Main Supplemental File.

Figure S1. Inter-individual variation in mean total genomic coverage and grouped by population. Total includes all autosomes, sex chromosomes, and the mitochondrial genome. Dashed lines represent means per population.

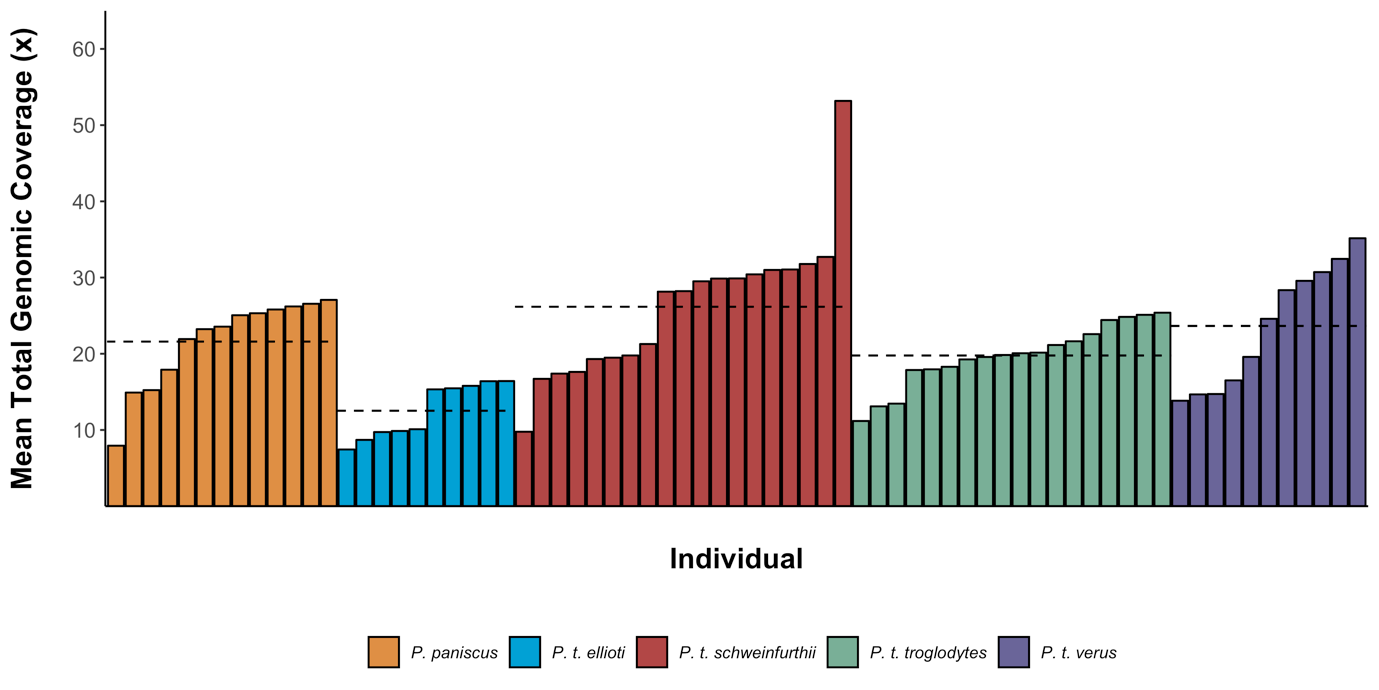


Figure S2. Inter-chromosomal variation in genomic coverage for all autosomes. Violin plots represent density at a particular coverage with the individual values plotted and jittered for each chromosome.

