

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

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

February 18, 2014

Contents

1	Samples	3
2	Read mapping	3
2.1	Variant filtering	4
3	SNPs after filtering	4
4	Diagnostic SNP mapping	5
4.1	Phylogenetic reconstruction	5
4.2	Mapping the SNP's in the BEAST tree	6
5	Session Information	16

List of Figures

1	Quality results for the mtDNA SNP calls before filtering	5
2	Quality results for the mtDNA SNP calls after filtering	6
3	Whole mtDNA genome scan for the P	7
4	BEAST Coalescent tree	8

List of Tables

1	Diagnostic SNP positions for the mtDNA genome after filtering.	4
---	--	---

1 Samples

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gathered from previous publications as well as data which is not yet available to the public (Judelson, unpublished).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

Note that both the Yoshida and Martin papers [9, 7] included ancient DNA in their analyses. Here we have omitted those samples and focused on modern samples. The sample ‘T30-4’ was the first sequenced genome and is considered the reference for nuclear work. This genome was assembled prior to high-throughput sequencing (i.e., Illumina and 454 technologies). The data presented here are not the sequences used for the paper but are part of a project by the Broad (http://www.broadinstitute.org/annotation/genome/phytophthora_infestans/MultiHome.html) to resequence this sample using Illumina and 454 technologies.

The Judelson data include a sample of US1 which was sampled at three different time points (us1_1, us1_2 and us1_3). We suspect that these were different samples and not necessarily the same clone. Therefore differences among these samples may either be due to biological or technical factors. It also includes a sample of US8 which has been characterized as having fungicide resistance [3]. This lineage was also sequenced by Martin et al. [7]. These are most likely different samples so differences among these samples may be interpreted as biological.

The Yoshida data includes one sample of *P. mirabilis* (p7722), this should jump out in the analyses.

For the mitochondrial data we used the type IIa form [1] because it was the longest sequence and we felt this would provide the best alignment.

For enigmatic reasons, not all of the samples from the Yoshida and Martin papers were actually available online. Therefore our numbers here do not match those presented in the papers.

2 Read mapping

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

2.1 Variant filtering

To check which of the calls are good calls, the variant files were filtered by quality, read depth and mapping quality (Figure 2). For this we used an in-house R package Brian wrote called [vcfR](#). The results of these statistics for the whole mitochondria can be seen in Figure 3. Note that quality here is for each individual call and ranges from 1-99. This is different from the per variant quality discussed above which ranges from 1-999. Most of these heterozygous sites are of low quality. The samples which included high quality heterozygote calls (p1362, p6096, p10650, p12204, p10127) were mostly from the Yoshida et al. [9] paper and were among the low sequencing depth samples they included. Because these samples are not among the US lineages we're interested in and because they are apparently of low sequencing depth I think we can ignore them for now. However, the sample nl07434 was among the high sequencing depth samples from this paper and is perhaps noteworthy. T30-4 was called as a heterozygote for one variant and is perplexing.

```
## Before filtering:
## [1] 247
## After filtering:
## [1] 70
## gt.m2sfs is commented out
```

3 SNPs after filtering

After filtering we proceeded to look at which of the SNP's were supported after the filtering (Table 1).

Table 1: 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

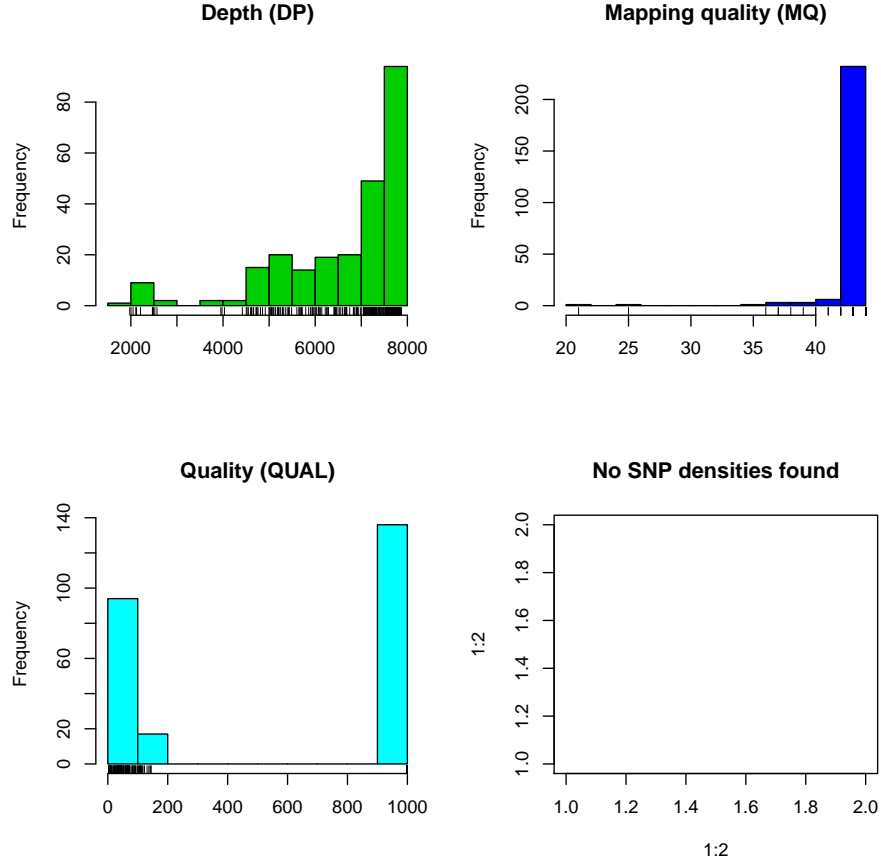


Figure 1: Quality results for the mtDNA SNP calls before filtering.

4 Diagnostic SNP mapping

4.1 Phylogenetic reconstruction

Using the whole genome alignment (28 sequences, 39870 nucleotides) we performed a whole-genome phylogeny using likelihood (RAxML) and bayesian inference (BEAST). We used RAxML using no partitions, 1000 bootstrap replicates, a GTR+I+G model of nucleotide evolution to obtain a bipartitioned tree with the bootstrap values mapped to the branches. For BEAST, we specified *p7722* (*P.mirabilis*) as the outgroup. We used a HKY+G+I model of nucleotide substitutions, a strict molecular clock, a constant population size prior, UPGMA starting tree and 10 million markov Chains. The best tree is shown in Figure 4.

