

Mitochondrial molecular markers for US  
lineages of *P. infestans*

Brian J. Knaus, Javier F. Tabima and Niklaus J. Grünwald

February 21, 2014

## Contents

|          |                                               |           |
|----------|-----------------------------------------------|-----------|
| <b>1</b> | <b>The sample</b>                             | <b>3</b>  |
| <b>2</b> | <b>Variant discovery</b>                      | <b>4</b>  |
| 2.1      | Variant filtering . . . . .                   | 4         |
| <b>3</b> | <b>Variant segregation</b>                    | <b>8</b>  |
| 3.1      | Phylogenetic reconstruction . . . . .         | 8         |
| 3.2      | Mapping the SNP's in the BEAST tree . . . . . | 8         |
| <b>4</b> | <b>Session information</b>                    | <b>20</b> |

## List of Figures

|   |                                                                                                       |   |
|---|-------------------------------------------------------------------------------------------------------|---|
| 1 | Quality control results for the mtDNA SNP calls before filtering                                      | 5 |
| 2 | Quality control results for the mtDNA SNP calls after filtering<br>and windowizing variants . . . . . | 6 |
| 3 | Whole mtDNA genome scan for the P . . . . .                                                           | 7 |
| 4 | BEAST Coalescent tree . . . . .                                                                       | 9 |

## List of Tables

|   |                                                                |    |
|---|----------------------------------------------------------------|----|
| 1 | Variants remaining after filtering. . . . .                    | 22 |
| 2 | Diagnostic SNP positions for the mtDNA genome after filtering. | 23 |

# 1 The sample

2 The sample includes data which was opportunistically gathered from previous  
3 publications as well as data which is not yet available to the public (Judelson,  
4 unpublished).

| Sample           | Count | Reference |
|------------------|-------|-----------|
| T30-4            | 1     | [4]       |
| PIC99189 & 90128 | 2     | [8]       |
| 13_a2            | 1     | [2]       |
| Yoshida et al.   | 13    | [9]       |
| Martin et al.    | 3     | [7]       |
| Judelson         | 8     | NA        |
| Total            | 28    |           |

- 6 • The sample ‘T30-4’ was the first sequenced genome and is considered the  
7 reference for nuclear work [4]. This genome was assembled prior to high-  
8 throughput sequencing (i.e., Illumina and 454 technologies). The data  
9 presented here are not the sequences used for the paper but are part of  
10 a project by The Broad to resequence this individual using Illumina and  
11 454 technologies.
- 12 • Note that both the Yoshida and Martin papers included ancient DNA in  
13 their analyses [9, 7]. Here we have omitted those samples and focused on  
14 modern samples.
- 15 • For enigmatic reasons, not all of the samples from the Yoshida and Martin  
16 papers were actually available online. Therefore our numbers here do not  
17 match those presented in the papers.
- 18 • The Judelson data include a sample of US1 which was sampled at three  
19 different time points (us1\_1, us1\_2 and us1\_3). We suspect that these  
20 were different samples and not necessarily the same clone. Therefore dif-  
21 ferences among these samples may either be due to biological or technical  
22 factors.
- 23 • The Judelson data includes a sample of US8 which has been characterized  
24 as having fungicide resistance [3]. This lineage was also sequenced by  
25 Martin et al. [7]. These are most likely different samples so differences  
26 among these samples may be interpreted as biological.
- 27 • The Yoshida data includes one sample of *P. mirabilis* (p7722), this should  
28 jump out in the analyses.
- 29 • For the mitochondrial data we used the type IIa form [1] because it was  
30 the longest sequence and we felt this would provide the best alignment.

31 We've used the term 'SNP' fairly loosely in this document. The term 'vari-  
32 ant' may be more appropriate. Until fairly recently the software tools we've  
33 been using could only handle SNPs. They now report short indels as well.  
34 We've included both variant types here.

## 35 2 Variant discovery

36 Reads were mapped to the type IIa mitochondrial reference "[AY898627.1](#)".  
37 Reads were mapped using bowtie2 [5]. Variants were called using SAMtools[6].

### 38 2.1 Variant filtering

39 As a quality control step, the variant files were filtered by quality, read depth  
40 and mapping quality (Figures 1,2). For this we used an in-house R package  
41 called [vcfR](#). Here, sequencing depth is cumulative over all samples. Quality  
42 here is for each variant over all samples and ranges from 1-999.

43 The genotype caller in Samtools assumes a diploid, bi-allelic model. Because  
44 mitochondria are assumed to be haploid we tried to filter out heterozygous  
45 calls. Samples which included high quality heterozygote calls (p1362, p6096,  
46 p10650, p12204, p10127) were mostly from the Yoshida et al. [9] paper and  
47 were among the low sequencing depth samples they included. Because these  
48 samples are not among the US lineages we're interested in, and because they  
49 are apparently of low sequencing depth, we omitted them for now. However, the  
50 sample nl07434 was among the high sequencing depth samples from this paper  
51 and is perhaps noteworthy. T30-4 was called as a heterozygote for one variant  
52 and is perplexing.

53 In an attempt to identify high quality variants we employed a filtering strategy.  
54 Filtering of the variant panel was based on quality (QUAL=999), cumulative  
55 sequencing depth (1st quartile >= DP >= 3rd quartile) and mapping quality  
56 (1st quartile >= MQ >= 3rd quartile). This resulted in 37 variants remaining  
57 after filtering (Table 1). We have identified a fraction of these as being diagnostic  
58 for a small group of samples (Table 2).

59 The variants remaining after filtering were visualized as a linear chromosome  
60 in Figure 3.

```
## gt.m2sfs is commented out  
## Before filtering:  
## [1] 247  
## After filtering:  
## [1] 37
```

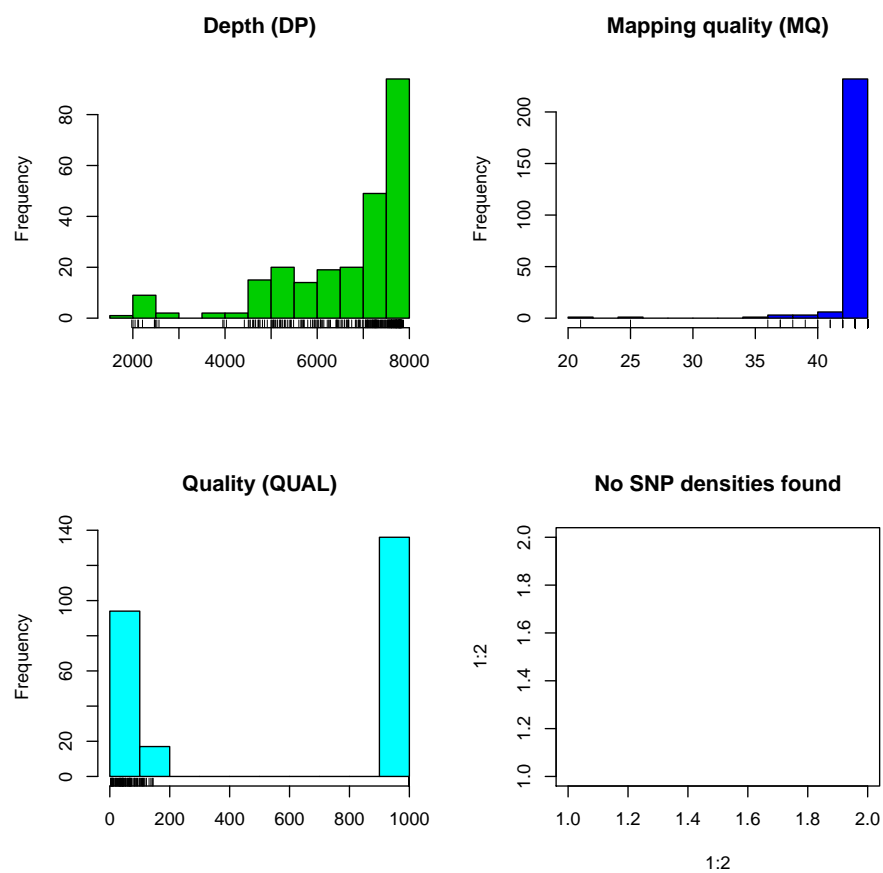


Figure 1: Quality control results for the mtDNA SNP calls before filtering.