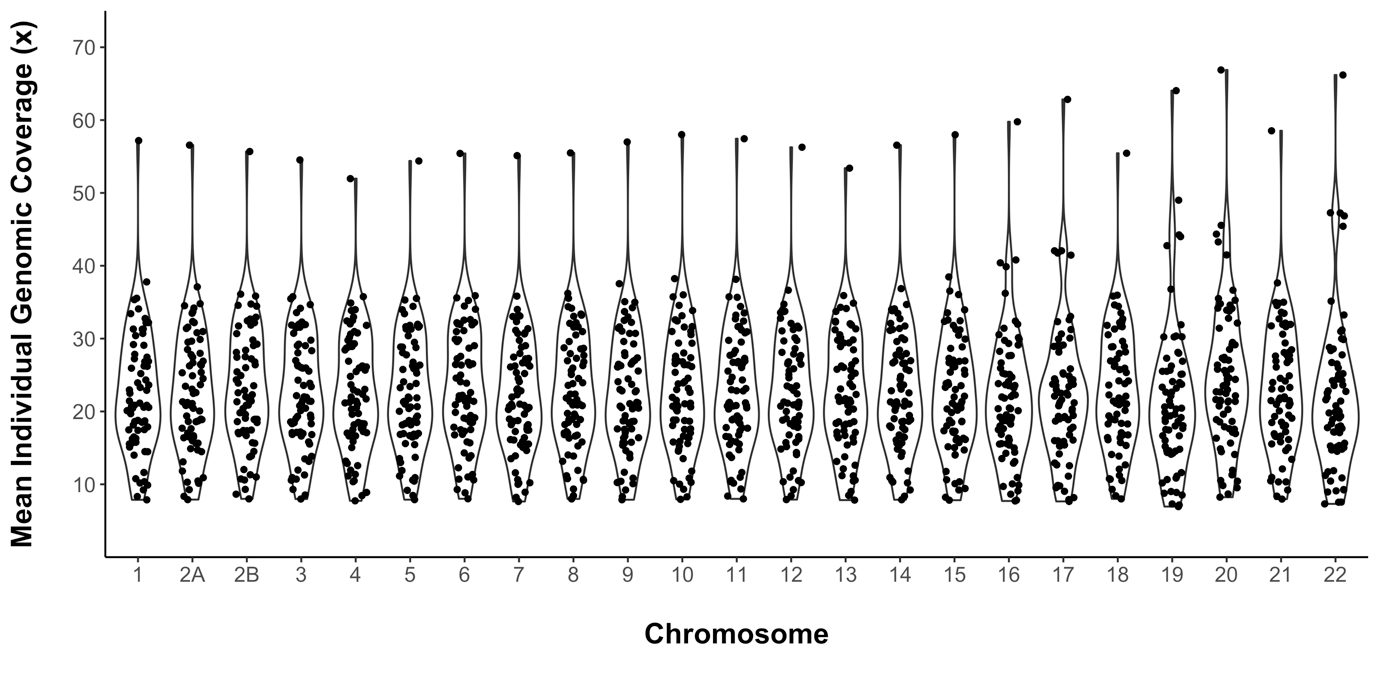


Figure S3. Number of snpEff annotations per SNP summed across all autosomes.

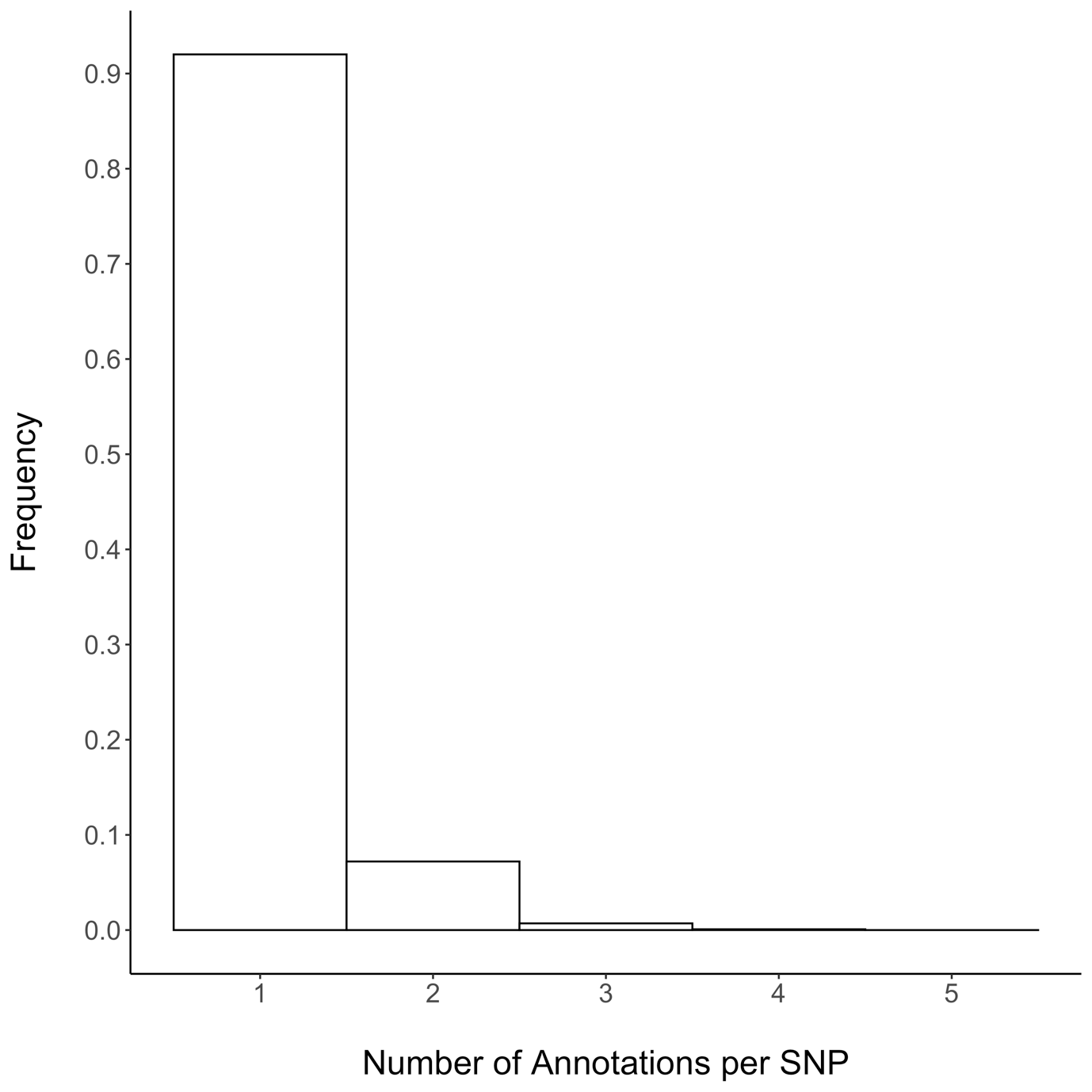


Figure S4. The rate of SNPs/bp per gene per lineage. The histogram represents the distribution of assessable genes and the dot plot displays the distribution of candidate genes under positive selection.

