One set of instances may be erroneous. Further investigation required.

PF4678, Z1865, Z1885, Z1889 and Z1892 also found in nearby clade J1a2b1-1 (line 731 on page 18). One set

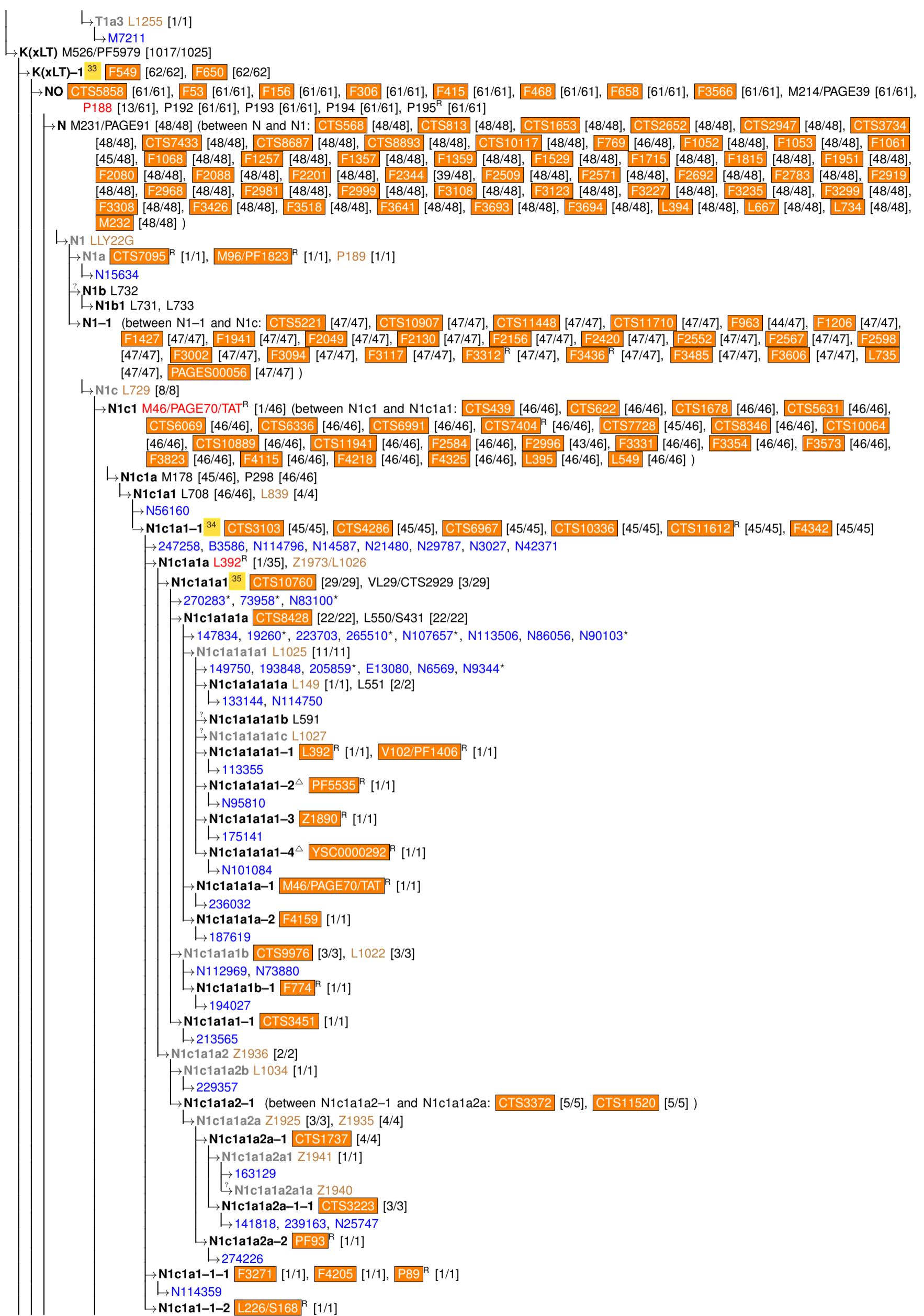
Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

