Text S3. Read depth

The inference of copy number, including gene loss, is dependent on sequence depth. For example, the loss of a gene may be determined by a lack of sequence coverage. Similarly, the presence of three copies of a gene may be inferred when the sequence depth for alleles at a locus are present in multiples of $1/3$. Insufficient sequence depth could result in the incorrect inference that a gene has been lost when it may actually be below the level of detection, perhaps because it is in a region of the genome that is difficult to sequence. Similarly, if a gene exists as two copies but is sequenced at 3X coverage it would only be possible to incorrectly infer that it has three copies. In order to address the issue of sufficient sequence depth we began by filtering the available genomes to only analyze genomes of relatively high sequence depth.

Sequenced read depth for genes was used to determine whether a sample had sufficient coverage for further analysis. Per position read depth was measured using SAMtools mpileup [1]. A GC corrected estimate of per gene sequence depth was then calculated following Raffaele and colleagues [4]:

$$^AARD(g) = ARD(g) * mARD/mARD_{GC}.$$

Here $ARD(g)$ is the median read depth for a gene. Raffaele and colleagues [4] reported average read depth but we have used median read depth to provide a robust estimate, but have retained the nomenclature of Raffaele and colleagues. $mARD$ is the median read depth over all genes and $mARD_{GC}$ is the median read depth for genes in the same GC percentile as the gene being evaluated. This results in a GC adjusted average read depth $^AARD(g)$.

A threshold of 12X coverage was used to determine suitability for analysis. Because our estimate of copy number relied on calculating the frequency of allele depth we chose a number that was greater than 10. Also, because we expected some samples to have three copies we chose a threshold that would divide evenly by two and three. Lastly, the choice of threshold was a compromise between attempting to select the highest coverage samples available, but also retaining as many samples as possible. The samples KM177497, M.0182897, and M.0182904 [6] were omitted because they did not appear to have enough data to process. Violin plots of were created of $^AARD(g)$ for each available sample using ggplot2 [5].

The samples T30-4, PIC99189, 90128, PIC99167, PIC99114, F18 [4], M.0182897, M.0182898, M.0182900, M.0182903, M.0182904, M.0182906, M.0182907, KM177497, KM177500, KM177509, KM177512, KM177513, KM177514, KM177517, KM177548, P11633, P12204, P1362, P6096 [6], Pi1845A, Pi1845B, Pi1876, Pi1882 [2], Kew123, P3681, P6629, P6635, PHU006, PCZ050, P10650, P3873, EC3394 [3], PIC97136, PIC97442 (this report) were considered but omitted based on the criterion that they had a genic AARD less than 12.

References

[1] H Li, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, R Durbin, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009.

[2] MD Martin, E Cappellini, JA Samaniego, ML Zepeda, PF Campos, A Seguin-Orlando, N Wales, L Orlando, SYW Ho, FS Dietrich, et al. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nature Communications*, 4, 2013.

[3] MD Martin, FG Vieira, SYW Ho, N Wales, M Schubert, A Seguin-Orlando, JB Ristaino, and MTP Gilbert. Genomic characterization of a South American *Phytophthora* hybrid mandates reassessment of the geographic origins of *Phytophthora infestans*. *Molecular Biology and Evolution*, 33(2):478–491, 2015.

[4] S Raffaele, RA Farrer, LM Cano, DJ Studholme, D MacLean, M Thines, RHY Jiang, MC Zody, SG Kunjeti, NM Donofrio, et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*, 330(6010):1540–1543, 2010.

[5] H Wickham. ggplot2: elegant graphics for data analysis. *Springer New York*, 1(2):3, 2009.