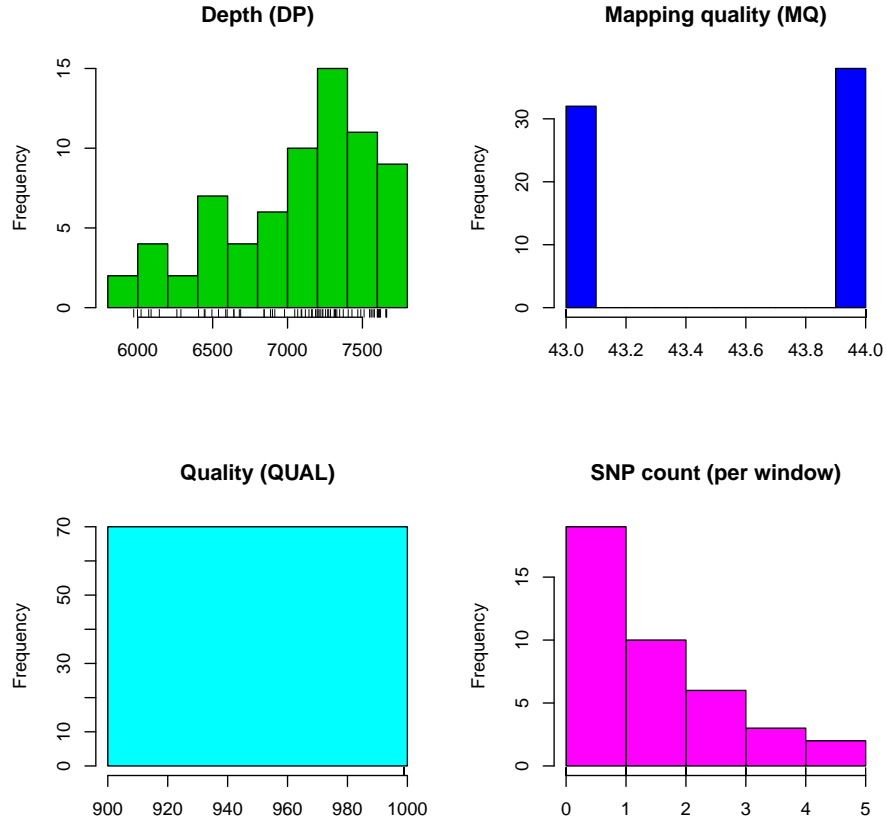
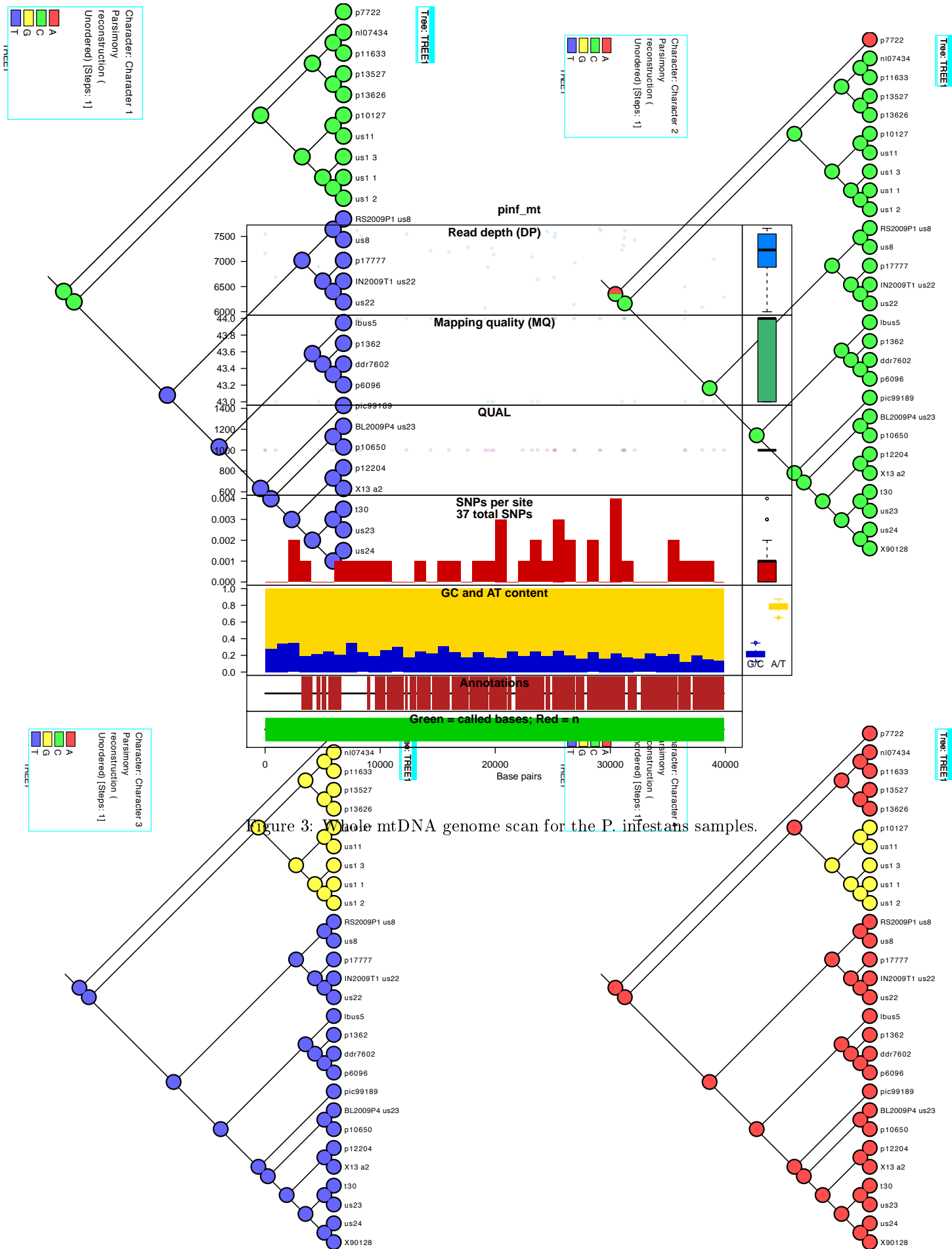


Figure 2: Quality results for the mtDNA SNP calls after filtering.

4.2 Mapping the SNP's in the BEAST tree

To map the SNP's found in the mtDNA genome to the coalescent tree, we used *Mesquite*. We did a removal of invariable regions and ancestral State reconstruction for all 31 SNPs using a parsimony reconstruction state (Figure ??).