<sup>29</sup> PF5169, PF5172, PF5176, PF5190, PF5191, PF5207 and PF5233 also found in nearby clade J2a1-1-1-1 (line 855 on page 19). PF5199, PF5215 and PF5216 also found in nearby clade J2a1-1-1-1-1 (line 857 on page 19). One set of instances may be erroneous. Further investigation required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.<sup>30</sup> L206 and M193 also found in nearby clade T1 (line 960 on page 21). PF5674 also found in nearby clade T1a1 (line 962 on page 21). One set of instances may be erroneous. Further investigation required.<sup>31</sup> L206 and M193 also found in nearby clade T (line 959 on page 21). One set of instances may be erroneous. Further investigation required.<sup>32</sup> PF5674 also found in nearby clade T (line 959 on page 21). One set of instances may be erroneous. Further investigation required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.



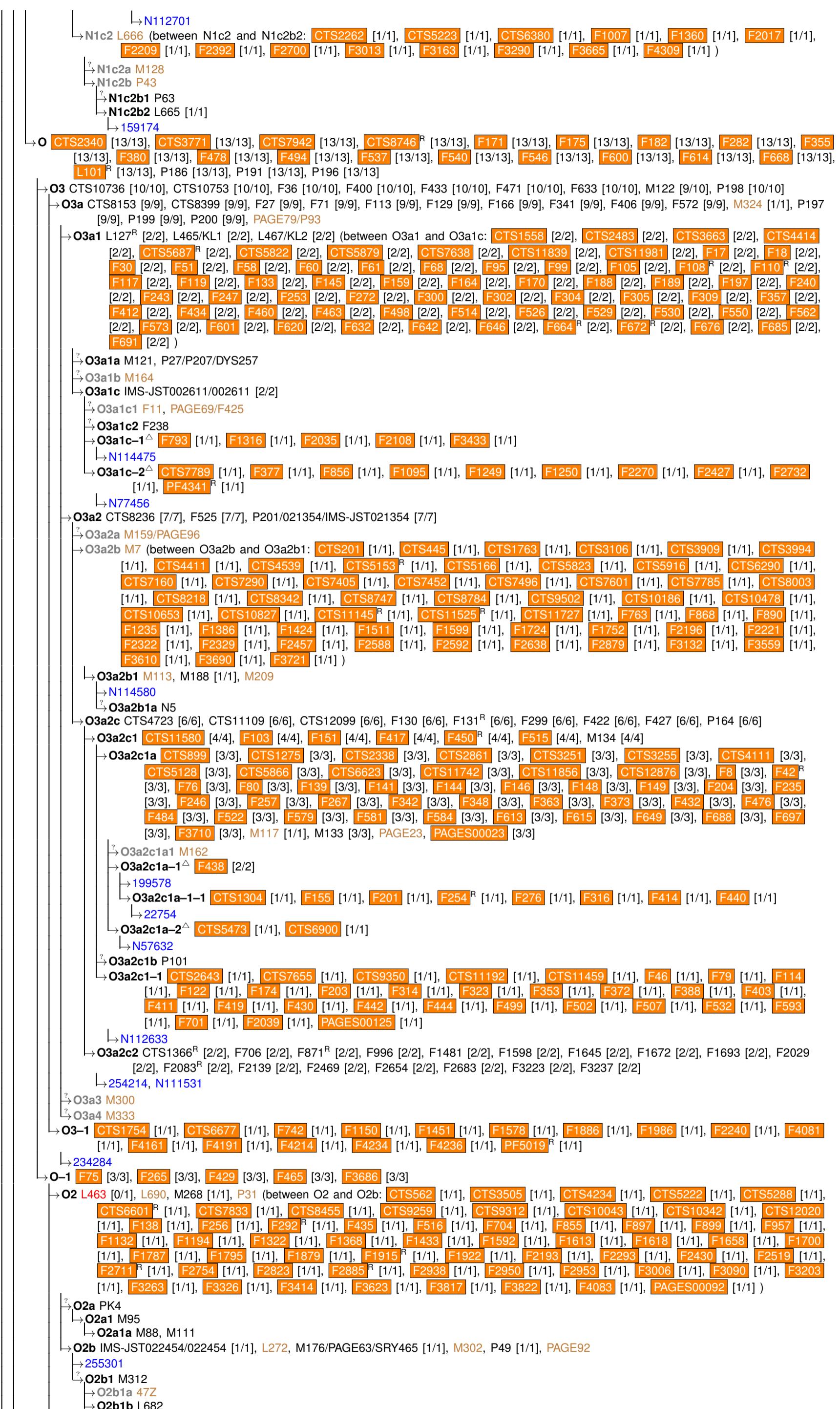
<sup>33</sup> If the classification of kit 213359 is correct, then this would suggest that haplogroup S is closer to macro-haplogroup NO than it is to macro-haplogroup P. Further testing is needed.