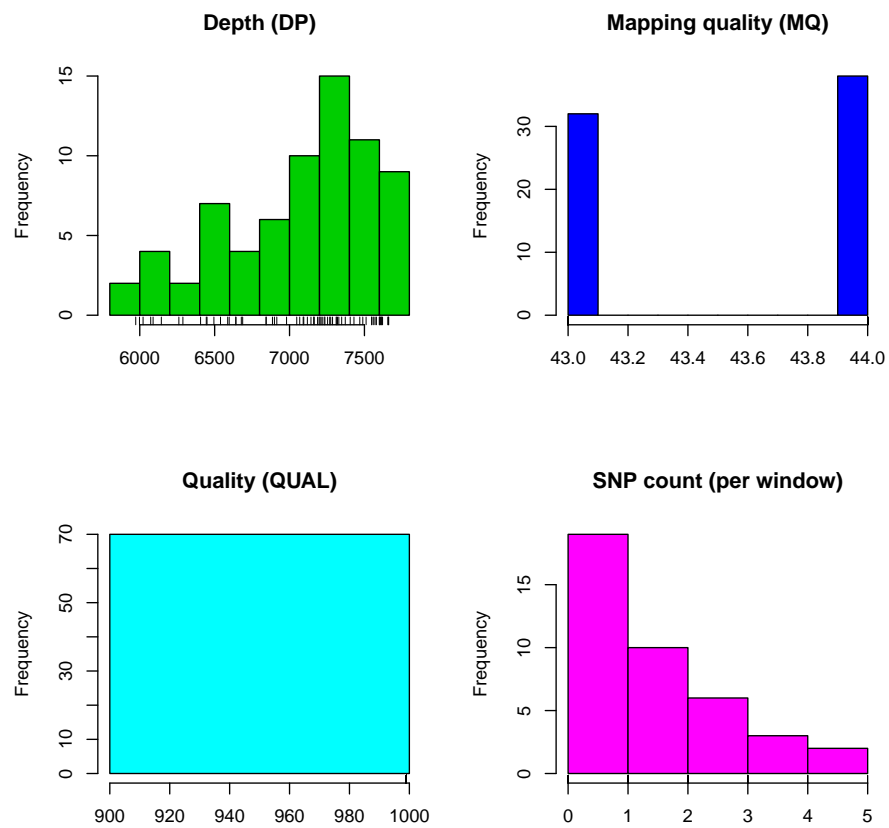


Figure 2: Quality control results for the mtDNA SNP calls after filtering and windowizing variants.

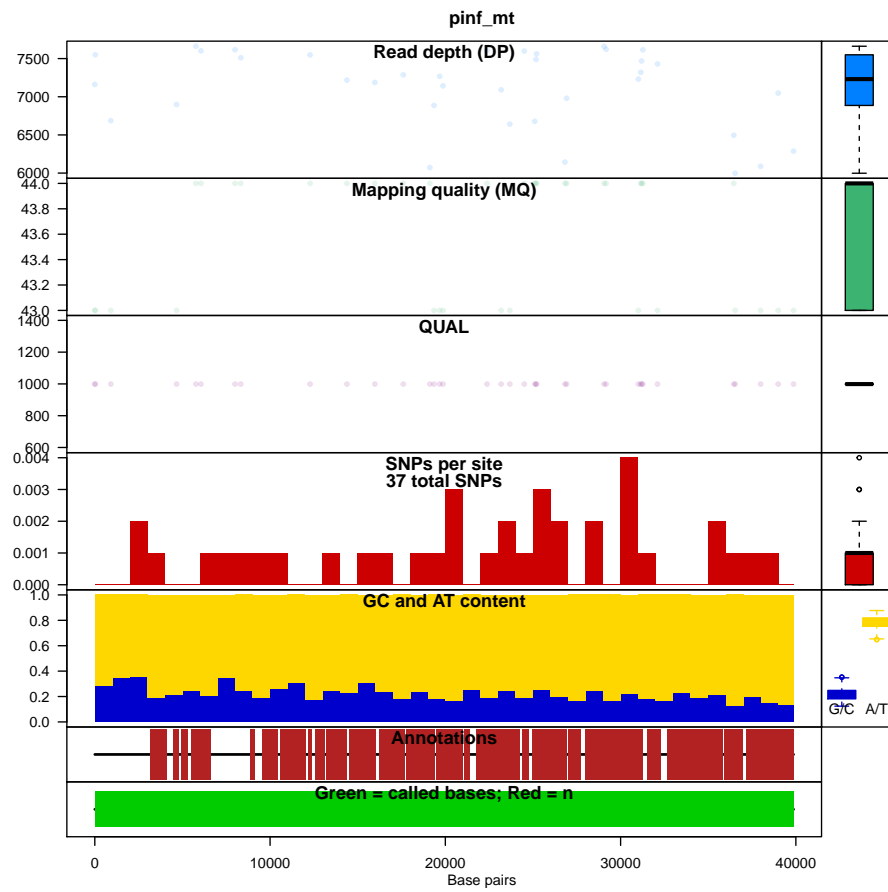


Figure 3: Whole mtDNA genome scan for the *P. infestans* samples.

## 61 3 Variant segregation

62 In order to visualize how variants segregated among the samples, a phylogeny  
63 was inferred. We then used ancestral state reconstruction to map the characters  
64 to the tree. At this time we're not trying to say anything bold about phylogeny  
65 or character evolution. We're simply using these tools to visualize how the  
66 variants segregate.

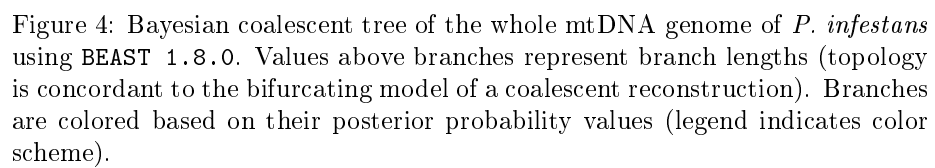
### 67 3.1 Phylogenetic reconstruction

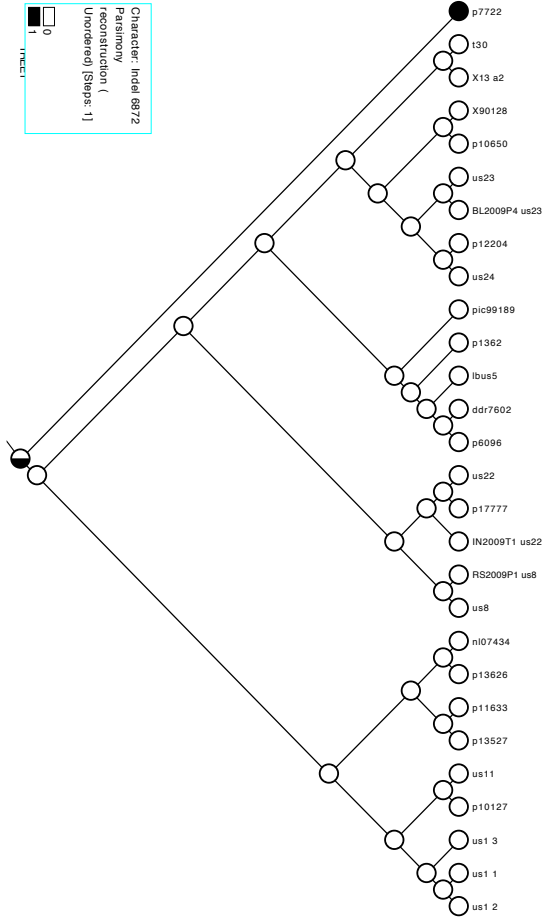
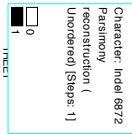
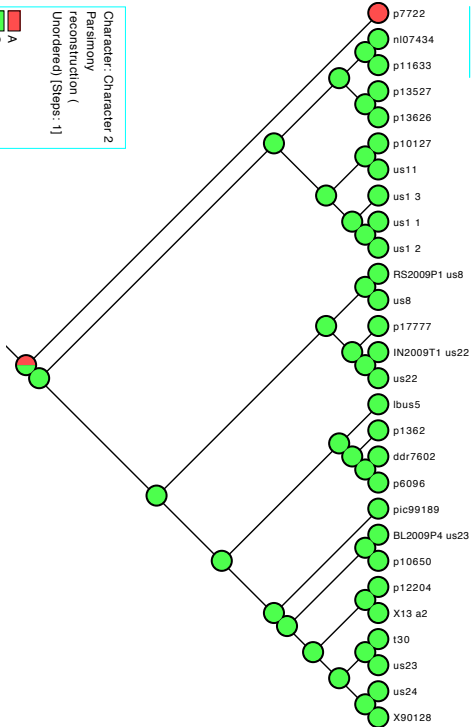
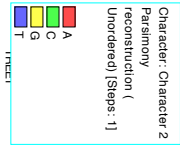
68 Using the whole genome alignment (28 sequences, 39,870 nucleotides) we per-  
69 formed a whole-genome phylogeny using maximum likelihood (RAxML) and  
70 Bayesian inference (BEAST). We used RAxML using no partitions, 1000 boot-  
71 strap replicates, a GTR+I+G model of nucleotide evolution to obtain a biparti-  
72 tioned tree with the bootstrap values mapped to the branches. For BEAST, we  
73 specified p7722 (*P.mirabilis*) as the outgroup. We used a HKY+G+I model of nu-  
74 cleotide substitutions, a strict molecular clock, a constant population size prior,  
75 UPGMA starting tree and 10 million Markov chains. The best tree is shown in  
76 Figure 4.

