# 1 Text S7. Allele balance: non-*P. infestans* samples

Allele balance, the frequency at which each allele was sequenced, was calculated for each available sample. A genomic VCF file (g.VCF) was created for each sample using GATK [4, 1]. Each g.VCF file was read into R[5] for processing with vcfR [2]. The allele depth (AD) and genotypes were extracted from the vcfR object and the first and second most abundant alleles were extracted from the allele depth. The genotype information was used to subset the allele depth information to only heterozygous positions. The 15th and 85th percentiles were calculated for each sample and each alelle (the first and second most abundant alleles) and used as an inclusion threshold for depth filtering. A frequency for each allele was then calculated by dividing its allele depth by the sum of the first and second most abundant alleles. This information was used to plot each histogram. These methods are also described in Knaus and Grünwald [3].
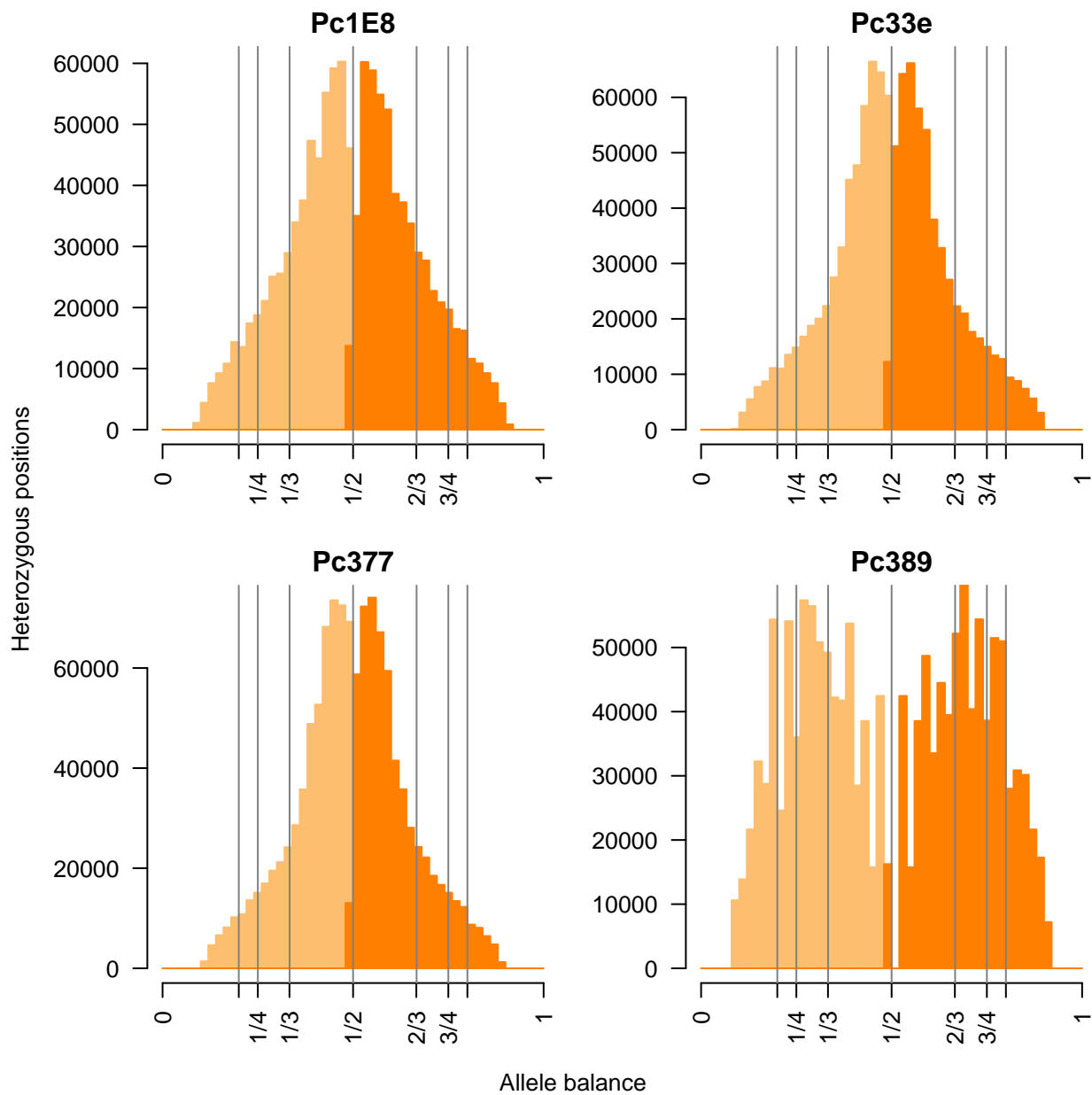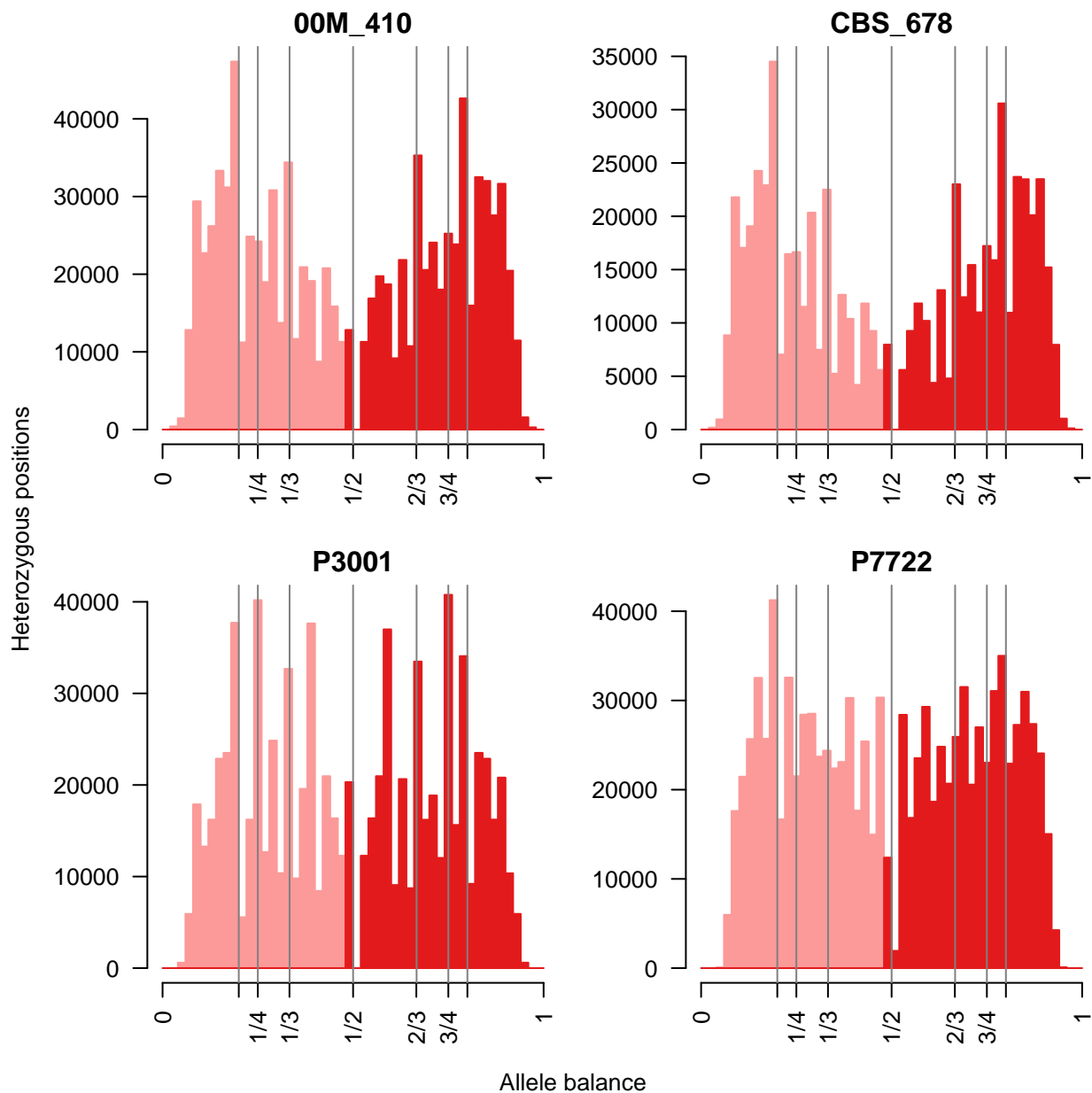
Supplemental item 4A Histograms of allele balance for *P. capsici*. The expectations for pentaploid (1/5, 4/5), tetraploid (1/4, 3/4), triploid (1/3, 2/3), and diploid (1/2) heterozygote sequenced allele frequency is indicated with vertical lines.