- to confirm kit 213359's membership in S-M226;
  - to confirm that haplogroup S is indeed derived for F549 and F650;
  - to determine the placement of haplogroups K1, K2, K3 and M relative to this potential novel macro-haplogroup

Positive findings will impact the haplogroup nomenclature system. The author favours the macro-haplogroup names "SON", "NOMS" (French for "names") and "MONKS" – the choice depending on the F549 and F650 statuses of haplogroups K1, K2, K3 and M – over alternatives such as "K(xLT)1" or "(K(xLT))(xM,P)".

<sup>34</sup> These kits have one or more positive SNPs known or proposed to be at the N1c1a1a2-1 level (line 1038 on page 22): 247258 (2), B3586 (2), N114796 (2), N14587 (2), N21480 (2), N29787 (2), N3027 (2) and N42371 (2). Further investigation is required.

<sup>35</sup> These kits have one or more SNPs proposed to be at the N1c1a1a1b level (line 1029 on page 22), but they all lack positive calls for any ISOGG-recognised SNPs at this level: 270283 (1), 73958 (1) and N83100 (1).

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.<sup>36</sup> Kit N9416 has 1 positive SNP from the R1a1a1b1a3a1a-1 level (line 1268 on page 26). Further investigation is required.

Version: 26th July 2013; see <http://Ytree.MoreleyDNA.com> for latest.