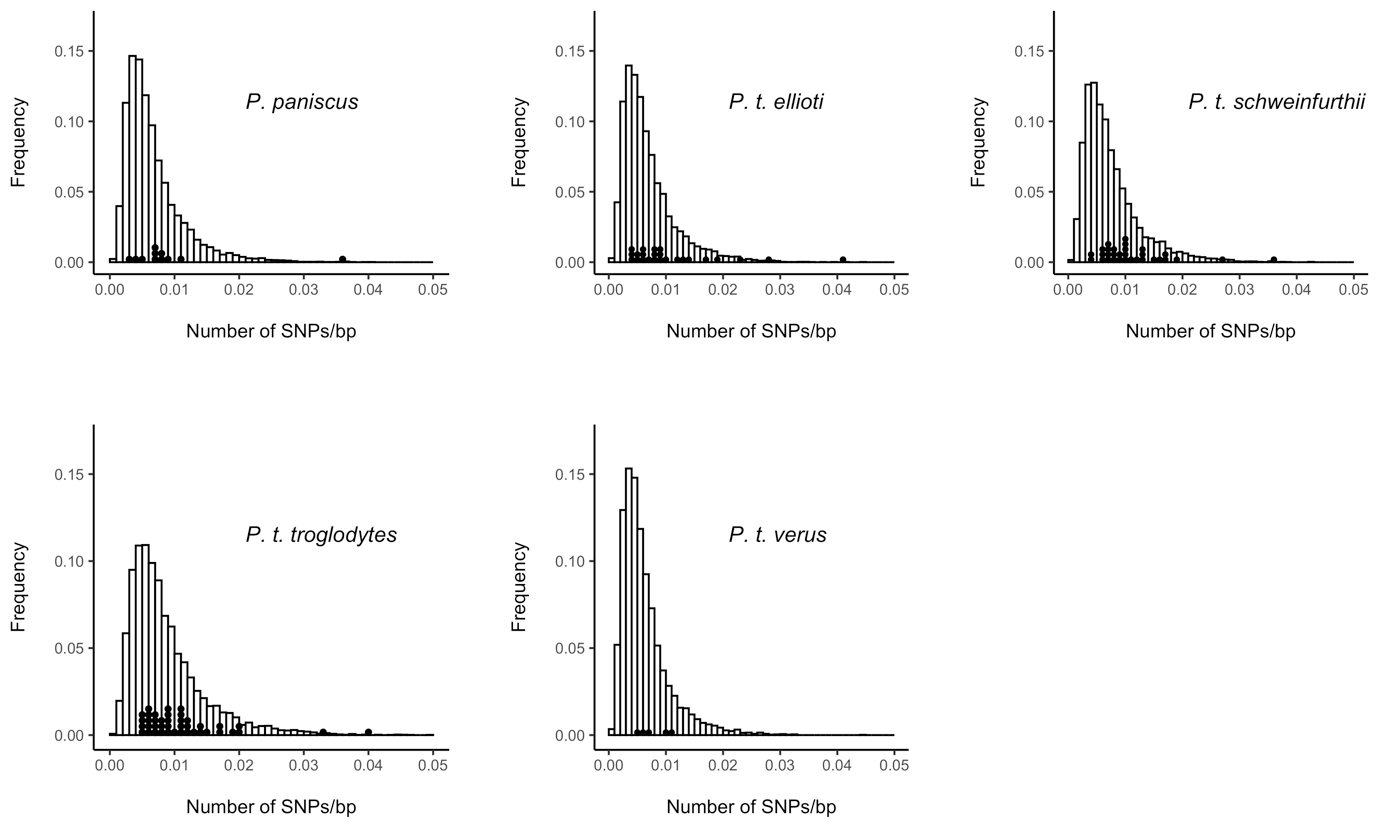


Figure S5. The rate of SNPs/bp per gene per lineage following MAF < 0.1 filtering. The histogram represents the distribution of assessable genes and the dot plot displays the distribution of candidate genes under positive selection.