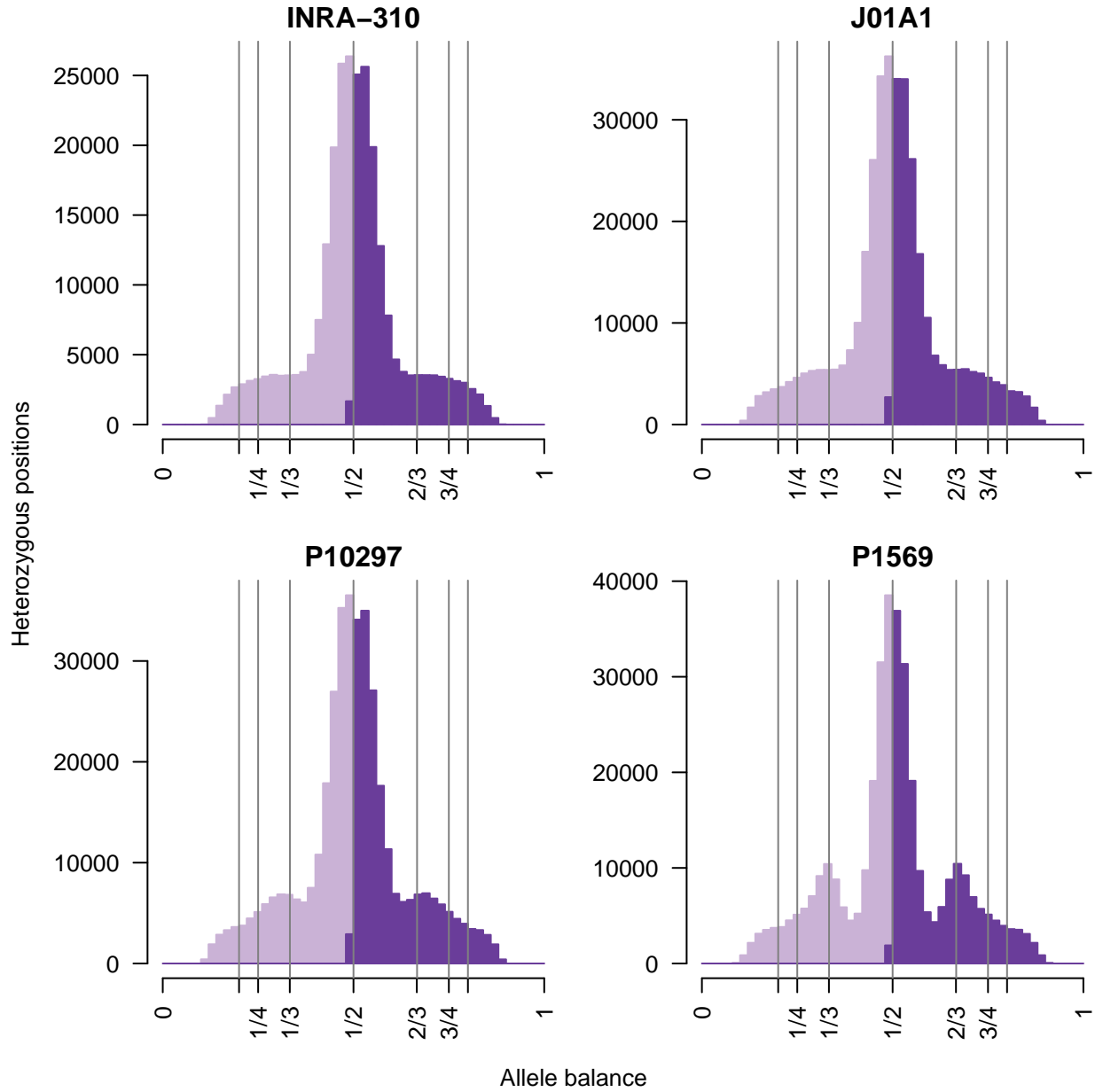
Supplemental item 4B. Histograms of allele balance for *P. parasitica*. The expectations for pentaploid (1/5, 4/5), tetraploid (1/4, 3/4), triploid (1/3, 2/3), and diploid (1/2) heterozygote sequenced allele frequency is indicated with vertical lines.