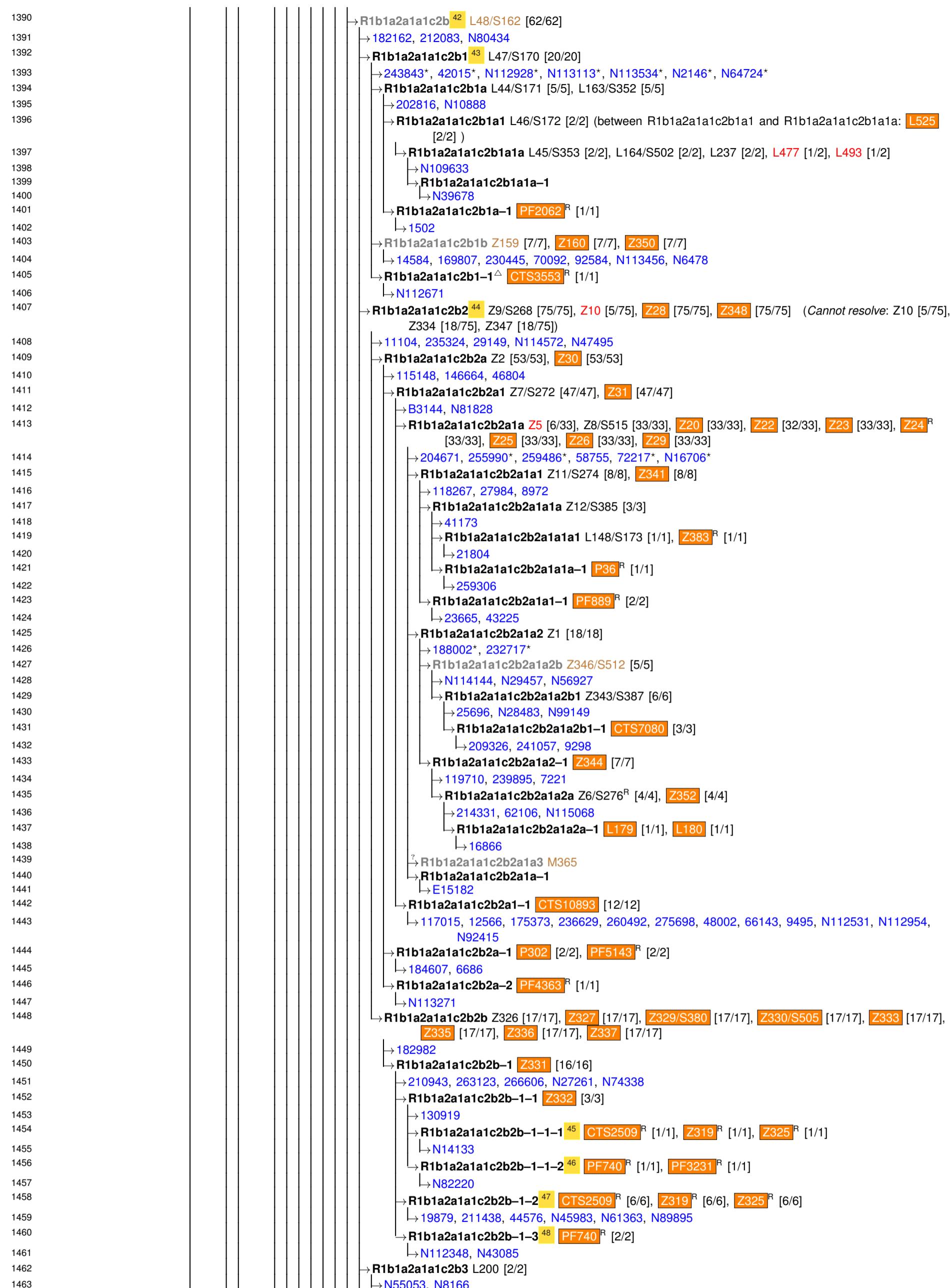
<sup>37</sup> These kits have one or more positive SNPs known or proposed to be at the R1b1a2a2c level (line 1858 on page 33): 26483 (1), N10795 (1), N112959 (1), N113118 (1), N114022 (1), N53795 (1), N79525 (1), N81217 (1) and N9475 (1). Further investigation is required. Kit N10795 has 1 positive SNP from the R1b1a2a2-2 level (line 1861 on page 33). Further investigation is required.

<sup>38</sup> These kits have one or more positive SNPs known or proposed to be at the R1b1a2a1a2-2 level (line 1831 on page 33): 16296 (1), 248663 (1), 249898 (1), N112805 (1) and N24160 (1). Further investigation is required. Kit 214080 has 1 positive SNP from the R1b1a2a1a1b1-1 level (line 1350 on page 27). Further investigation is required. Kit 222323 has 1 positive SNP from the R1b1a2a1a2a-2 level (line 1551 on page 30). Further investigation is required.

<sup>39</sup> Z19 also found in nearby clade R1b1a2a1a1b1a-2 (line 1348 on page 27). One set of instances may be erroneous. Further investigation required.

<sup>40</sup> These kits have one or more positive SNPs known or proposed to be at the R1b1a2a1a1b1a-2 level (line 1348 on page 27): N65375 (1) and N85364 (1). Further investigation is required.

<sup>41</sup> Z19 also found in nearby clade R1b1a2a1a1b (line 1337 on page 27). One set of instances may be erroneous. Further investigation required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

<sup>42</sup> Kit N80434 has 2 positive SNPs from the R1b1a2a1a1c2b1a1a level (line 1397 on page 28). Further investigation is required.

<sup>43</sup> These kits have one or more SNPs proposed to be at the R1b1a2a1a1c2b1b level (line 1403 on page 28), but they all lack positive calls for any ISOOGG-recognised SNPs at this level: 243843 (2), 42015 (2), N112928 (2), N113113 (2), N113534 (2), N2146 (2) and N64724 (2).

Kit N113534 has 2 positive SNPs from the R1b1a2a1a1c2b1a1a level (line 1397 on page 28). Further investigation is required.