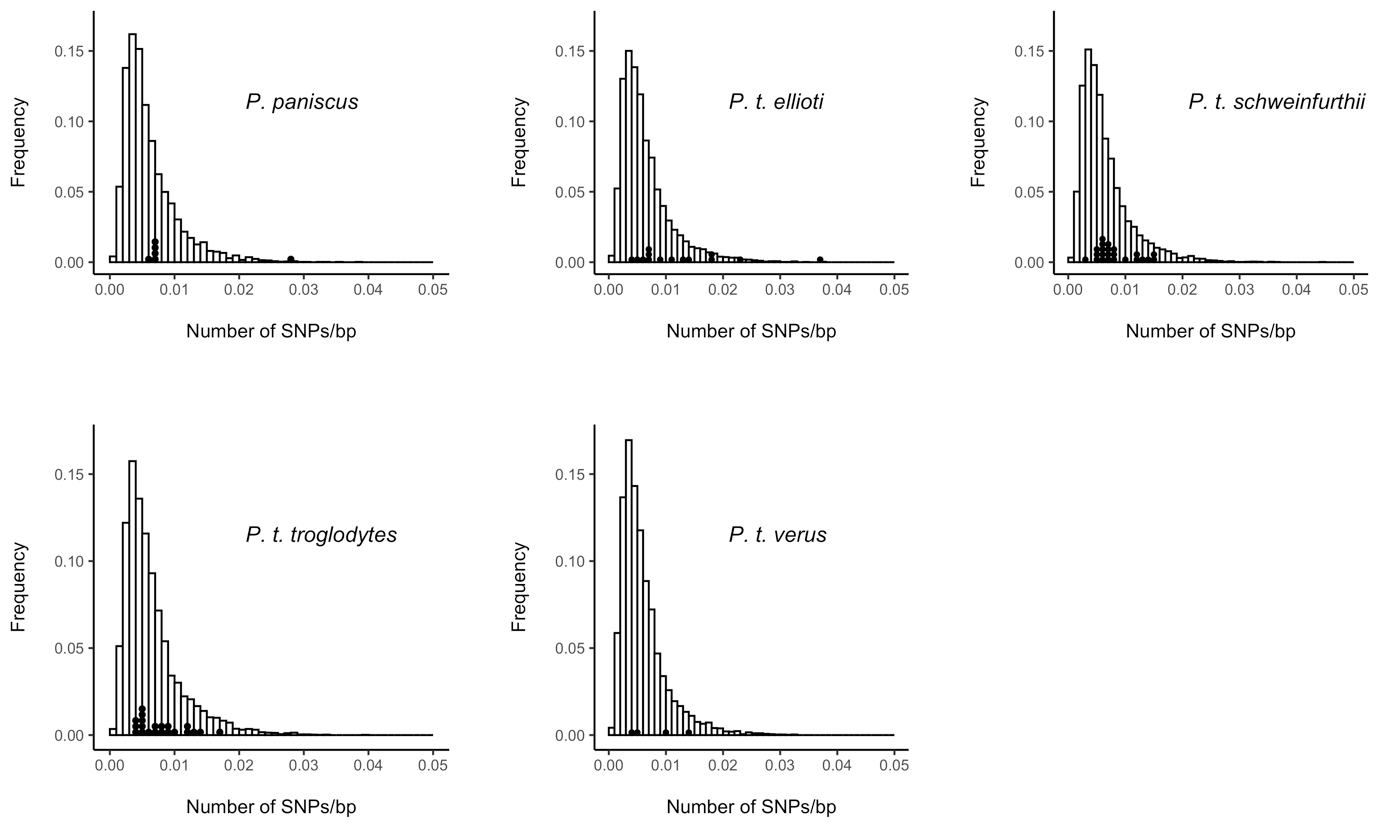


Figure S6. The rate of SNPs/bp per exon per lineage following MAF < 0.1 filtering. The histogram represents the distribution of assessable exon and the dot plot displays the distribution of candidate exons under positive selection.

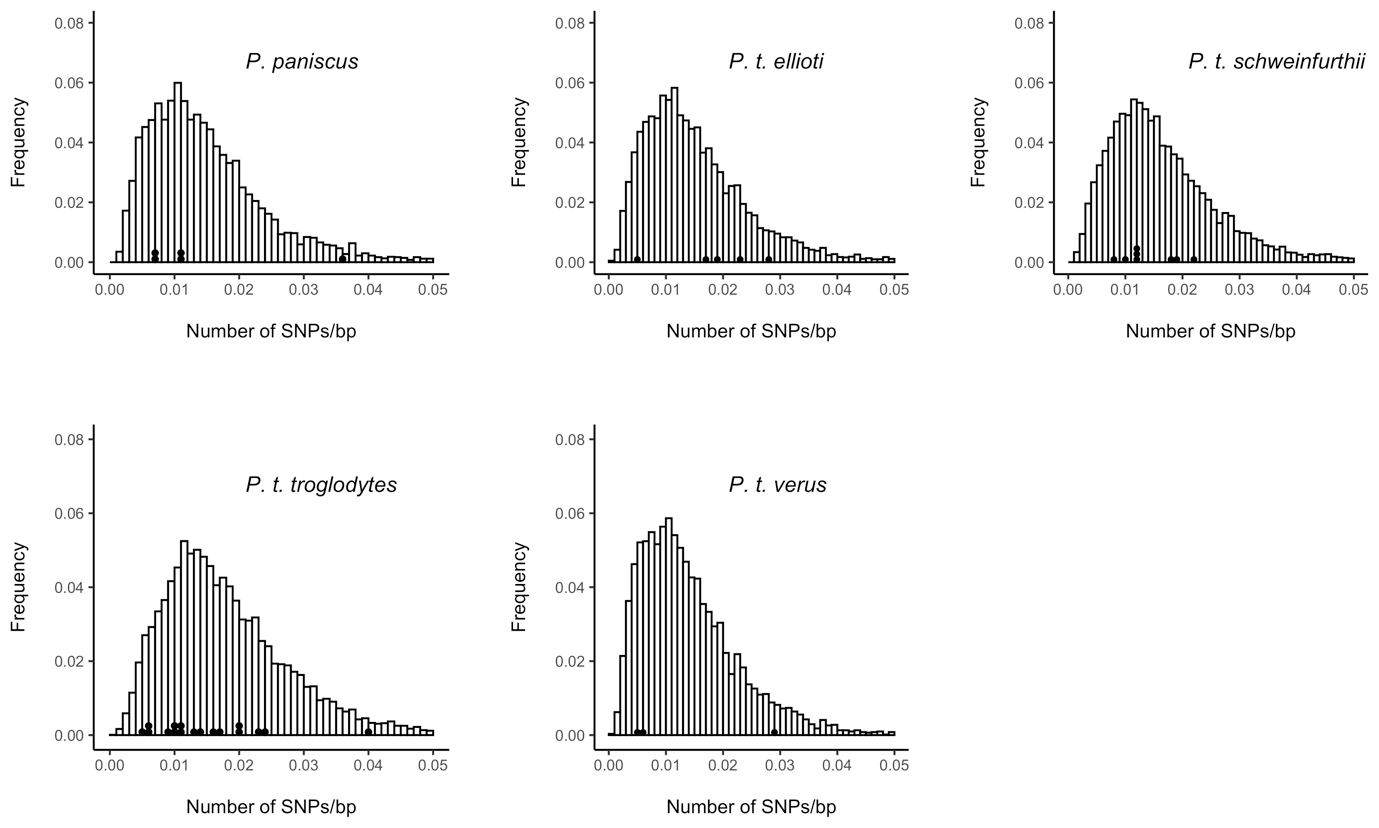


Figure S7. Upset plot of unique and shared candidate genes for positive selection without filtering SNPs with a minor allele frequency of < 0.1.