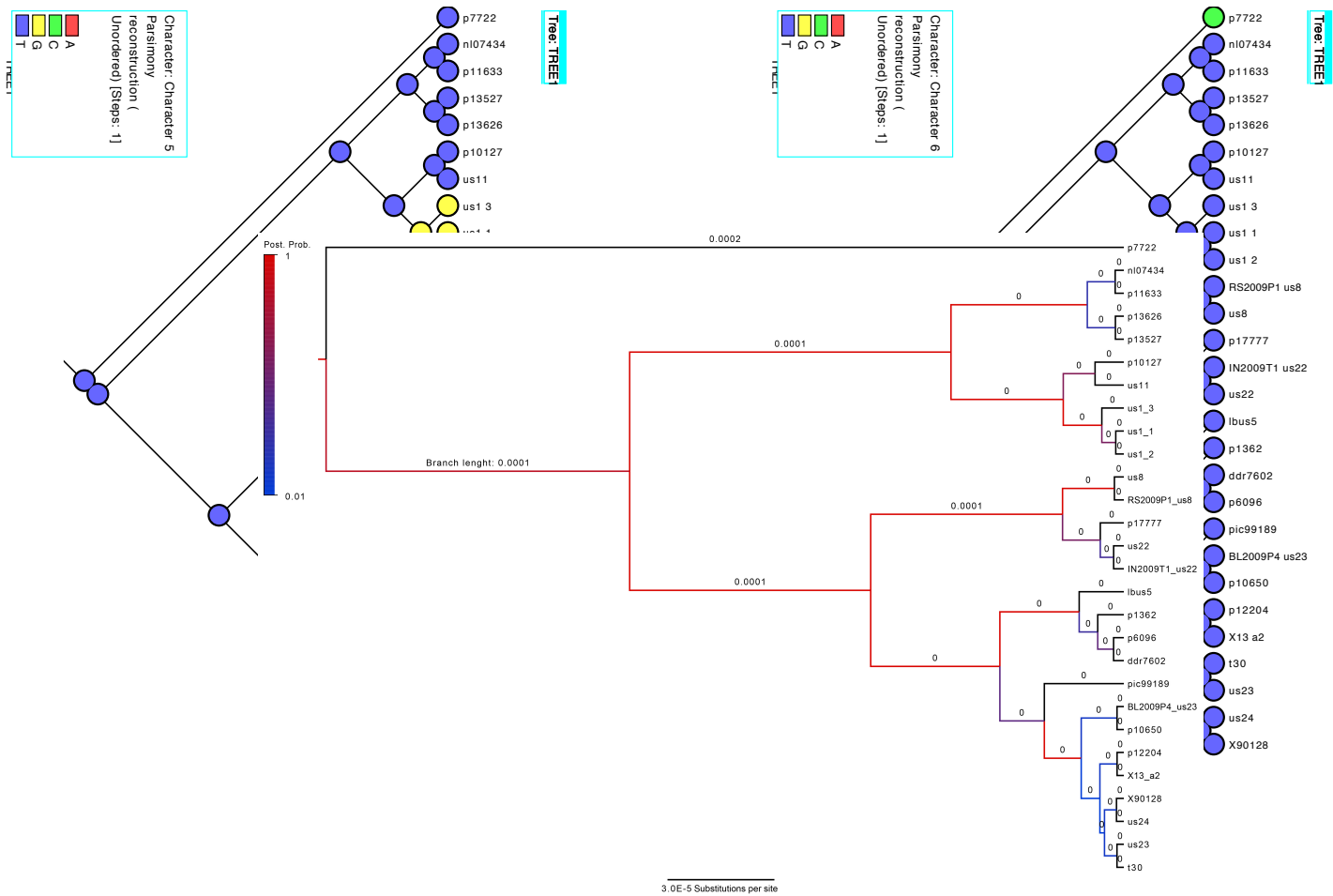
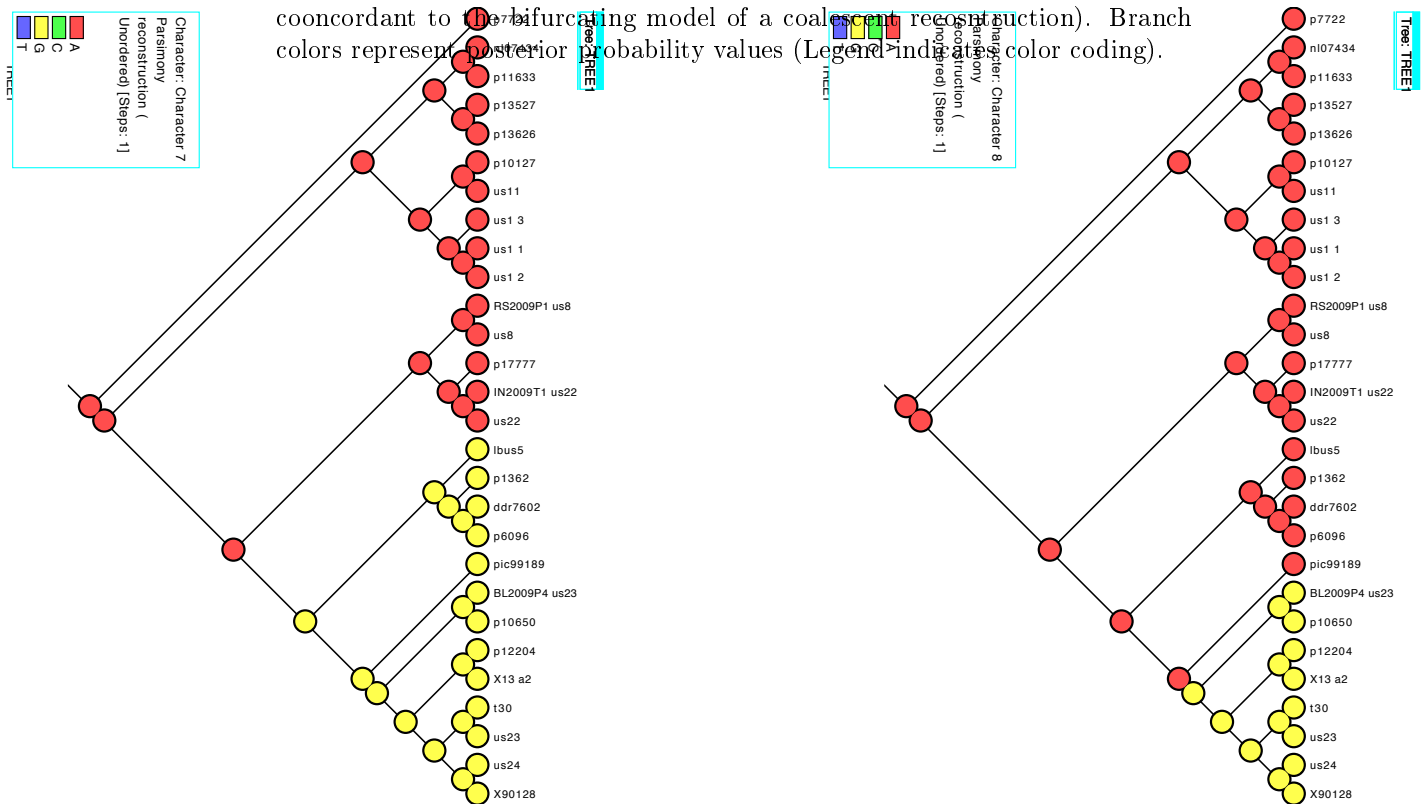
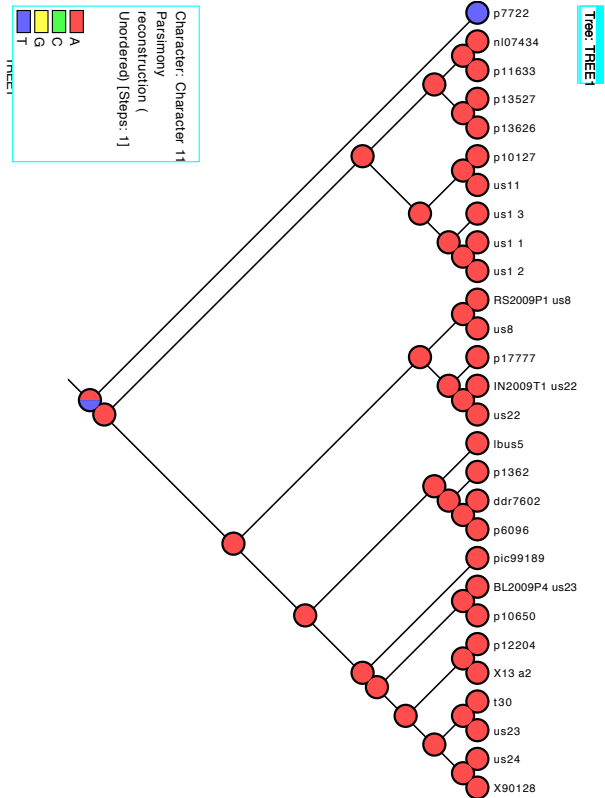
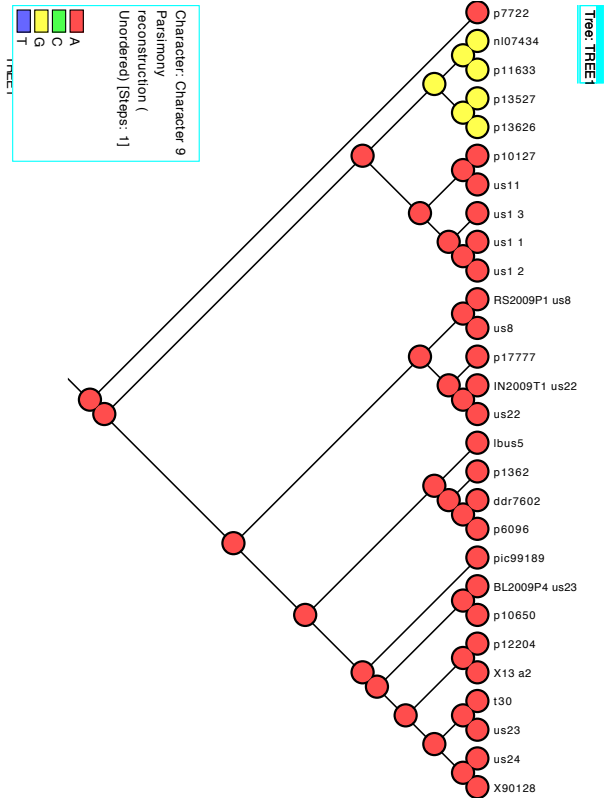
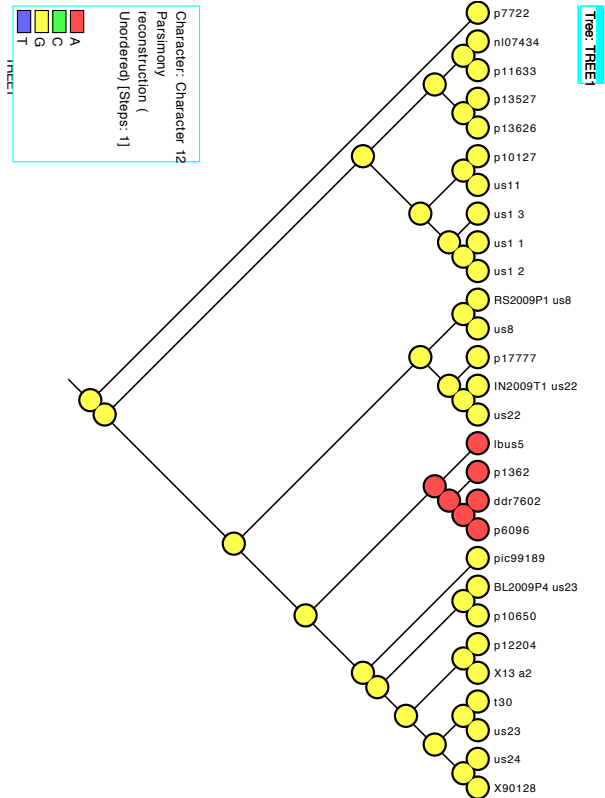
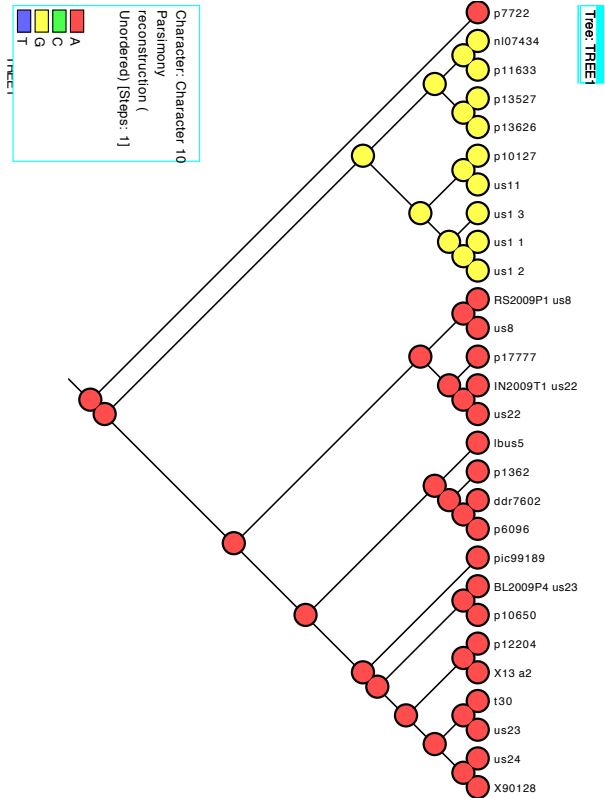