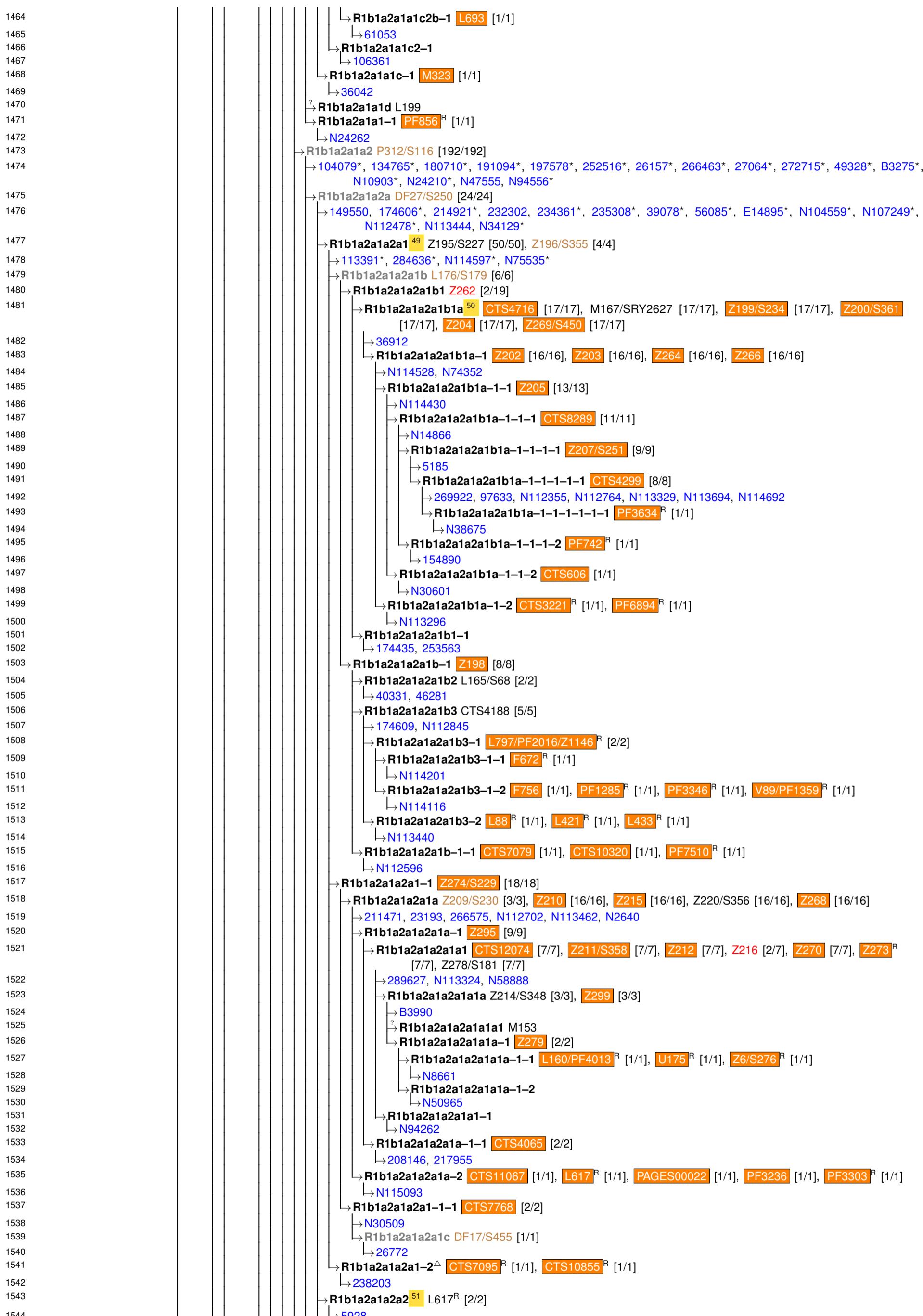
<sup>44</sup> Kit 29149 has 1 positive SNP from the R1b1a2a1a1c2b2b-1 level (line 1450 on page 28). Further investigation is required.

<sup>45</sup> CTS2509, Z319 and Z325 also found in nearby clade R1b1a2a1a1c2b2b-1-2 (line 1458 on page 28). One set of instances may be erroneous. Further investigation required.