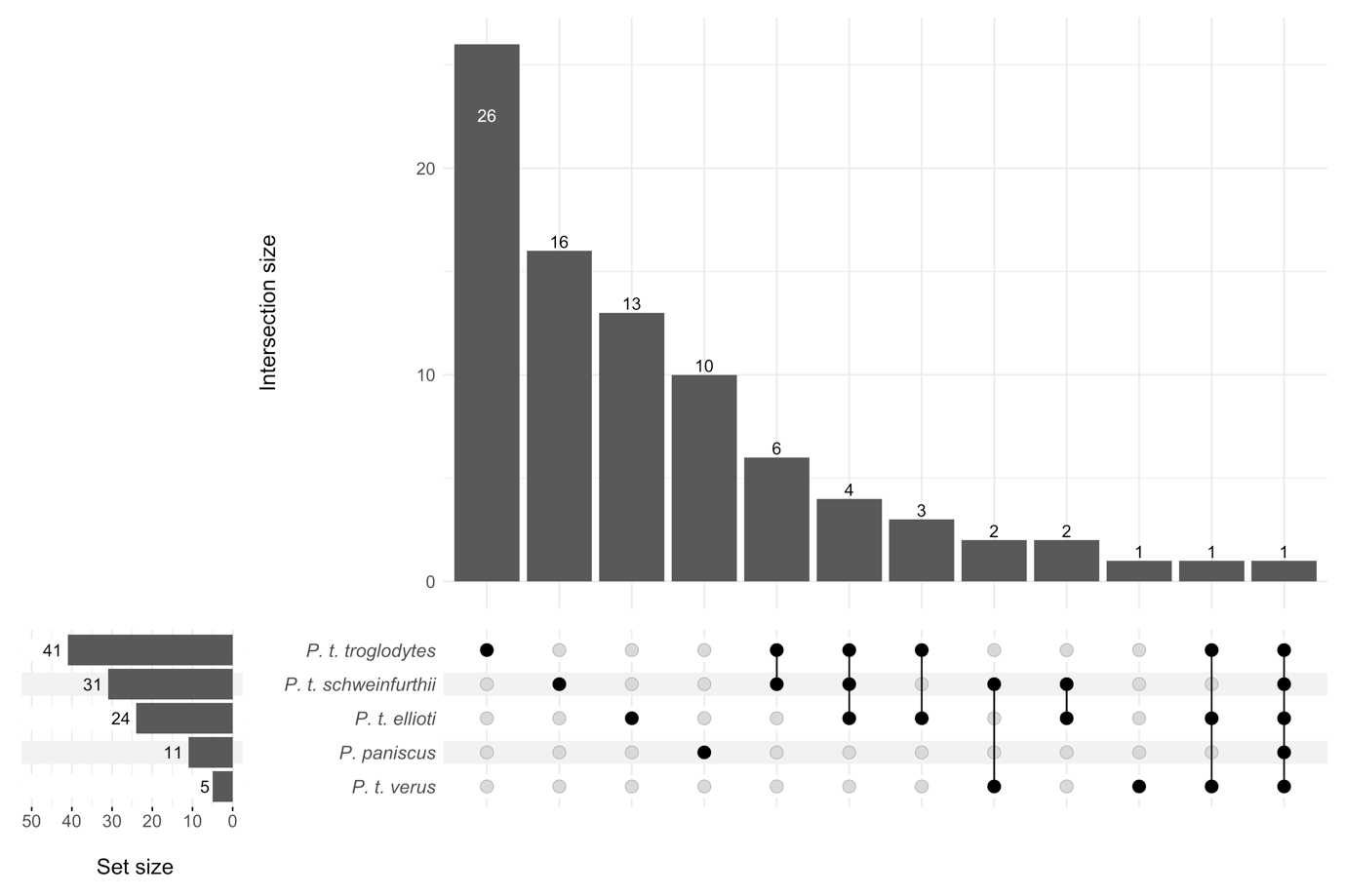


Figure S8. Upset plot of unique and shared candidate genes for positive selection after filtering SNPs with a minor allele frequency of < 0.1.