### 77 3.2 Mapping the SNP's in the BEAST tree

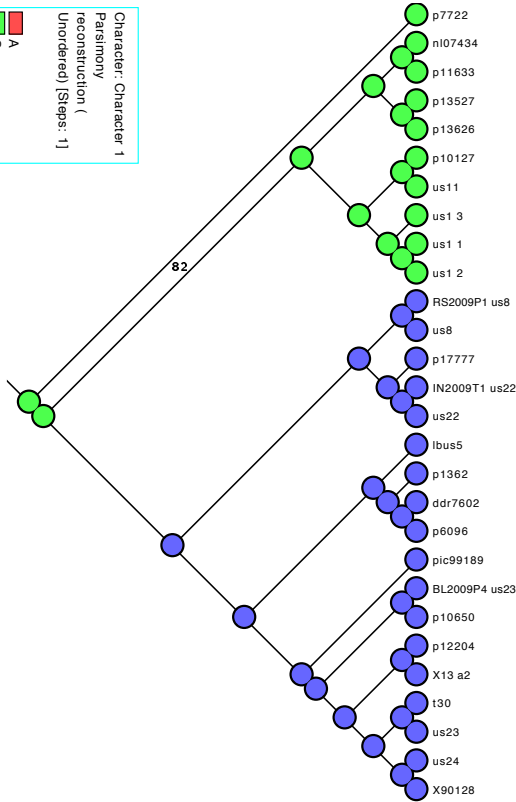
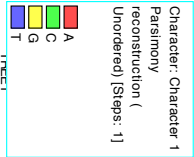
78 To map the variants found in the mtDNA genome to the coalescent tree, we  
79 used Mesquite. We did a removal of invariable regions and ancestral state  
80 reconstruction for all 37 SNPs using a parsimony reconstruction state (Figure  
81 ??).



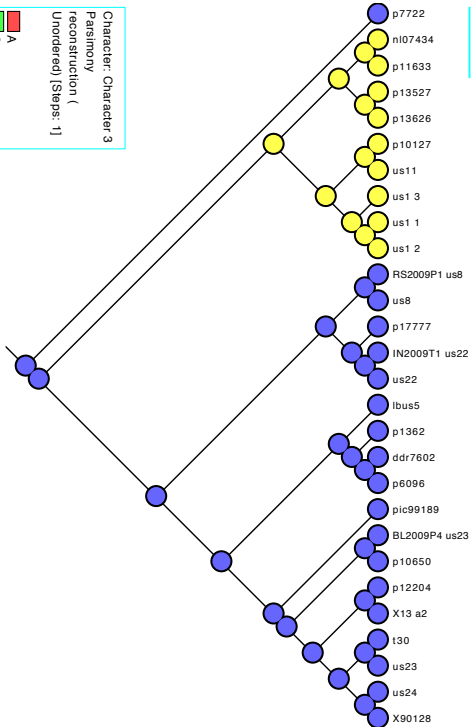
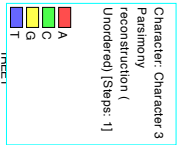


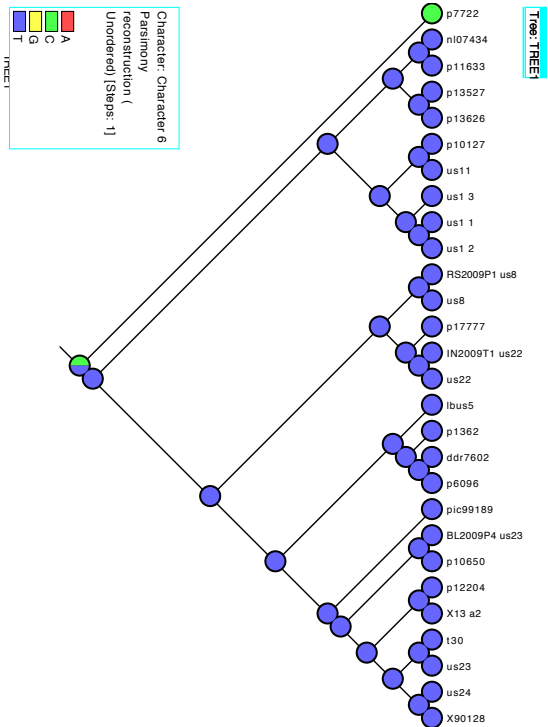
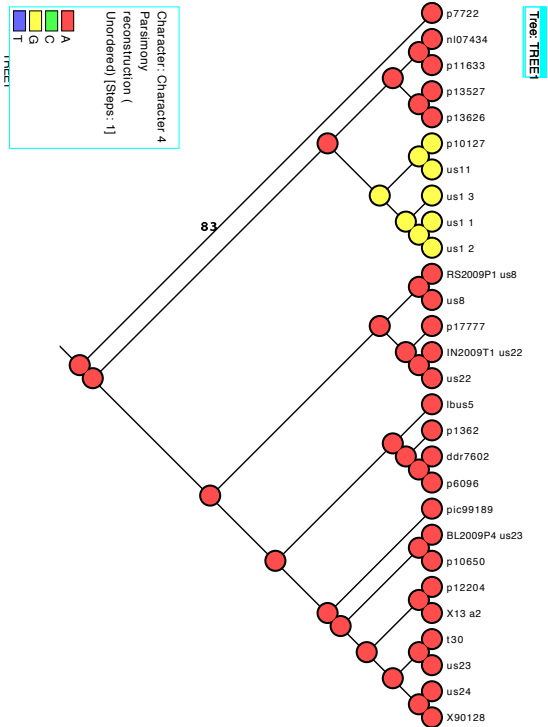
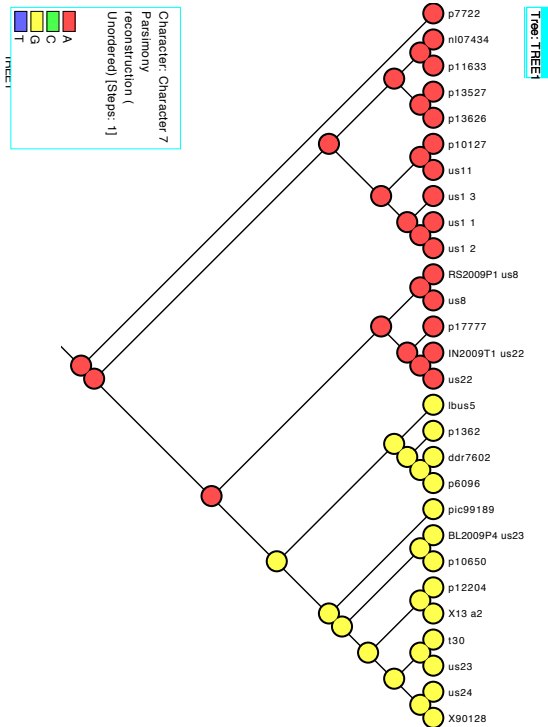
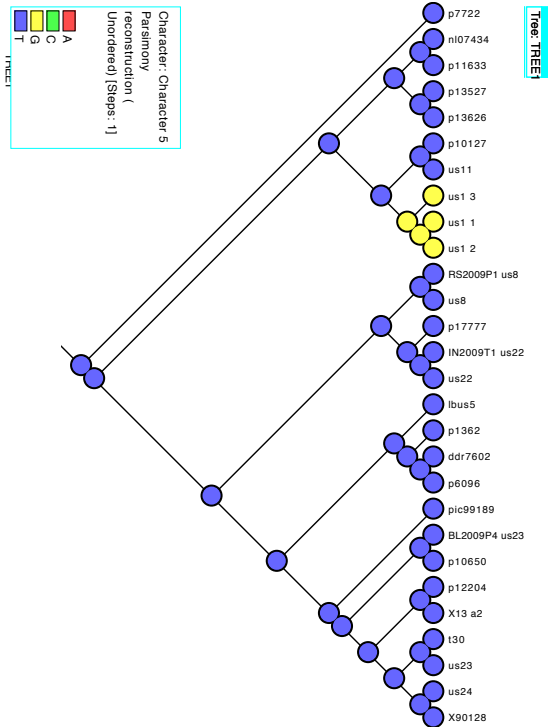