<sup>46</sup> PF740 also found in nearby clade R1b1a2a1a1c2b2b-1-3 (line 1460 on page 28). One set of instances may be erroneous. Further investigation required.

<sup>47</sup> CTS2509, Z319 and Z325 also found in nearby clade R1b1a2a1a1c2b2b-1-1-1 (line 1454 on page 28). One set of instances may be erroneous. Further investigation required.

<sup>48</sup> PF740 also found in nearby clade R1b1a2a1a1c2b2b-1-1-2 (line 1456 on page 28). One set of instances may be erroneous. Further investigation required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest

<sup>49</sup> These kits have one or more positive SNPs known or proposed to be at the R1b1a2a1a2a1b-1 level (line 1503 on page 29): 113391 (1), 284636 (1), N114597 (1) and N75535 (1). Further investigation is required.

<sup>50</sup> Z201 results have been discarded. This marker seems to be prone to back-mutations.

<sup>51</sup> L617 also found in nearby clade R1b1a2a1a2b1-2 (line 1607 on page 30). One set of instances may be erroneous. Further investigation required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.<sup>52</sup> These kits have one or more positive SNPs known or proposed to be at the R1b1a2a1a2b3-1 level (line 1625 on page 30): B1735 (2) and N111980 (2). Further investigation is required.<sup>53</sup> L617 also found in nearby clade R1b1a2a1a2a2 (line 1543 on page 29). One set of instances may be erroneous. Further investigation required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.<sup>54</sup> Results for David F Reynolds' blacklisted SNPs downstream of R-L21 have been discarded.

These kits have one or more positive SNPs known or proposed to be at the R1b1a2a1a2c1-3 level (line 1765 on page 32): N113768 (1) and N113851 (1). Further investigation is required. Kit N114320 has 1 positive SNP from the R1b1a2a1a2c1-5 level (line 1769 on page 32). Further investigation is required. Kit N115037 has 1 positive SNP from the R1b1a2a1a2c1b level (line 1665 on page 31). Further investigation is required.

<sup>55</sup> Kit N112614 has 1 positive SNP from the R1b1a2a1a2c1a1a1-8 level (line 1663 on page 31). Further investigation is required.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.

Version: 26th July 2013; see <http://Ytree.MorleyDNA.com> for latest.



1874           └→ **R2** M479/PF6107 [7/7] (between R2 and R2a: **F761<sup>R</sup>** [7/7], **F802** [7/7], **F1018** [7/7], **F1087** [7/7], **F1115** [7/7], **F1281** [7/7], **F1376** [7/7], **F1684** [7/7], **F1766** [7/7], **F1768** [7/7], **F2342** [7/7], **F2585** [7/7], **F2810** [7/7], **F2979** [7/7], **F3005** [7/7], **F3029** [7/7], **F3118** [7/7], **F3171** [7/7], **F3265** [7/7], **F3302** [7/7], **F3379** [7/7], **F3435** [7/7], **F3484** [7/7], **F4235** [7/7], **L381** [7/7], **L722** [7/7], **L726** [7/7], **PF7490** [7/7], **PF7492** [7/7], **PF7493** [7/7], **PF7498** [7/7], **PF7500** [7/7], **PF7504** [7/7], **PF7510<sup>R</sup>** [7/7] )

1875           └→ **R2a** L266/PF6108 [7/7], M124 [7/7], P249 [7/7], PF6109/P267 [7/7]

1876            └→ **R2a1** L295<sup>R</sup> [1/1] (between R2a1 and R2a1b: **CTS10033** [1/1] )

1877            ?└→ **R2a1a** L294

1878            └→ **R2a1b** L723 [1/1]

1879            └→ 239090

1880           ?└→ **R2a2** L263

1881           └→ **R2a3** L1069 [1/1]

1882            └→ 130481

1883           └→ **R2a-1<sup>△</sup>** **F1092** [4/4], **F1159** [4/4], **F1758** [4/4], **F3604** [4/4]

1884            └→ 250837, 274421, N112631, N40042

1885           └→ **R2a-2<sup>△</sup>** **L288** [1/1]

1886            └→ N8823