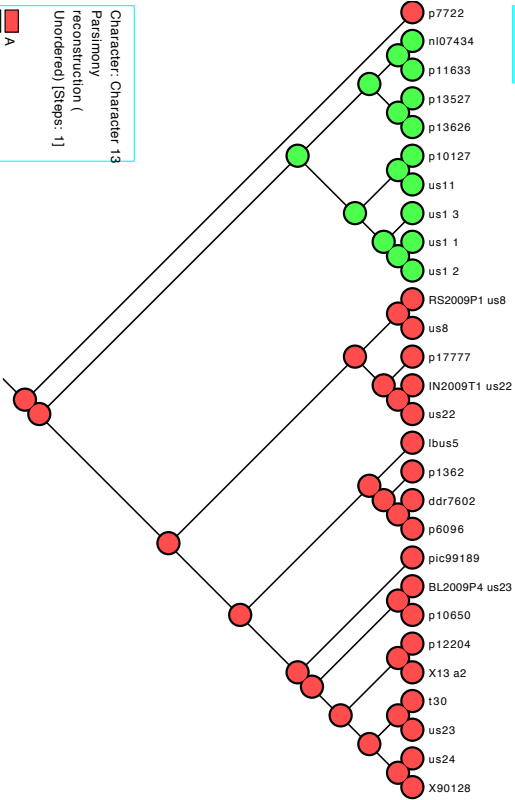
Figure 4: Bayesian coalescent of the whole mtDNA genome of *P. infestans* using BEAST 1.8.0. Values above branches represent branch lengths (Topology is concordant to the bifurcating model of a coalescent reconstruction). Branch colors represent posterior probability values (Legend indicates color coding).





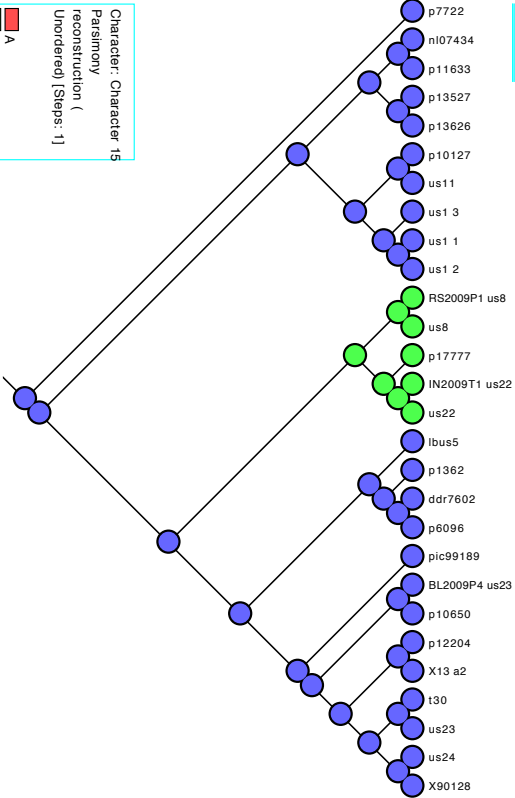
Character: Character 13
Parsimony
reconstruction (Unordered) [Steps: 1]

A
C
G
T



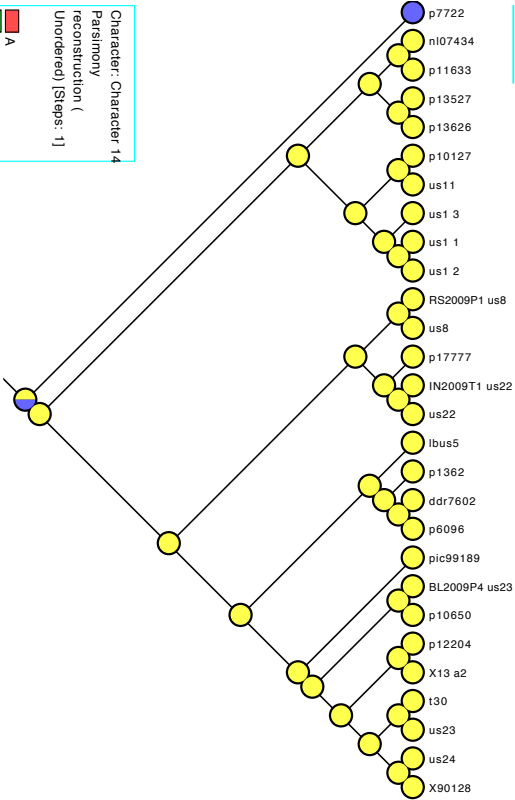
Character: Character 15
Parsimony
reconstruction (Unordered) [Steps: 1]

A
C
G
T



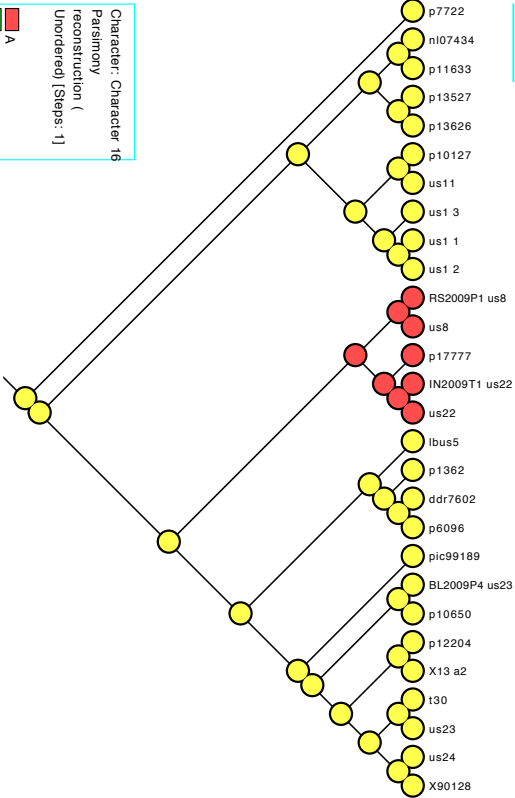
Character: Character 14
Parsimony
reconstruction (Unordered) [Steps: 1]