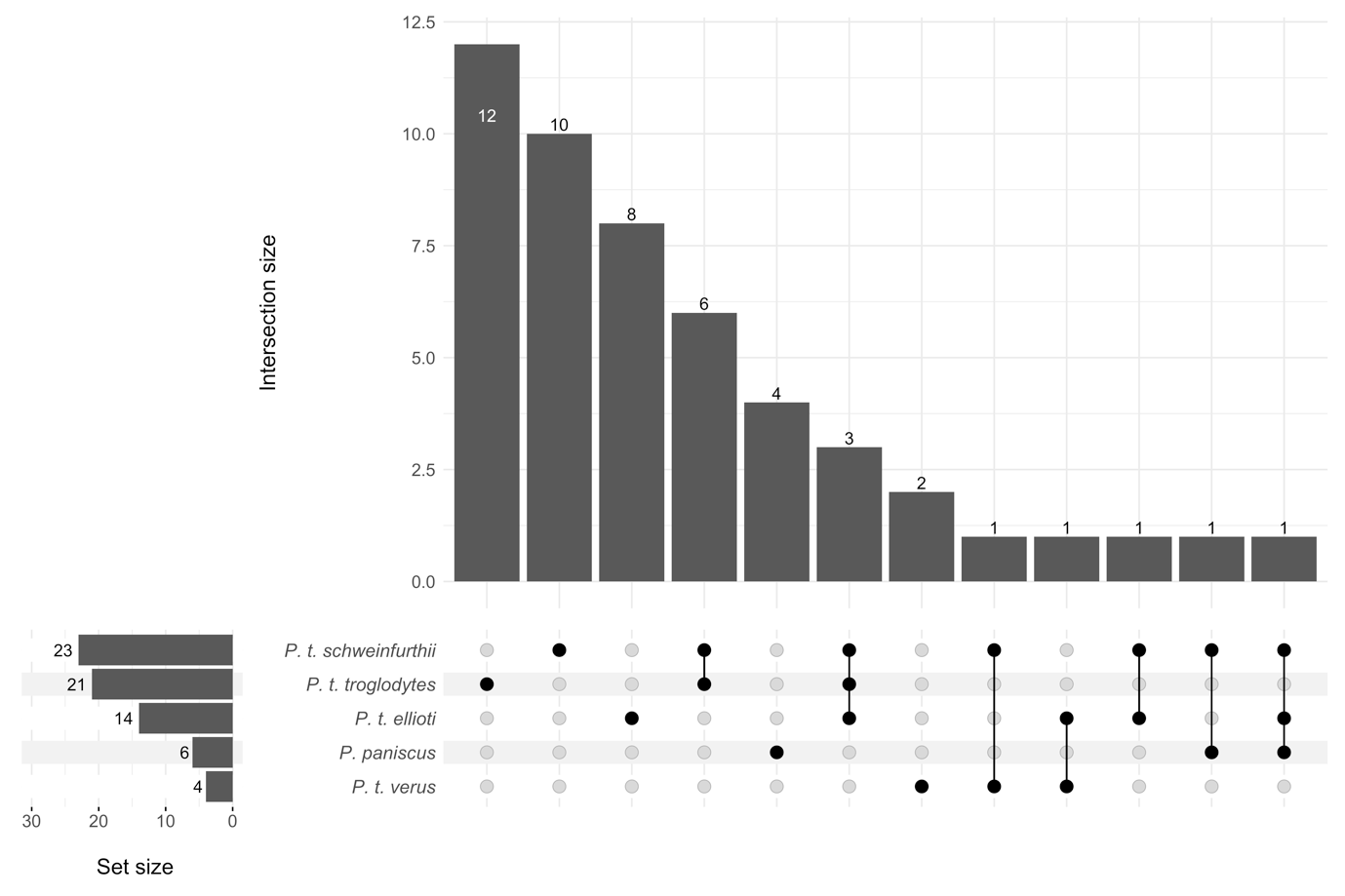


Figure S9. Plot of demographic strings used in simulated data. Demography data adapted from de Manuel et al. 2016. Data points are smoothed here using a spline function.