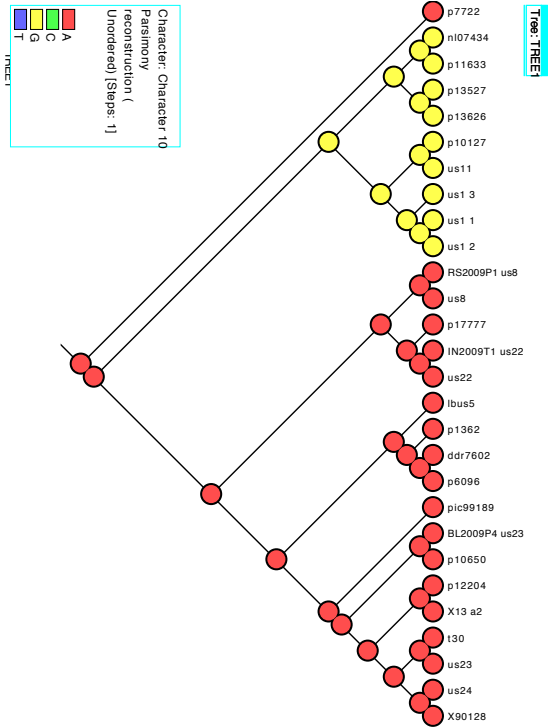
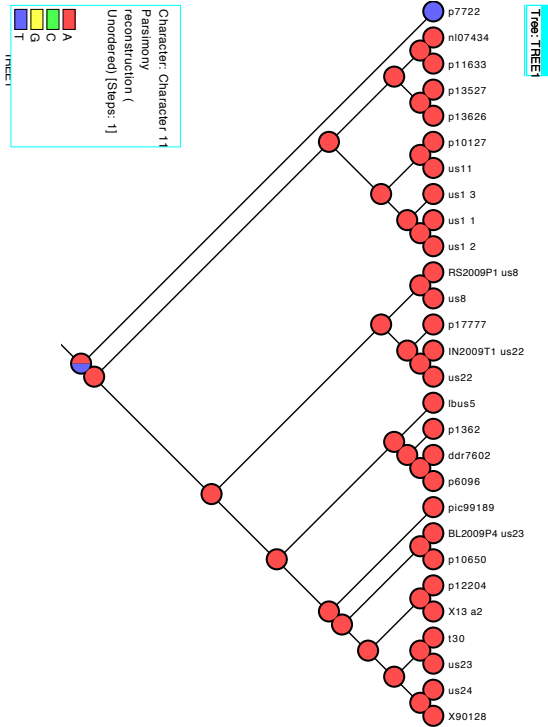
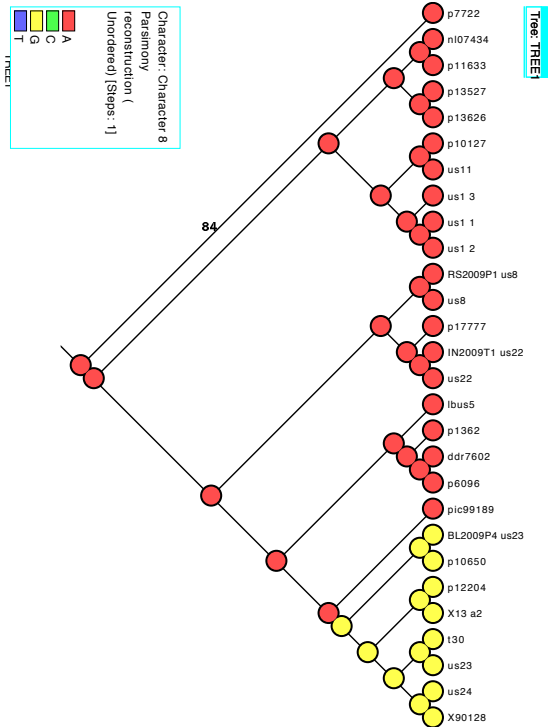
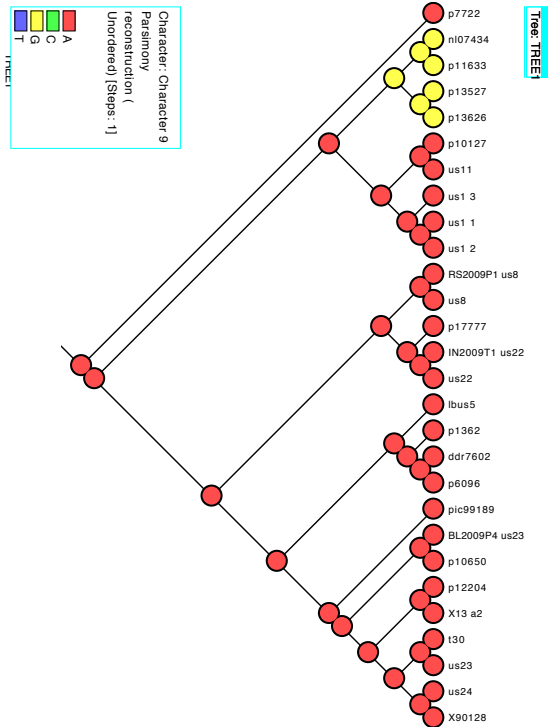
Tree: TREE1