A
C
G
T



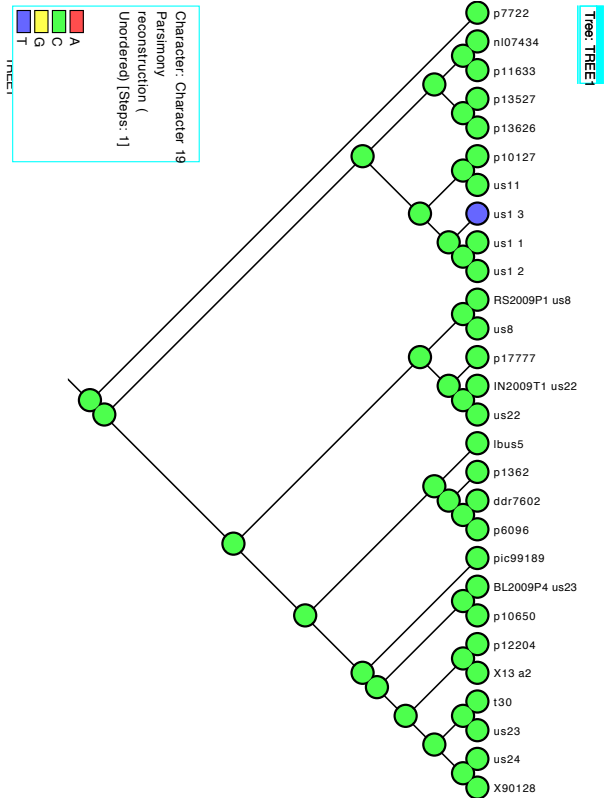
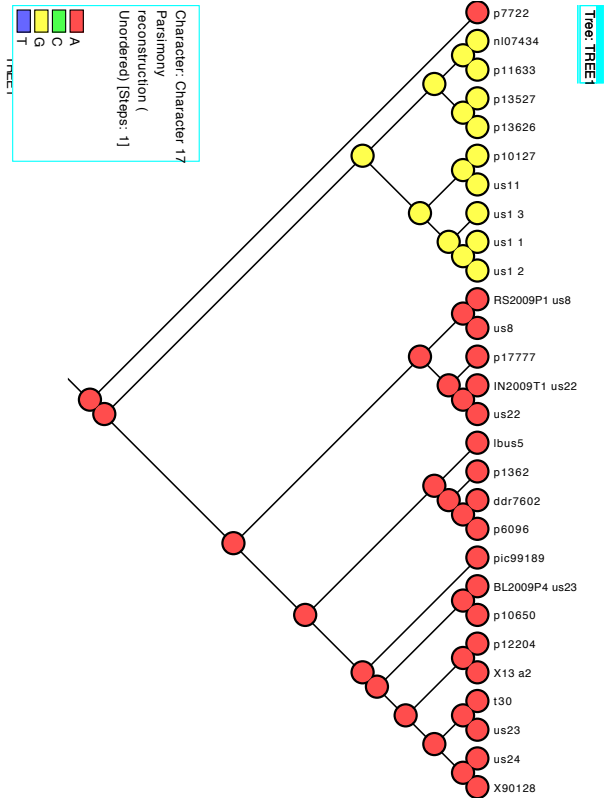
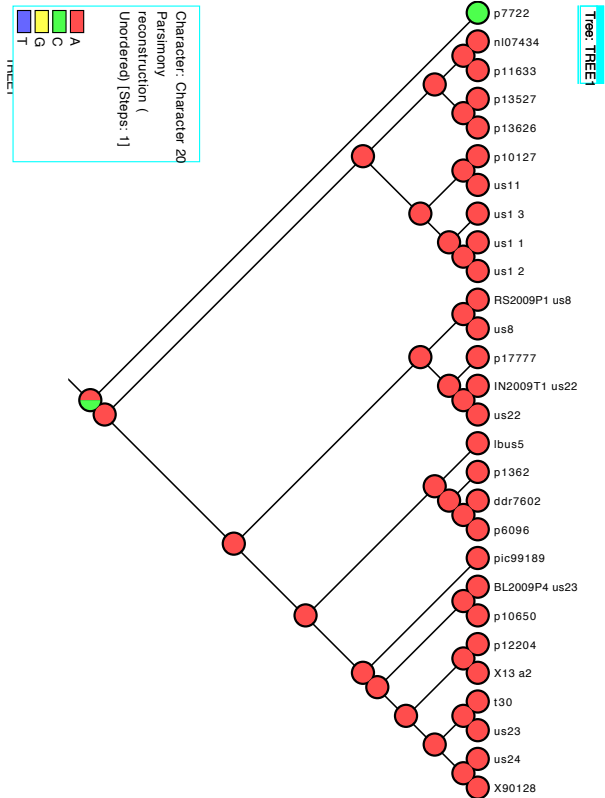
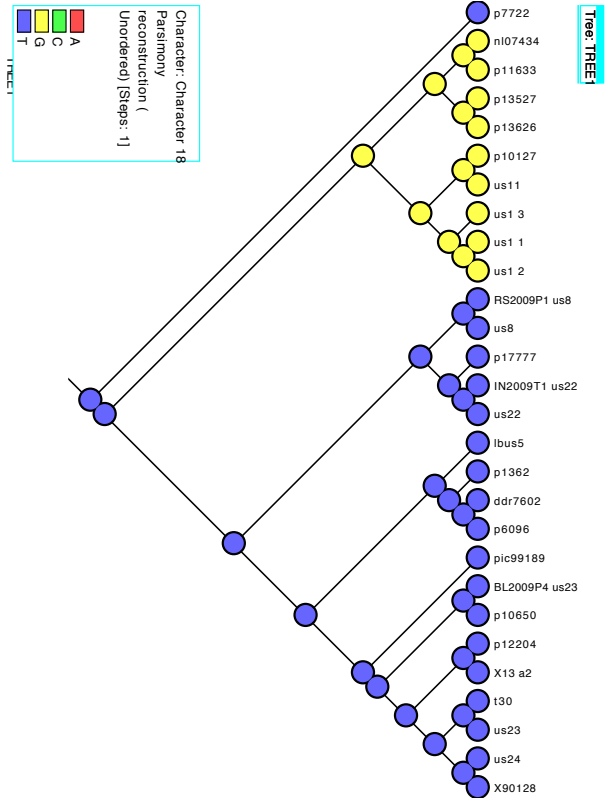
Character: Character 16
Parsimony
reconstruction (Unordered) [Steps: 1]