P1976
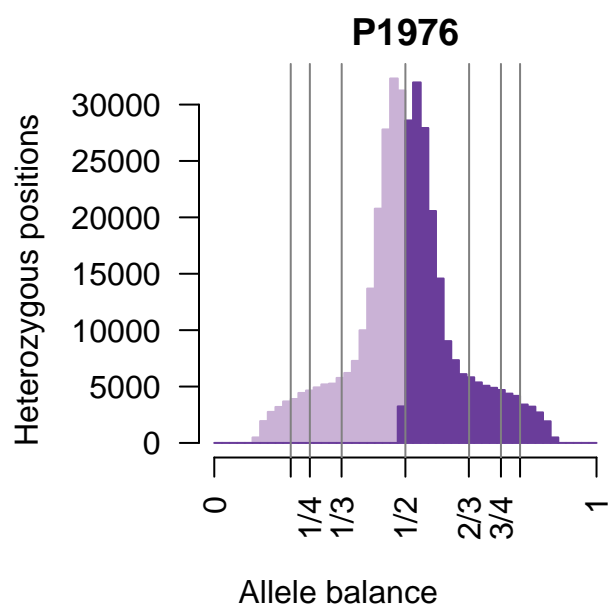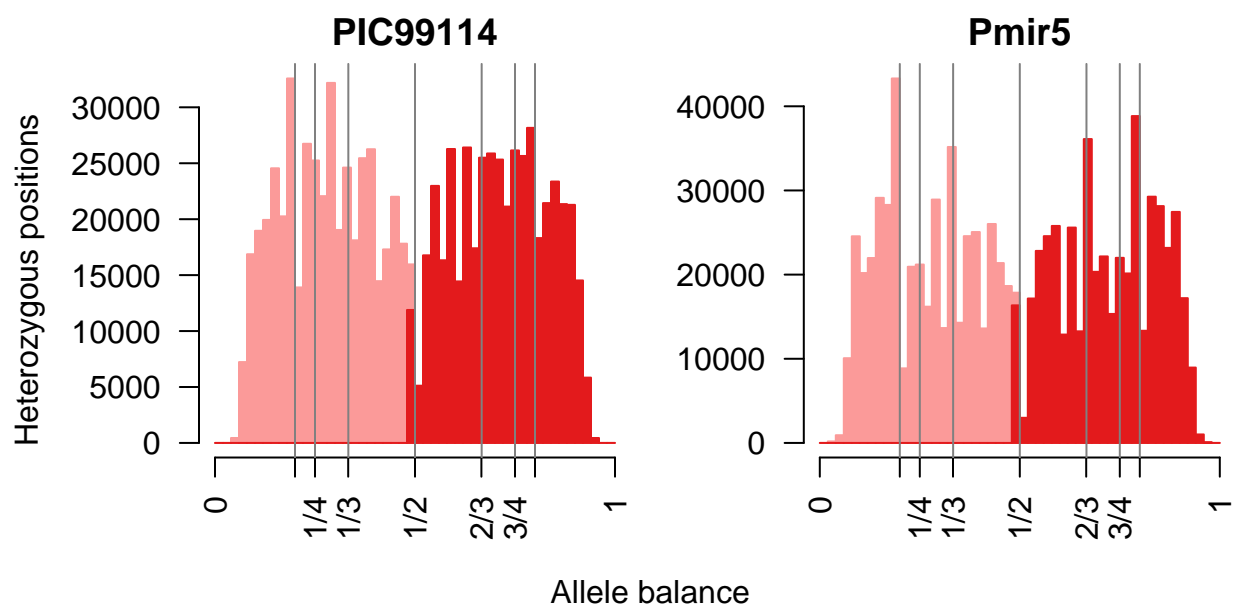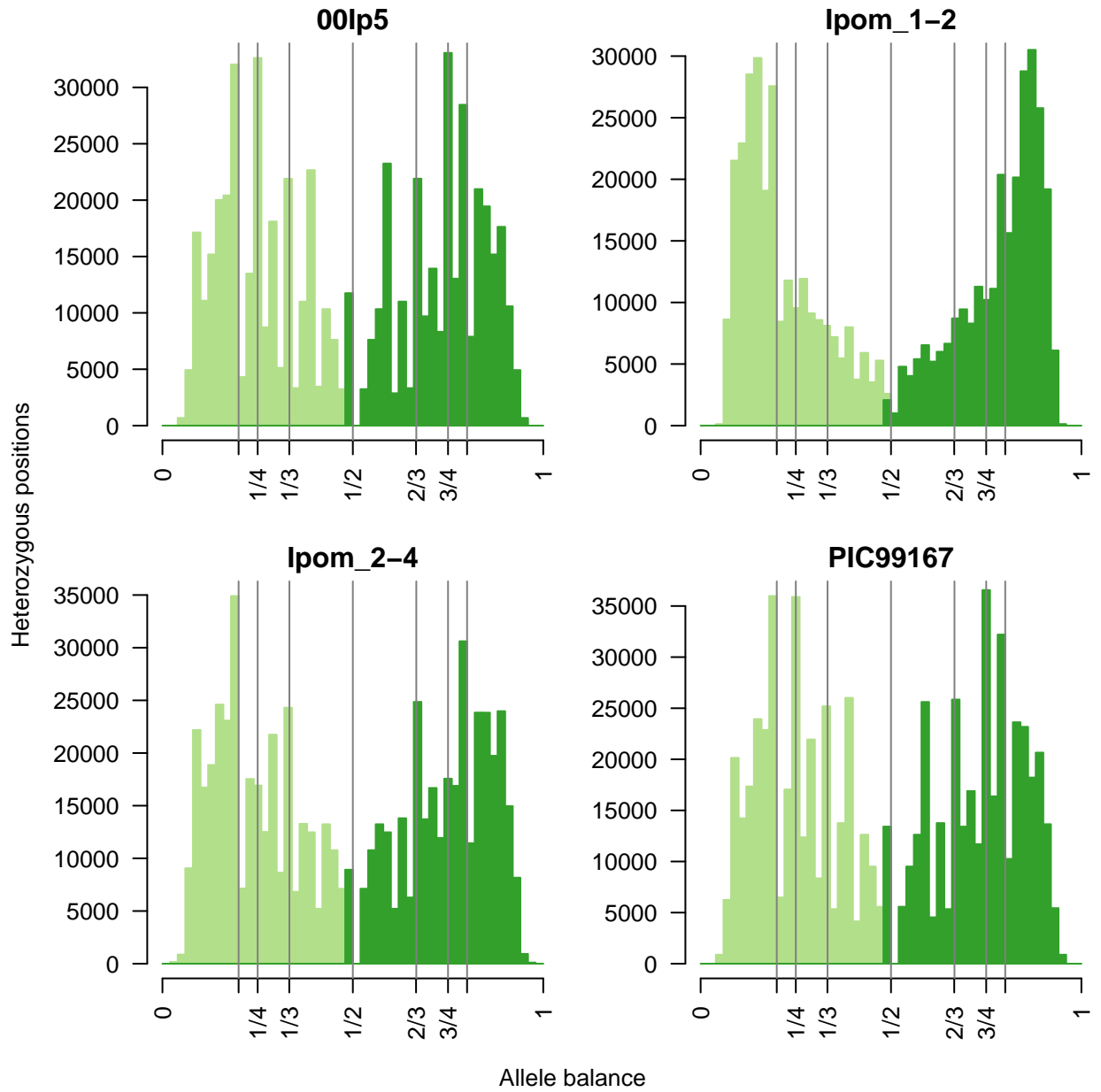
Supplemental item 4C. Histogram of allele balance for *P. parasitica*. The expectations for pentaploid (1/5, 4/5), tetraploid (1/4, 3/4), triploid (1/3, 2/3), and diploid (1/2) heterozygote sequenced allele frequency is indicated with vertical lines.

Supplemental item 4D. Histograms of allele balance for *P. mirabilis*. The expectations for pentaploid (1/5, 4/5), tetraploid (1/4, 3/4), triploid (1/3, 2/3), and diploid (1/2) heterozygote sequenced allele frequency is indicated with vertical lines.