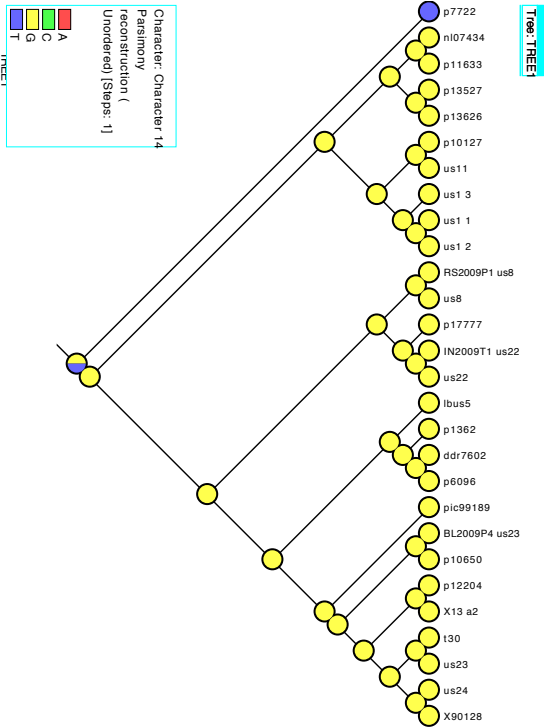
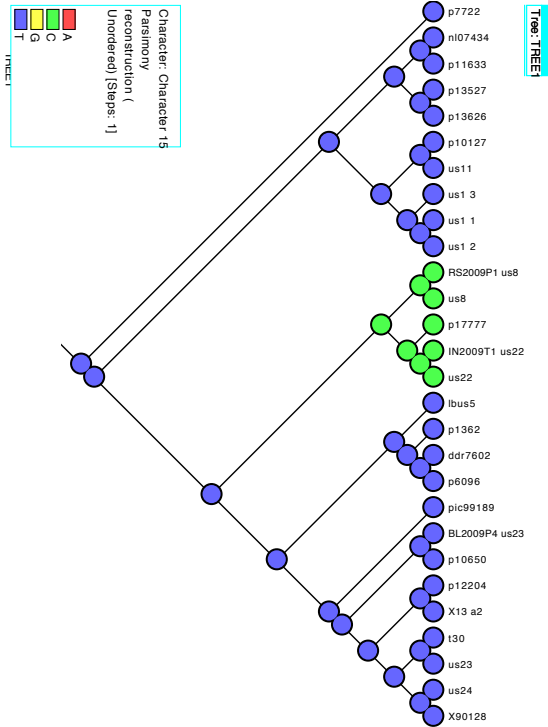
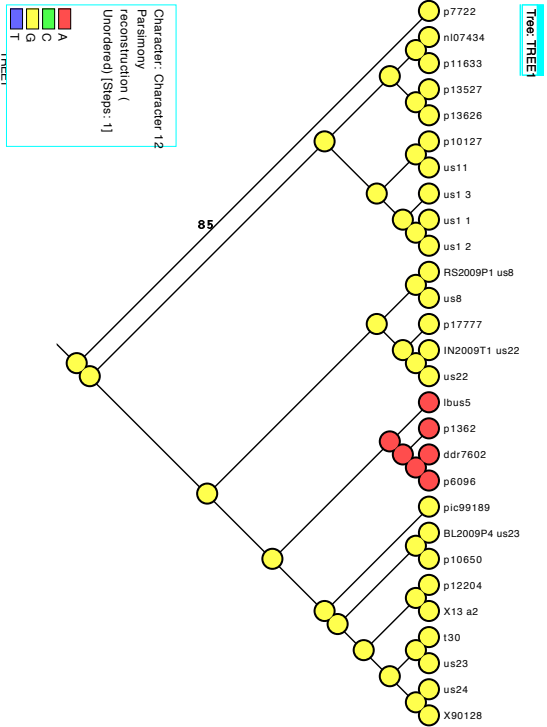
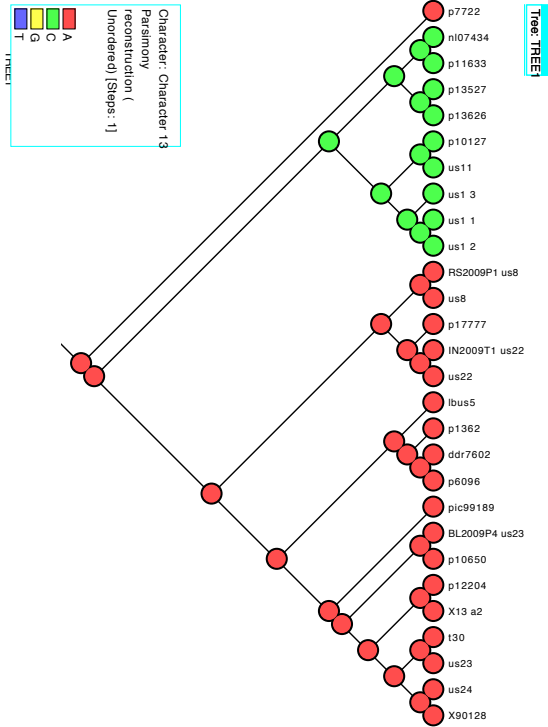


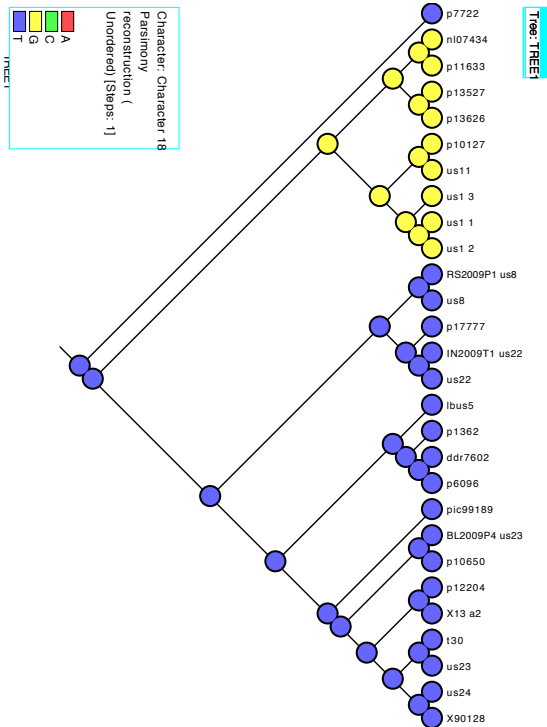
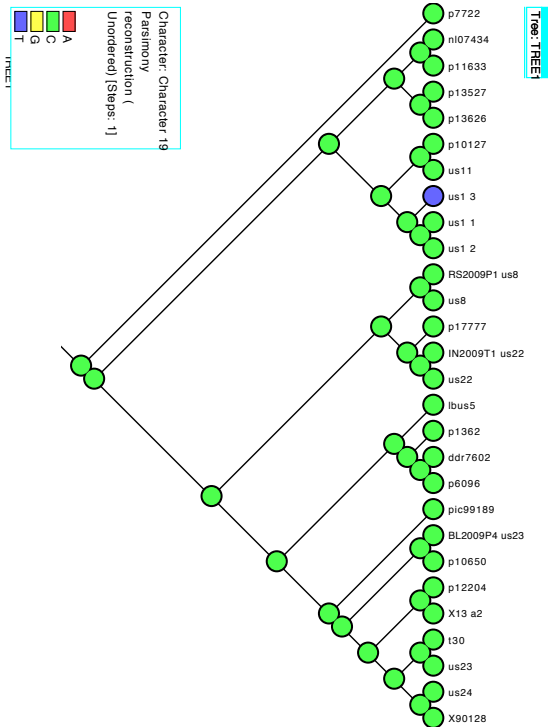
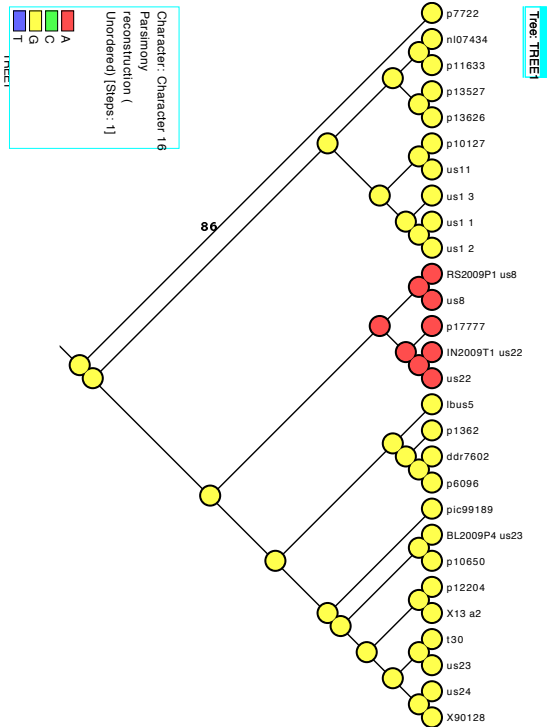
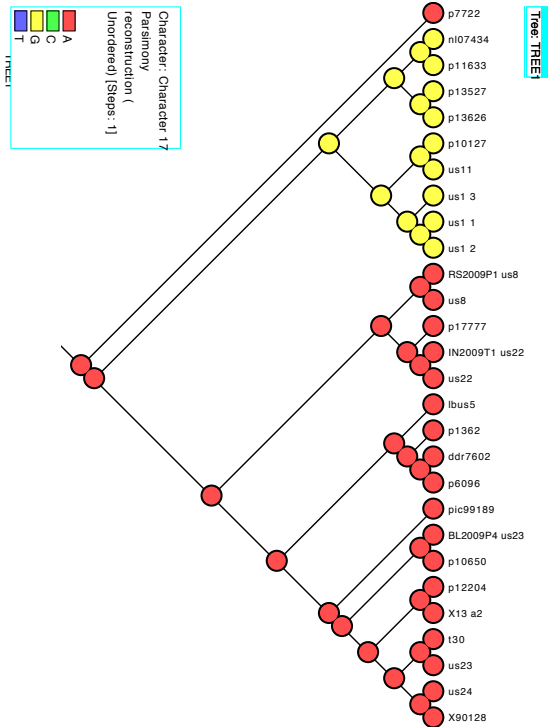
Tree: TREE1