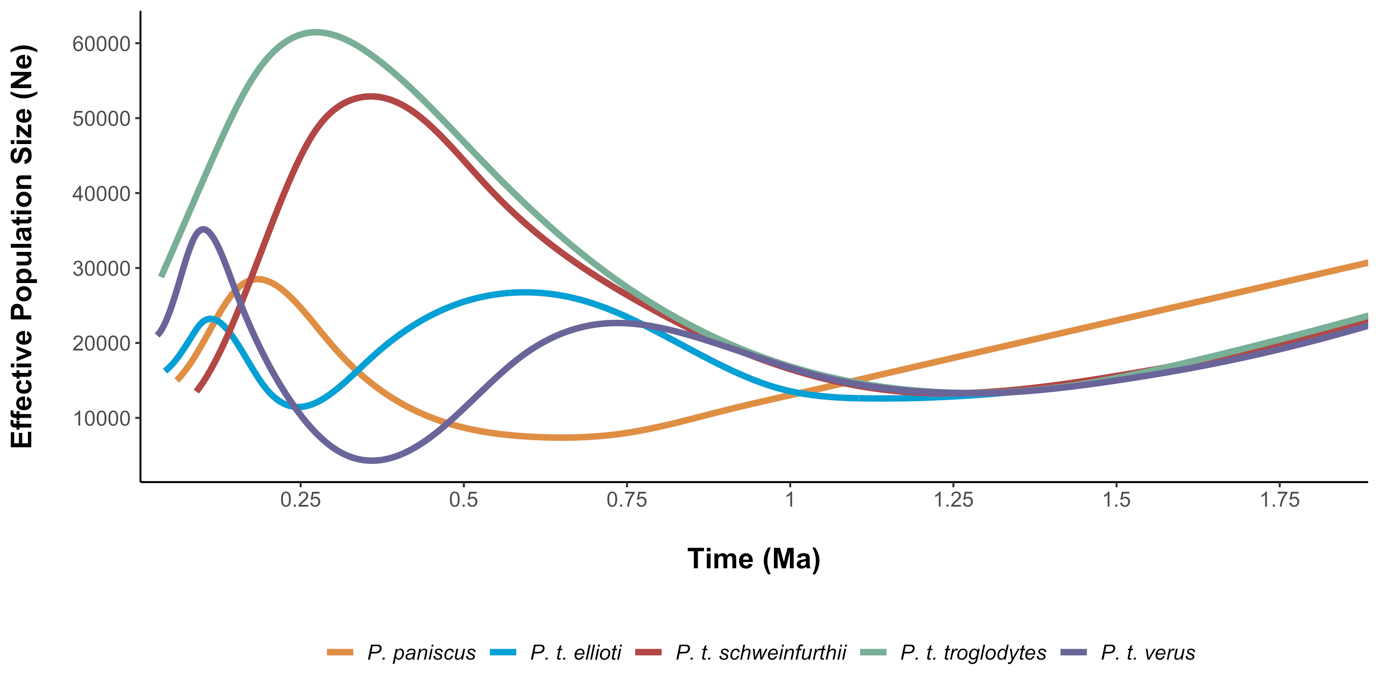


Figure S10. Correlations between parameters from the best fit model (δ) as determined by Legofit. delta = the admixture proportion, ec = ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*, nw = ancestor of *P. t. ellioti* and *P. t. verus*, ecnw = common ancestor of all *P. troglodytes* lineages, becnw = *Pan* common ancestor, T = time parameter, twoN = population size parameter.

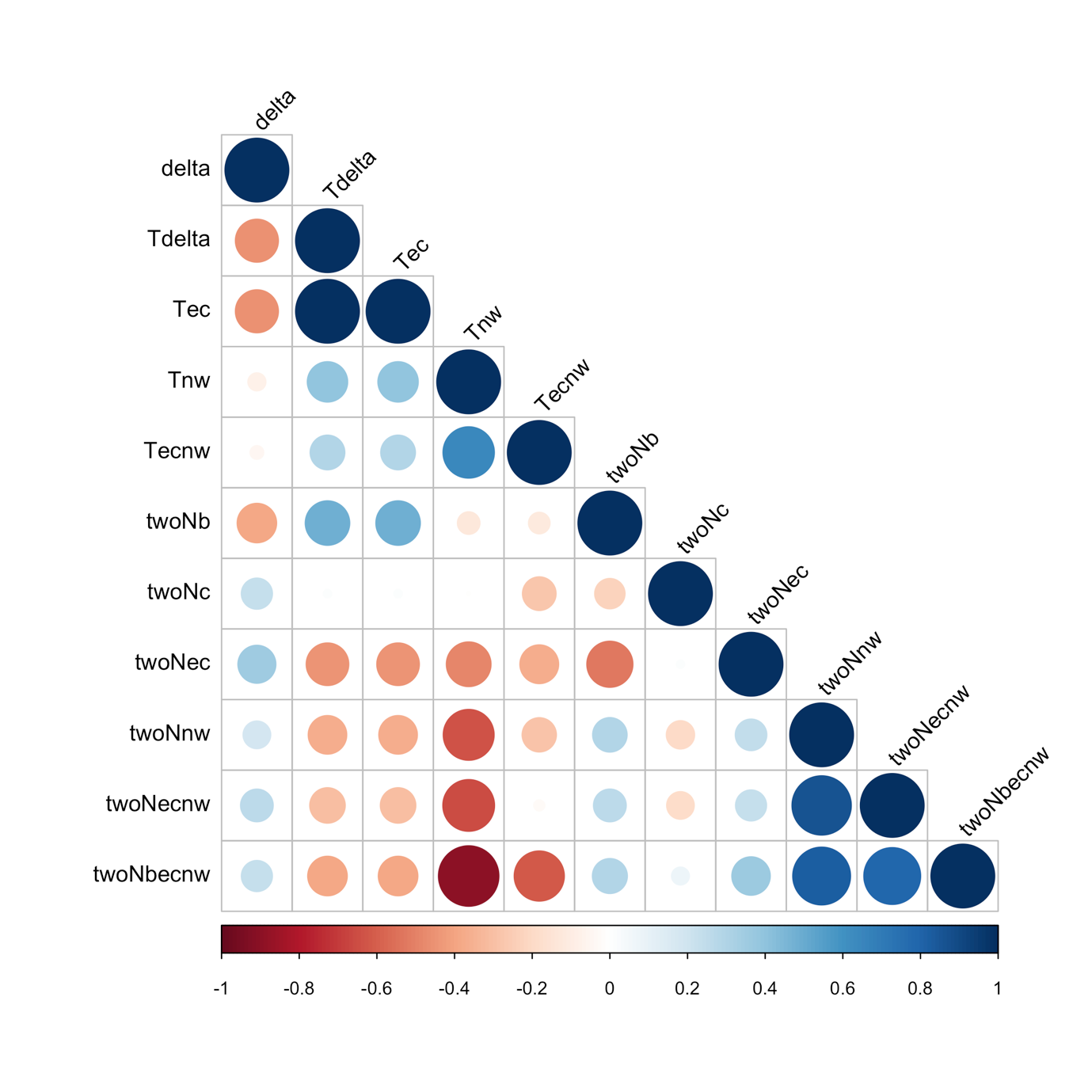


Table S1. Number of SNPs for MK analysis post filtering per chromosome.