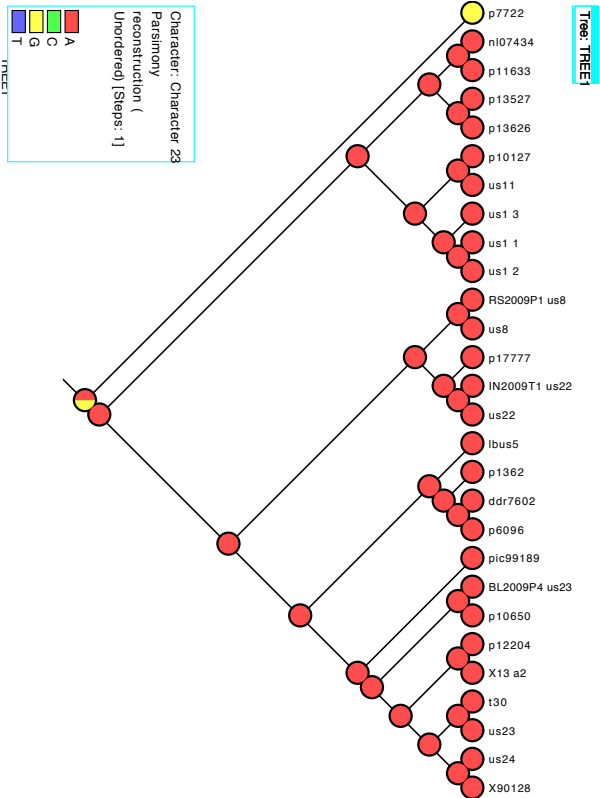
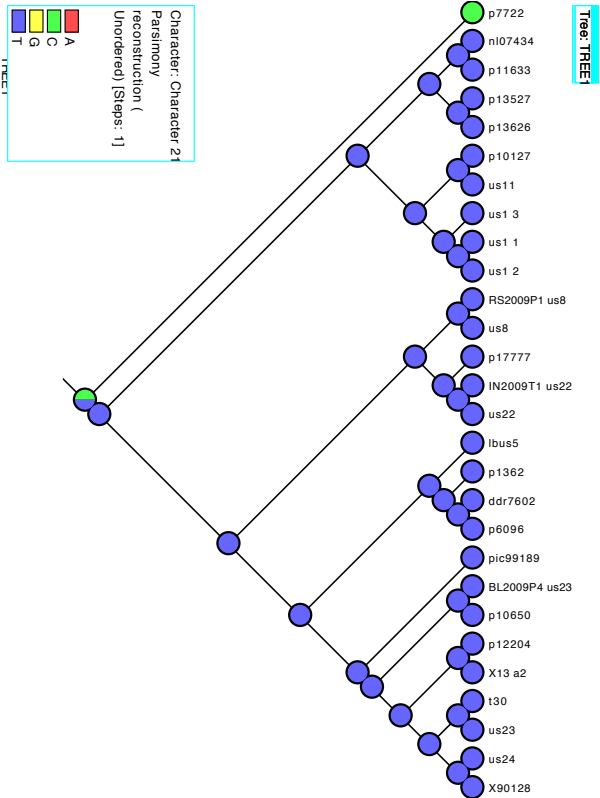
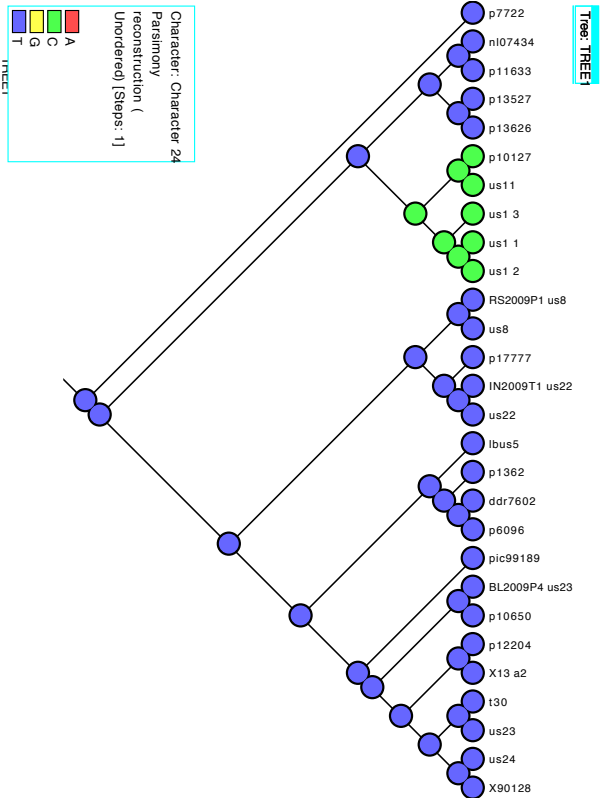
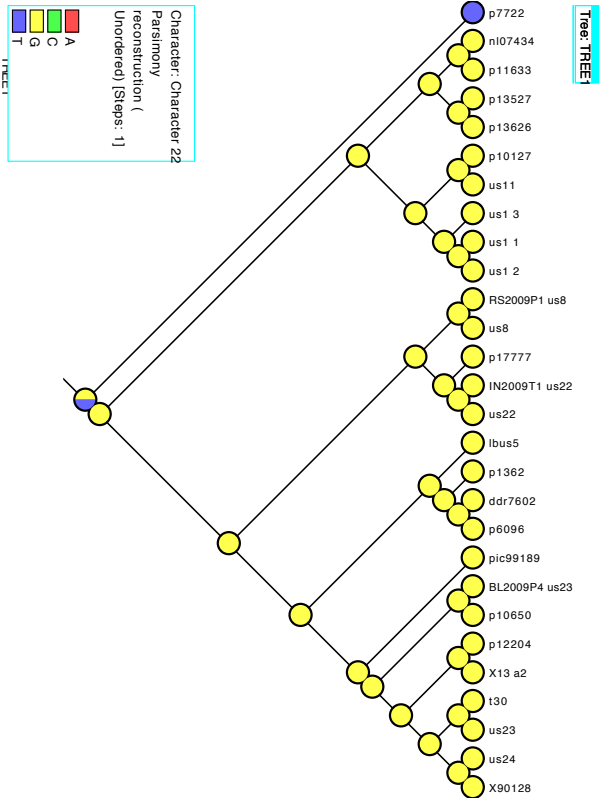
A
C
G
T