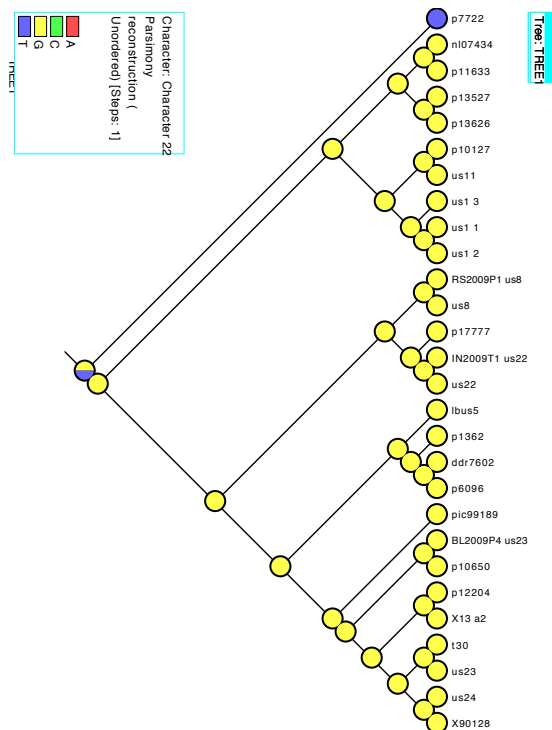
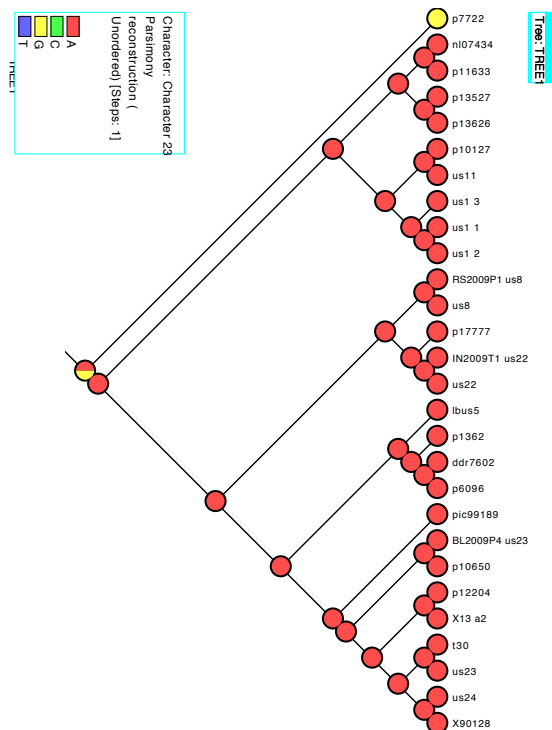
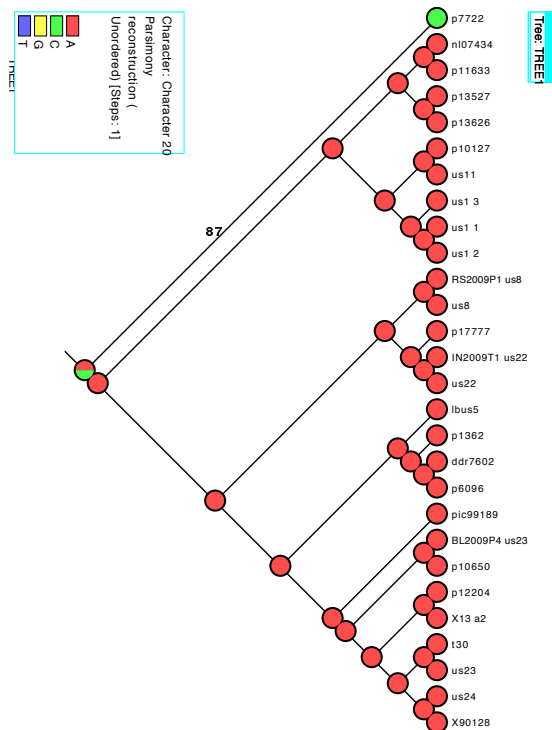
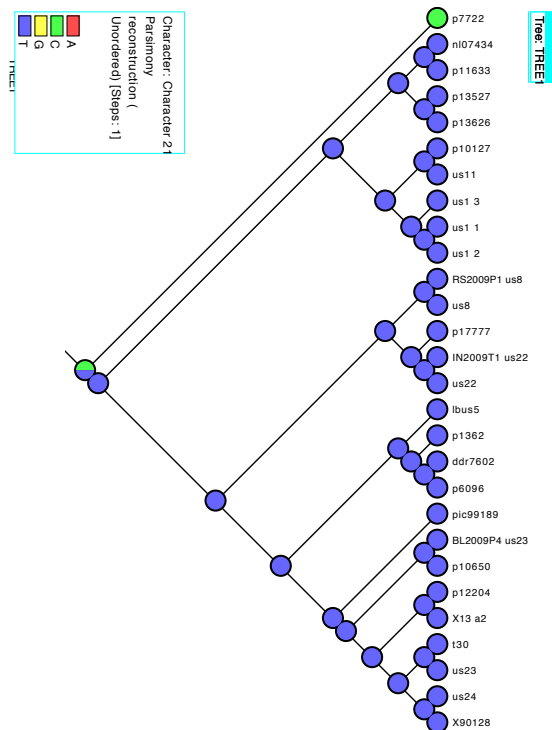