|  |  |
| --- | --- |
| chromosome | SNPs post filtering (N) |
| 1 | 29,505 |
| 2A | 10,249 |
| 2B | 13,470 |
| 3 | 18,413 |
| 4 | 15,397 |
| 5 | 13,833 |
| 6 | 17,773 |
| 7 | 15,375 |
| 8 | 10,413 |
| 9 | 11,914 |
| 10 | 12,852 |
| 11 | 21,445 |
| 12 | 16,094 |
| 13 | 7,344 |
| 14 | 9,967 |
| 15 | 9,722 |
| 16 | 8,232 |
| 17 | 12,471 |
| 18 | 6,192 |
| 19 | 17,584 |
| 20 | 6,082 |
| 21 | 3,070 |
| 22 | 4,385 |

Table S2. Number of SNPs post filtering per chromosome for selective sweep analysis.

|  |  |
| --- | --- |
| chromosome | N SNPs post filtering |
| 1 | 3,332,313 |
| 2A | 1,774,559 |
| 2B | 2,024,437 |
| 3 | 3,156,661 |
| 4 | 3,086,706 |
| 5 | 2,585,495 |
| 6 | 2,741,379 |
| 7 | 2,451,758 |
| 8 | 2,415,460 |
| 9 | 1,815,711 |
| 10 | 2,086,543 |
| 11 | 2,056,912 |
| 12 | 2,024,301 |
| 13 | 1,570,832 |
| 14 | 1,373,336 |
| 15 | 1,246,426 |
| 16 | 1,202,275 |
| 17 | 1,016,097 |
| 18 | 1,254,126 |
| 19 | 687,464 |
| 20 | 971,965 |
| 21 | 560,810 |
| 22 | 434,326 |
| Total | 41,869,892 |

Table S3. Simulated vs predicted sweep type classes for *P. paniscus* classifier on a second, independent set of simulations with identical parameters. We generated 2 x 103 simulations for each class. Shaded cells reflect accuracy as a binary classifier: positive if hard or soft sweep, negative if linked sweep or neutral. Green = true positive, blue = false negative, purple = false positive, red = true negative. The false positive rate for this classifier is 0.043. Accuracy for identifying sweeps vs non-sweeps is 0.958 and 0.848 for correctly identifying the specific class.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Predicted Class** | | | | |
| Hard | Linked-hard | Neutral | Linked-soft | Soft |
| **Simulated Class** | Hard | 1881 | 26 | 0 | 0 | 93 |
| Linked-hard | 65 | 1877 | 0 | 57 | 1 |
| Neutral | 0 | 0 | 1676 | 217 | 107 |
| Linked-soft | 6 | 294 | 220 | 1403 | 77 |
| Soft | 220 | 4 | 100 | 33 | 1643 |

Table S4. Simulated vs predicted sweep type classes for *P. t. ellioti* classifier on a second, independent set of simulations with identical parameters. We generated 2 x 103 simulations for each class. Shaded cells reflect accuracy as a binary classifier: positive if hard or soft sweep, negative if linked sweep or neutral. Green = true positive, blue = false negative, purple = false positive, red = true negative. The false positive rate for this classifier is 0.042. Accuracy for identifying sweeps vs non-sweeps is 0.941 and 0.816 for correctly identifying the specific class.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Predicted Class** | | | | |
| Hard | Linked-hard | Neutral | Linked-soft | Soft |
| **Simulated Class** | Hard | 1847 | 69 | 0 | 3 | 81 |
| Linked-hard | 58 | 1911 | 0 | 30 | 1 |
| Neutral | 0 | 0 | 1611 | 278 | 111 |
| Linked-soft | 6 | 279 | 395 | 1243 | 77 |
| Soft | 186 | 2 | 196 | 69 | 1547 |

Table S5. Simulated vs predicted sweep type classes for *P. t. schweinfurthii* classifier on a second, independent set of simulations with identical parameters. We generated 2 x 103 simulations for each class. Shaded cells reflect accuracy as a binary classifier: positive if hard or soft sweep, negative if linked sweep or neutral. Green = true positive, blue = false negative, purple = false positive, red = true negative. The false positive rate for this classifier is 0.031. Accuracy for identifying sweeps vs non-sweeps is 0.966 and 0.862 for correctly identifying the specific class.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Predicted Class** | | | | |
| Hard | Linked-hard | Neutral | Linked-soft | Soft |
| **Simulated Class** | Hard | 1911 | 43 | 0 | 0 | 46 |
| Linked-hard | 15 | 1904 | 0 | 81 | 0 |
| Neutral | 0 | 0 | 1360 | 522 | 118 |
| Linked-soft | 1 | 136 | 114 | 1700 | 49 |
| Soft | 140 | 1 | 68 | 43 | 1748 |

Table S6. Simulated vs predicted sweep type classes for *P. t. troglodytes* classifier on a second, independent set of simulations with identical parameters. We generated 2 x 103 simulations for each class. Shaded cells reflect accuracy as a binary classifier: positive if hard or soft sweep, negative if linked sweep or neutral. Green = true positive, blue = false negative, purple = false positive, red = true negative. The false positive rate for this classifier is 0.014. Accuracy for identifying sweeps vs non-sweeps is 0.983 and 0.921 for correctly identifying the specific class.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Predicted Class** | | | | |
| Hard | Linked-hard | Neutral | Linked-soft | Soft |
| **Simulated Class** | Hard | 1950 | 13 | 0 | 0 | 37 |
| Linked-hard | 20 | 1949 | 0 | 31 | 0 |
| Neutral | 0 | 0 | 1769 | 181 | 50 |
| Linked-soft | 0 | 155 | 100 | 1731 | 14 |
| Soft | 123 | 1 | 49 | 19 | 1808 |