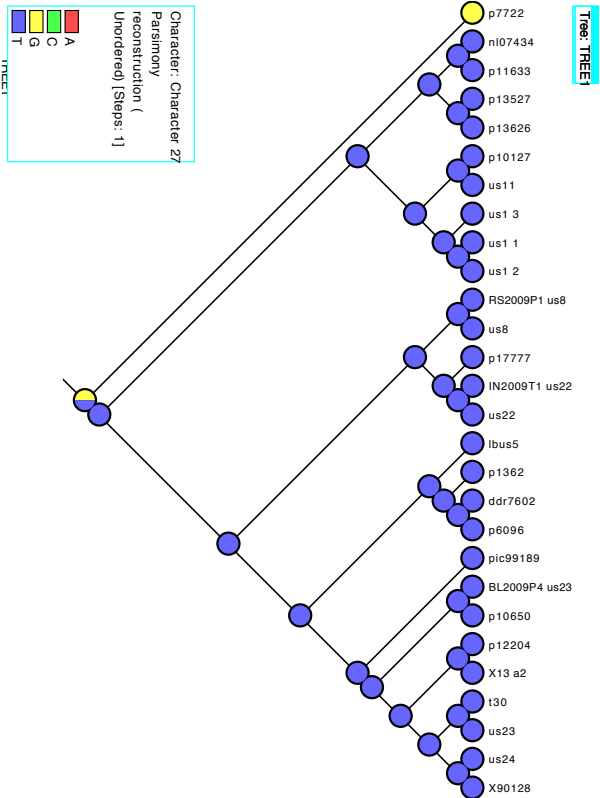
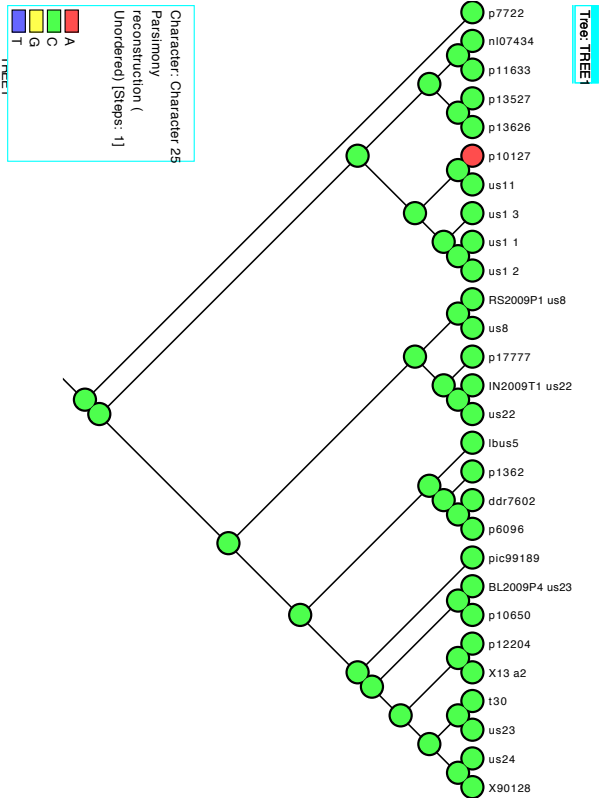
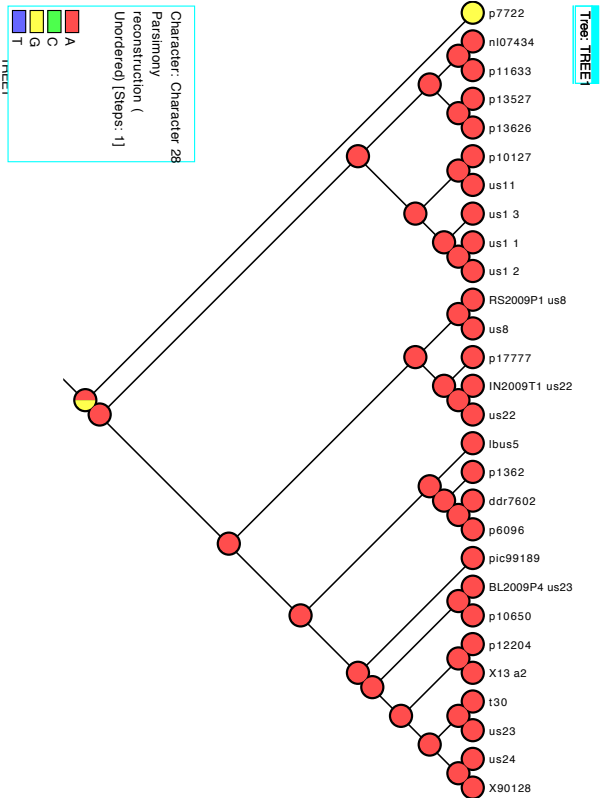
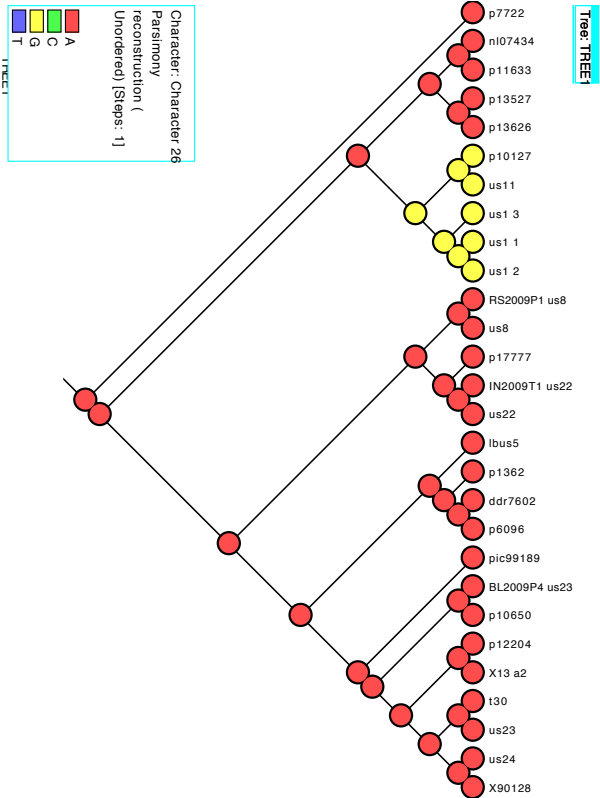


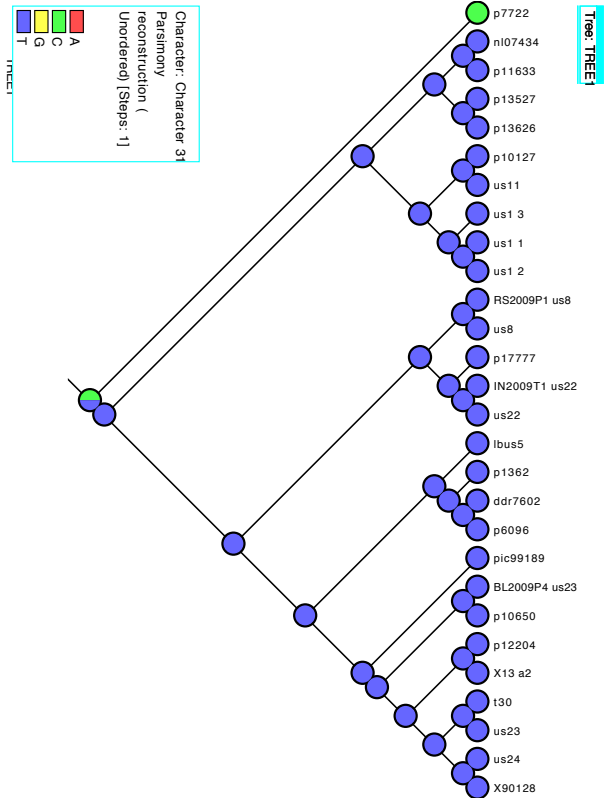
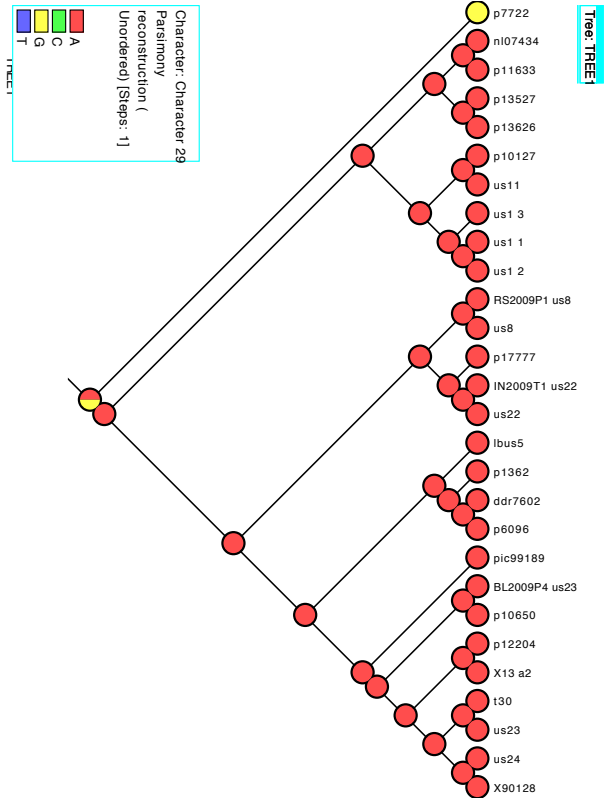
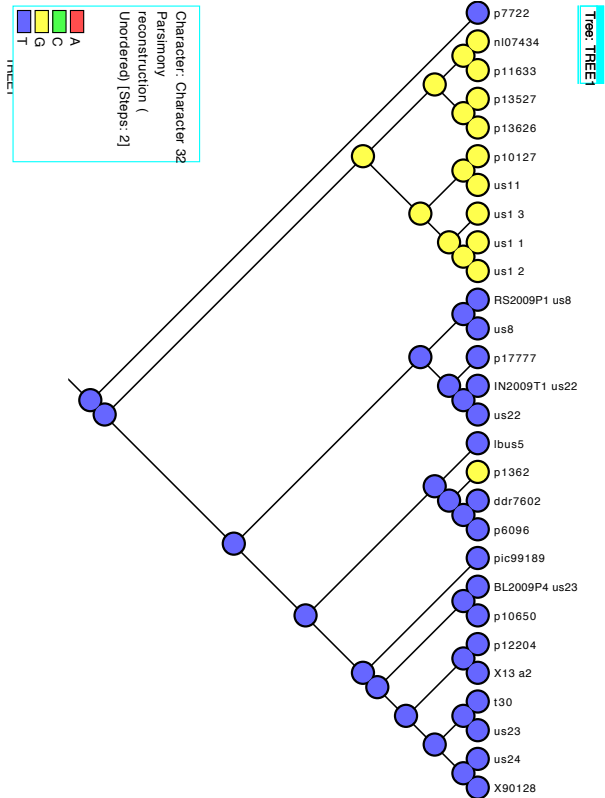
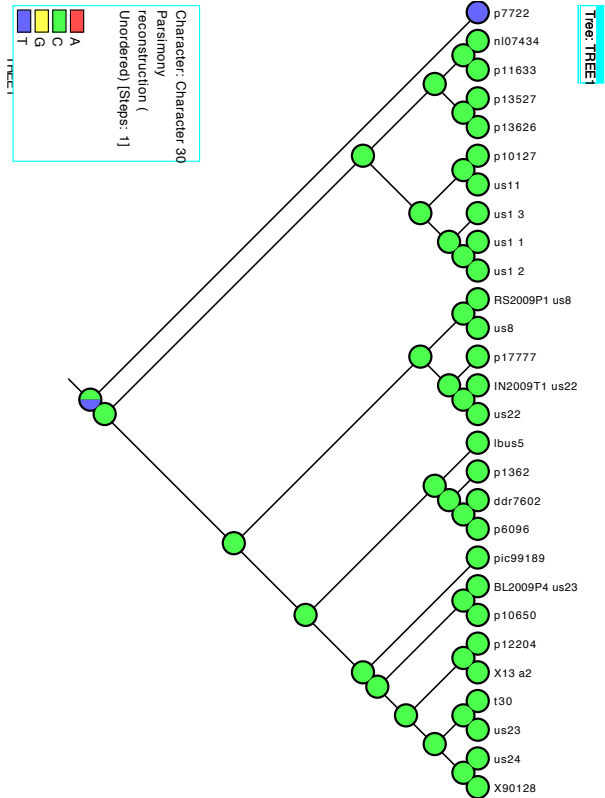
Tree: TREE1