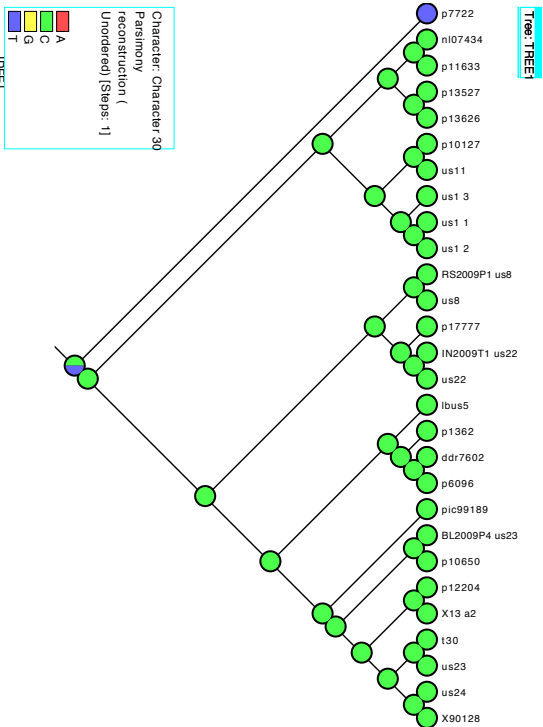
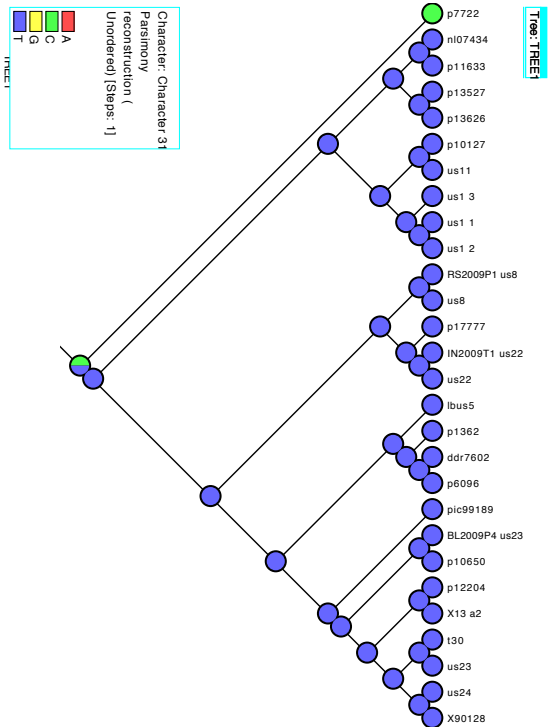
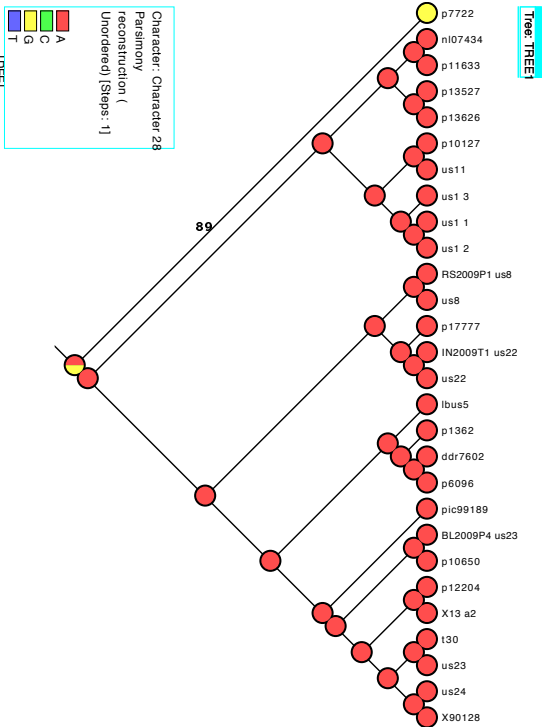
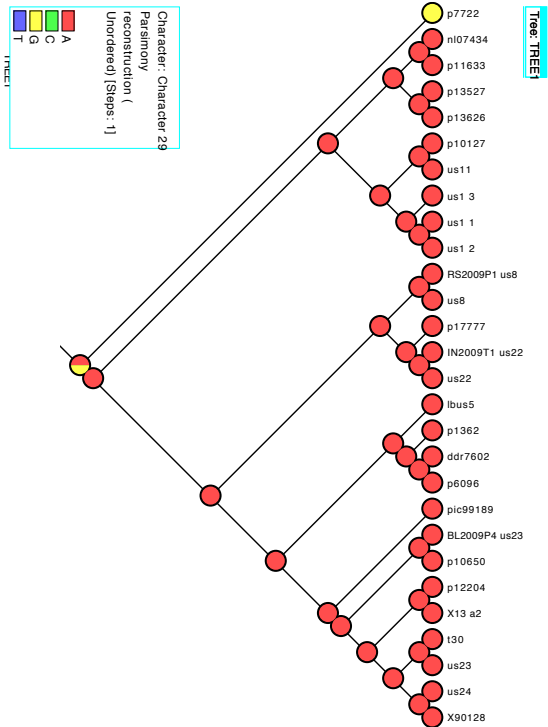




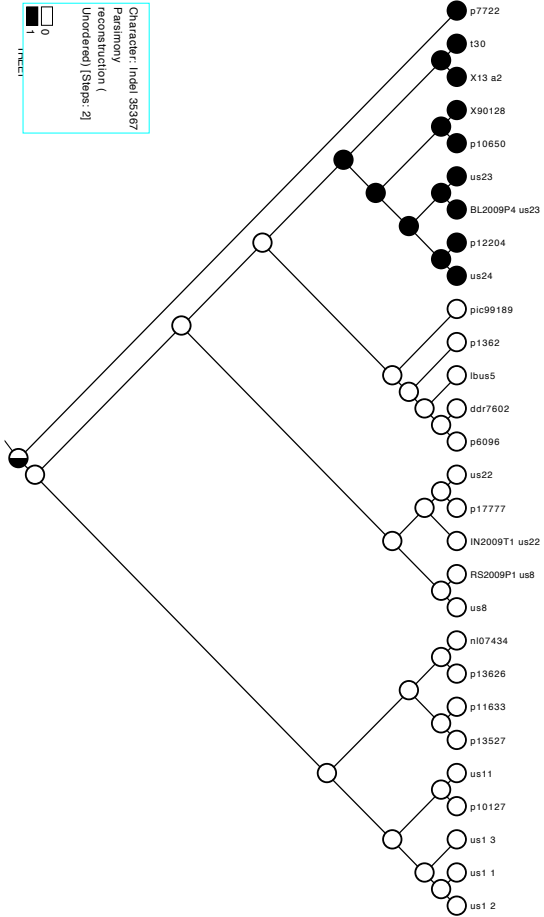




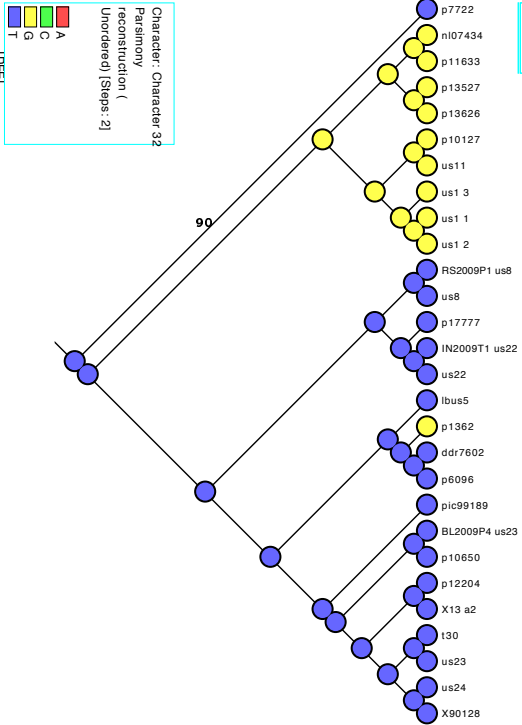




Character: Indel 35367  
Parimony  
reconstruction (Unordered) [Steps: 2]  
0  
1  
Tree: TREE1

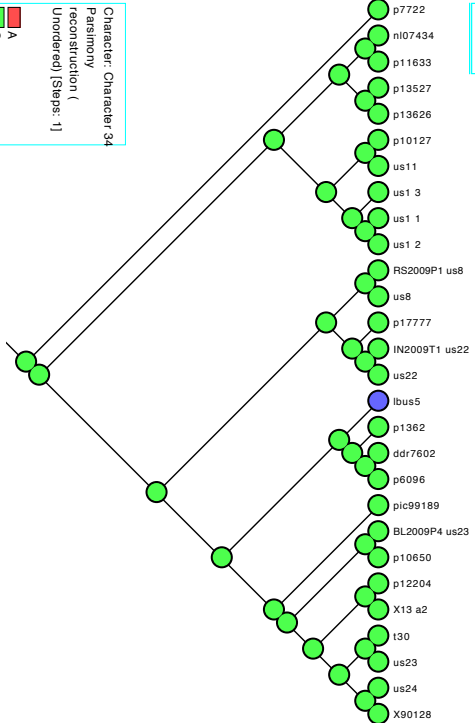


Tree: TREE1

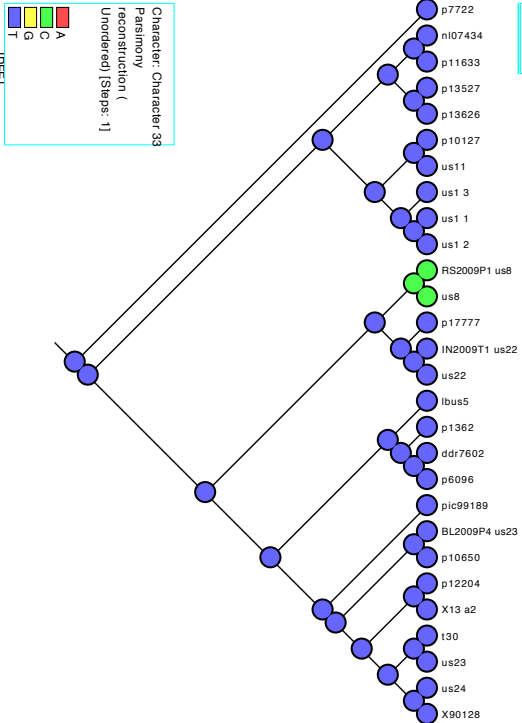


Character: Character 32  
Parimony  
reconstruction (Unordered) [Steps: 2]  
A  
C  
G  
T  
Tree: TREE1