[6] K Yoshida, VJ Schuenemann, LM Cano, M Pais, B Mishra, R Sharma, C Lanz, FN Martin, S Kamoun, J Krause, et al. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife*, 2:e00731, 2013.
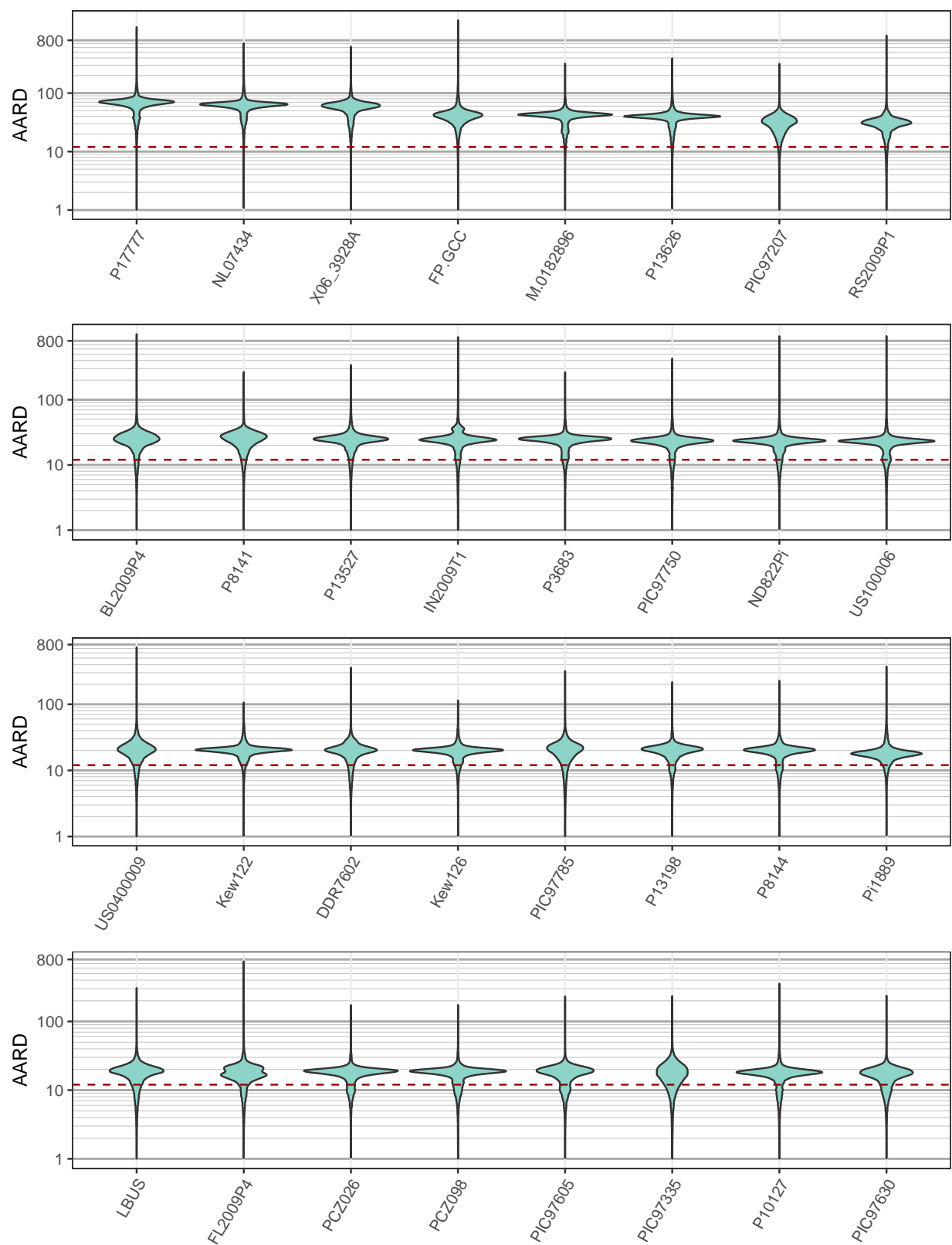
Supplemental item 2A. Violin plot of adjusted average read depth of for all *P. infestans* genes. The threshold for inclusion (12X) is indicated by a dashed red line.