Tree: TREE1



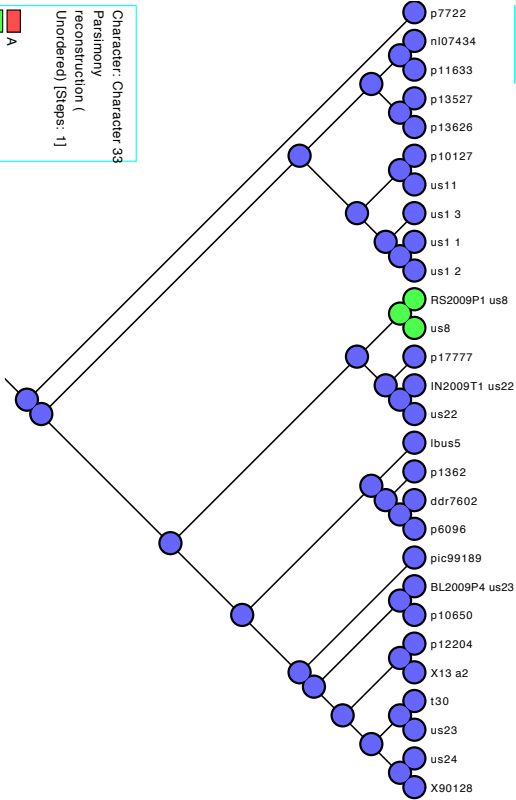






Character: Character 33
Parsimony
reconstruction (Unordered) [Steps: 1]

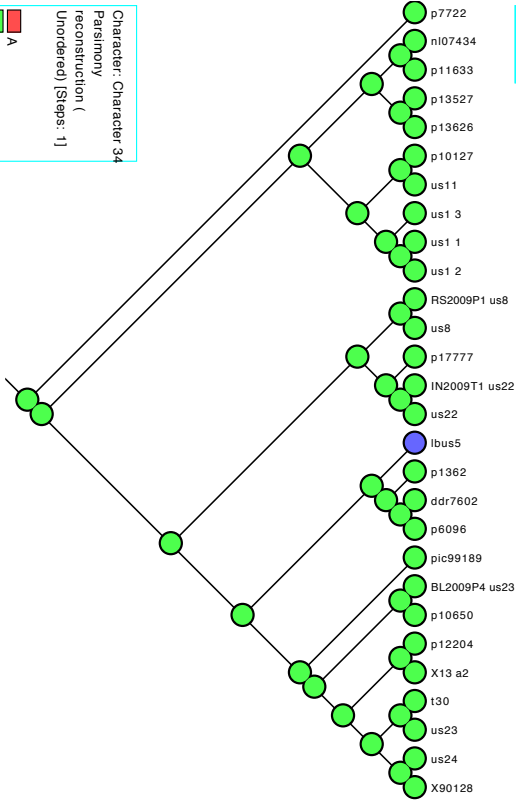
A
C
G
T



Tree: TREE1

Character: Character 34
Parsimony
reconstruction (Unordered) [Steps: 1]

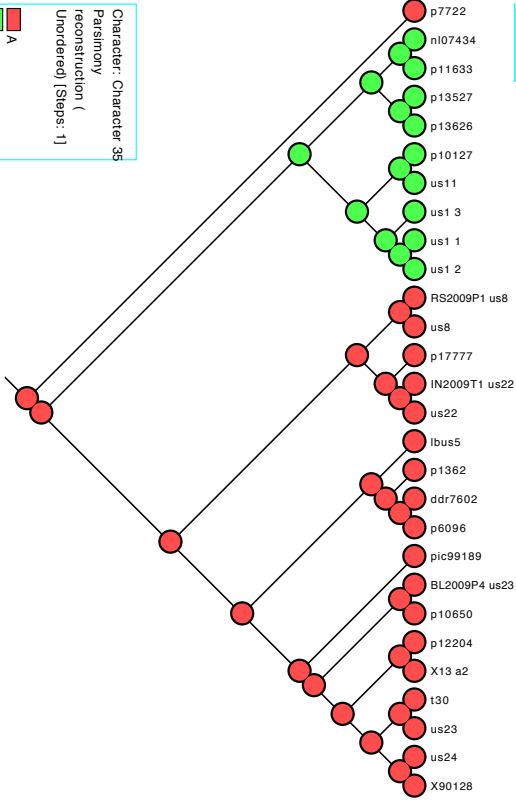
A
C
G
T



Tree: TREE1

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

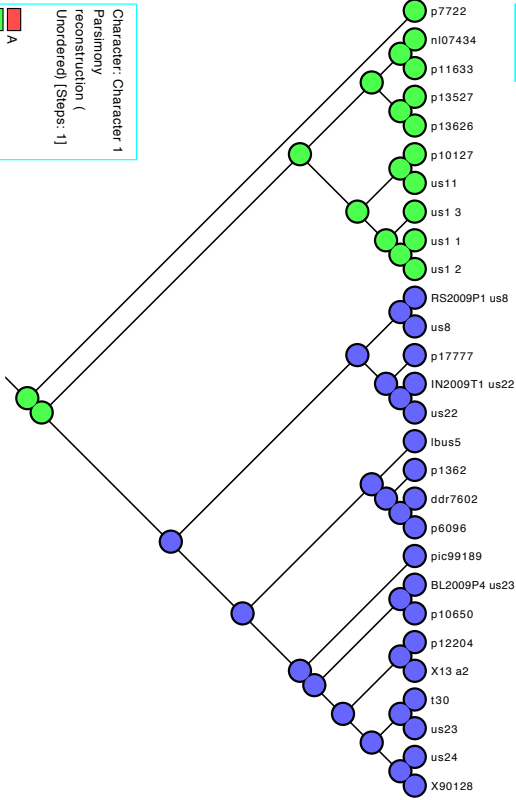
A
C
G
T



Tree: TREE1

Character: Character 1
Parsimony
reconstruction (Unordered) [Steps: 1]

A
C
G
T



Tree: TREE1

5 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] 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
```

References

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