Character: Character 34  
Parimony  
reconstruction (Unordered) [Steps: 1]  
A  
C  
G  
T  
Tree: TREE1



Tree: TREE1



Character: Character 33  
Parimony  
reconstruction (Unordered) [Steps: 1]  
A  
C  
G  
T  
Tree: TREE1

Character: Character 35  
Parimony  
reconstruction (Unordered) [Steps: 1]

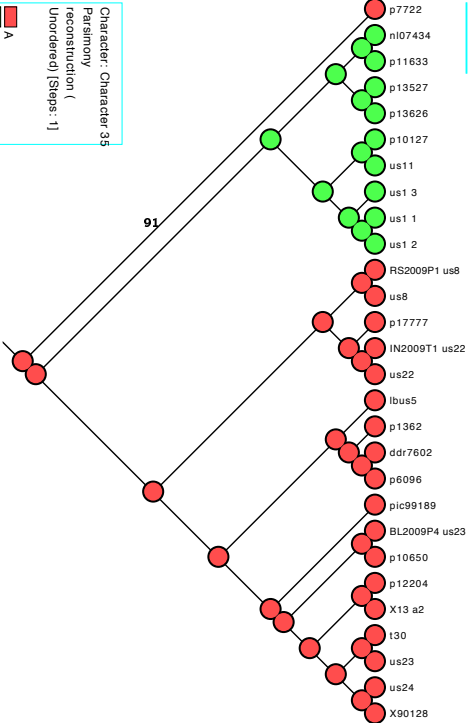
A

C

G

T

Tree



## 92 4 Session information

```

sessionInfo()

## R version 3.0.2 (2013-09-25)
## Platform: x86_64-pc-linux-gnu (64-bit)
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] xtable_1.7-1 vcfR_0.1      knitr_1.5
##
## loaded via a namespace (and not attached):
## [1] ape_3.0-11      evaluate_0.5.1  formatR_0.10    grid_3.0.2
## [5] lattice_0.20-24 nlme_3.1-111    stringr_0.6.2   tools_3.0.2

```

## 93 References

- 94 [1] Cruz Avila-Adame, Luis Gómez-Alpizar, Victoria Zismann, Kristine M  
95 Jones, C Robin Buell, and Jean Beagle Ristaino. Mitochondrial genome  
96 sequences and molecular evolution of the Irish potato famine pathogen, *Phy-*  
97 *tophthora infestans*. *Current genetics*, 49(1):39–46, 2006.
- 98 [2] David EL Cooke, Liliana M Cano, Sylvain Raffaele, Ruairidh A Bain,  
99 Louise R Cooke, Graham J Etherington, Kenneth L Deahl, Rhys A Farrer,  
100 Eleanor M Gilroy, Erica M Goss, et al. Genome analyses of an aggressive  
101 and invasive lineage of the irish potato famine pathogen. *PLoS pathogens*,  
102 8(10):e1002940, 2012.
- 103 [3] G Danies, IM Small, K Myers, R Childers, and William E Fry. Phenotypic  
104 characterization of recent clonal lineages of *Phytophthora infestans* in the  
105 united states. *Plant Disease*, 97(7):873–881, 2013.
- 106 [4] Brian J Haas, Sophien Kamoun, Michael C Zody, Rays HY Jiang, Robert E  
107 Handsaker, Liliana M Cano, Manfred Grabherr, Chinnappa D Kodira, Syl-  
108 vain Raffaele, Trudy Torto-Alalibo, et al. Genome sequence and analy-

- sis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature*, 461(7262):393–398, 2009.
- [5] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–359, 2012.
- [6] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009.
- [7] Michael D Martin, Enrico Cappellini, Jose A Samaniego, M Lisandra Zepeda, Paula F Campos, Andaine Seguin-Orlando, Nathan Wales, Ludovic Orlando, Simon YW Ho, Fred S Dietrich, et al. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nature communications*, 4, 2013.
- [8] Sylvain Raffaele, Joe Win, Liliana M Cano, and Sophien Kamoun. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC genomics*, 11(1):637, 2010.
- [9] Kentaro Yoshida, Verena J Schuenemann, Liliana M Cano, Marina Pais, Bagdevi Mishra, Rahul Sharma, Chirsta Lanz, Frank N Martin, Sophien Kamoun, Johannes Krause, et al. Correction: The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife*, 2, 2013.

|     | CHROM           | POS   | REF         | ALT         | Tree Number |
|-----|-----------------|-------|-------------|-------------|-------------|
| 6   | Supercontig_1.1 | 2701  | C           | T           | 1           |
| 7   | Supercontig_1.1 | 2728  | C           | A           | 2           |
| 12  | Supercontig_1.1 | 3519  | G           | T           | 3           |
| 36  | Supercontig_1.1 | 6872  | accc        | acc         | 4           |
| 42  | Supercontig_1.1 | 7857  | A           | G           | 5           |
| 44  | Supercontig_1.1 | 8118  | T           | G           | 6           |
| 59  | Supercontig_1.1 | 9857  | T           | C           | 7           |
| 62  | Supercontig_1.1 | 10146 | A           | G           | 8           |
| 85  | Supercontig_1.1 | 13684 | A           | G           | 9           |
| 106 | Supercontig_1.1 | 15563 | G           | A           | 10          |
| 114 | Supercontig_1.1 | 16986 | G           | A           | 11          |
| 119 | Supercontig_1.1 | 18440 | A           | T           | 12          |
| 123 | Supercontig_1.1 | 19793 | G           | A           | 13          |
| 124 | Supercontig_1.1 | 20005 | C           | A           | 14          |
| 125 | Supercontig_1.1 | 20290 | G           | T           | 15          |
| 128 | Supercontig_1.1 | 20464 | T           | C           | 16          |
| 141 | Supercontig_1.1 | 22711 | G           | A           | 17          |
| 147 | Supercontig_1.1 | 23431 | G           | A           | 18          |
| 150 | Supercontig_1.1 | 23872 | G           | T           | 19          |
| 154 | Supercontig_1.1 | 24611 | C           | T           | 20          |
| 159 | Supercontig_1.1 | 25142 | A           | C           | 21          |
| 160 | Supercontig_1.1 | 25204 | T           | C           | 22          |
| 161 | Supercontig_1.1 | 25237 | G           | T           | 23          |
| 169 | Supercontig_1.1 | 26684 | A           | G           | 24          |
| 171 | Supercontig_1.1 | 26767 | T           | C           | 25          |
| 186 | Supercontig_1.1 | 28680 | C           | A           | 26          |
| 187 | Supercontig_1.1 | 28783 | A           | G           | 27          |
| 192 | Supercontig_1.1 | 30427 | T           | G           | 28          |
| 193 | Supercontig_1.1 | 30552 | A           | G           | 29          |
| 194 | Supercontig_1.1 | 30591 | A           | G           | 30          |
| 195 | Supercontig_1.1 | 30660 | C           | T           | 31          |
| 201 | Supercontig_1.1 | 31403 | T           | C           | 32          |
| 221 | Supercontig_1.1 | 35296 | G           | T           | 33          |
| 222 | Supercontig_1.1 | 35367 | taaaaaaaaaa | taaaaaaaaaa | 34          |
| 233 | Supercontig_1.1 | 36663 | T           | C           | 35          |
| 239 | Supercontig_1.1 | 37562 | C           | T           | 36          |
| 242 | Supercontig_1.1 | 38345 | C           | A           | 37          |

Table 1: Variants remaining after filtering.

Table 2: Diagnostic SNP positions for the mtDNA genome after filtering.

| Position | SNP | Diagnostic for       |
|----------|-----|----------------------|
| 7857     | A/G | p10127, us11 and us1 |
| 8118     | T/G | us1                  |
| 20464    | T/C | p17777, us22 and us8 |
| 22711    | G/A | p17777, us22 and us8 |
| 26767    | A/G | p10127, us11 and us1 |
| 28783    | A/G | p10127, us11 and us1 |
| 36663    | T/C | us8                  |