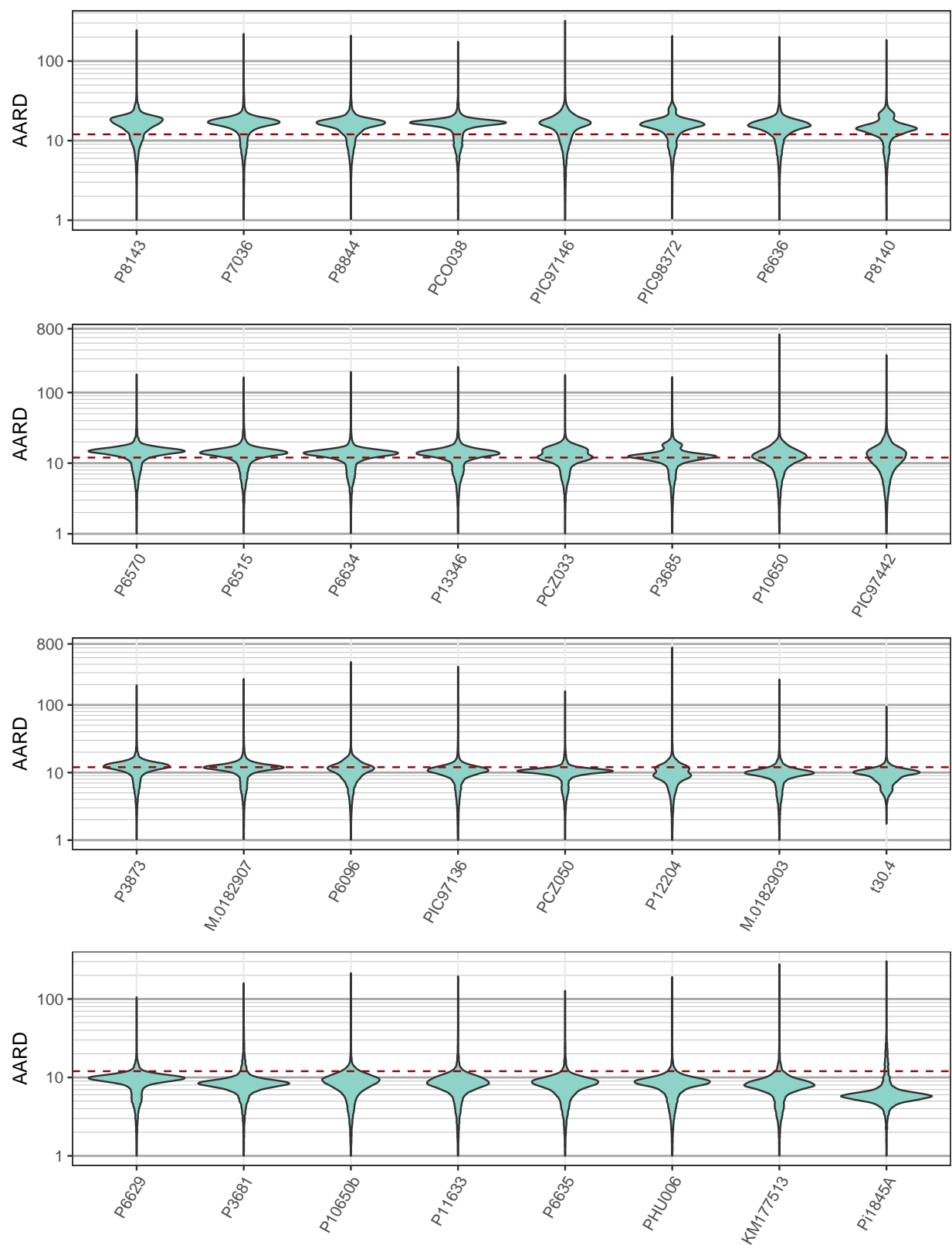
Supplemental item 2B. Violin plot of adjusted average read depth of for all *P. infestans* genes. The threshold for inclusion (12X) is indicated by a dashed red line.
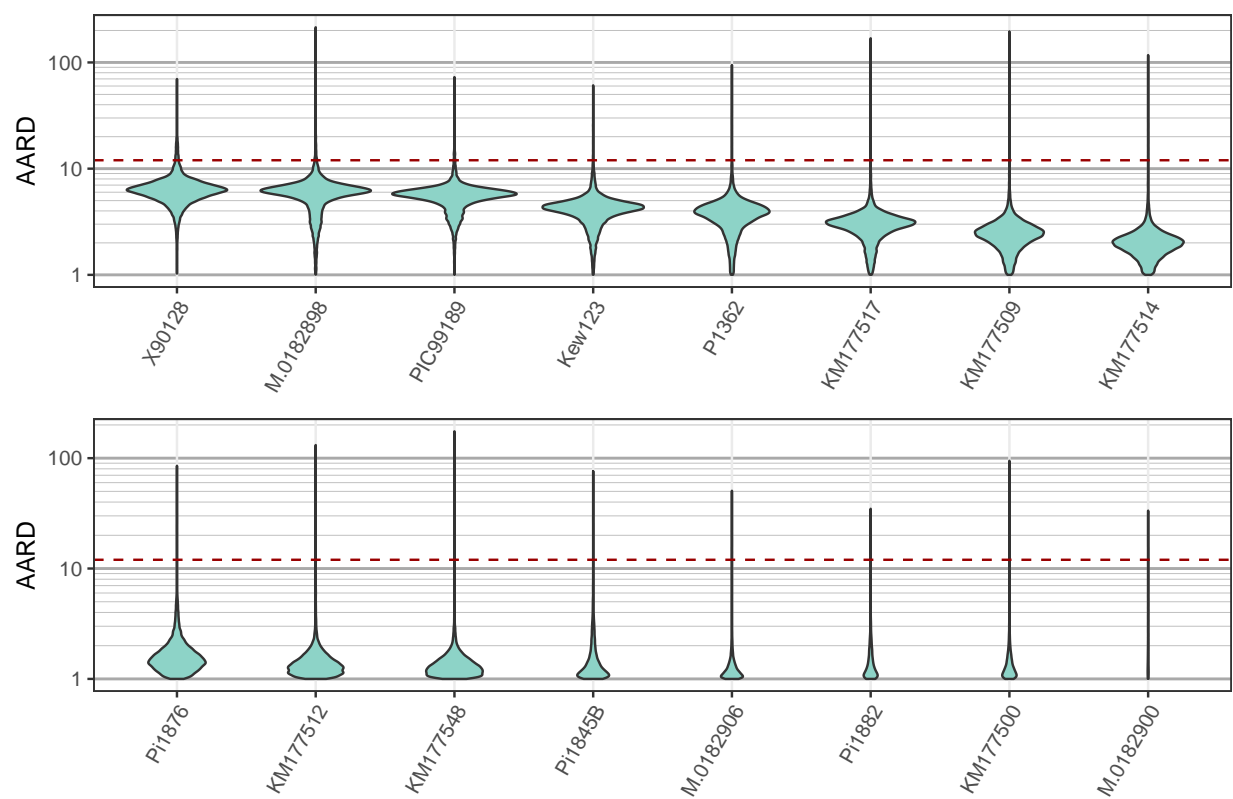
Supplemental item 2C. Violin plot of adjusted average read depth of for all *P. infestans* genes. The threshold for inclusion (12X) is indicated by a dashed red line.