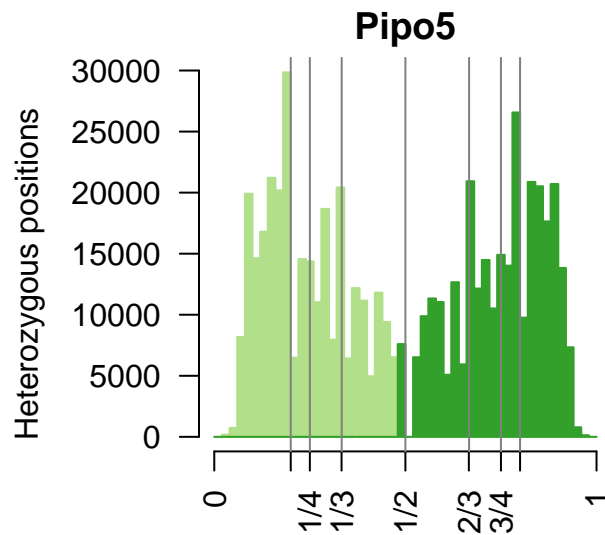
Supplemental item 4E. Histograms of allele balance for *P. mirabilis*. The expectations for pentaploid (1/5, 4/5), tetraploid (1/4, 3/4), triploid (1/3, 2/3), and diploid (1/2) heterozygote sequenced allele frequency is indicated with vertical lines.

Supplemental item 4F. Histograms of allele balance for *P. ipomoeae*. The expectations for pentaploid (1/5, 4/5), tetraploid (1/4, 3/4), triploid (1/3, 2/3), and diploid (1/2) heterozygote sequenced allele frequency is indicated with vertical lines.

Pipo5
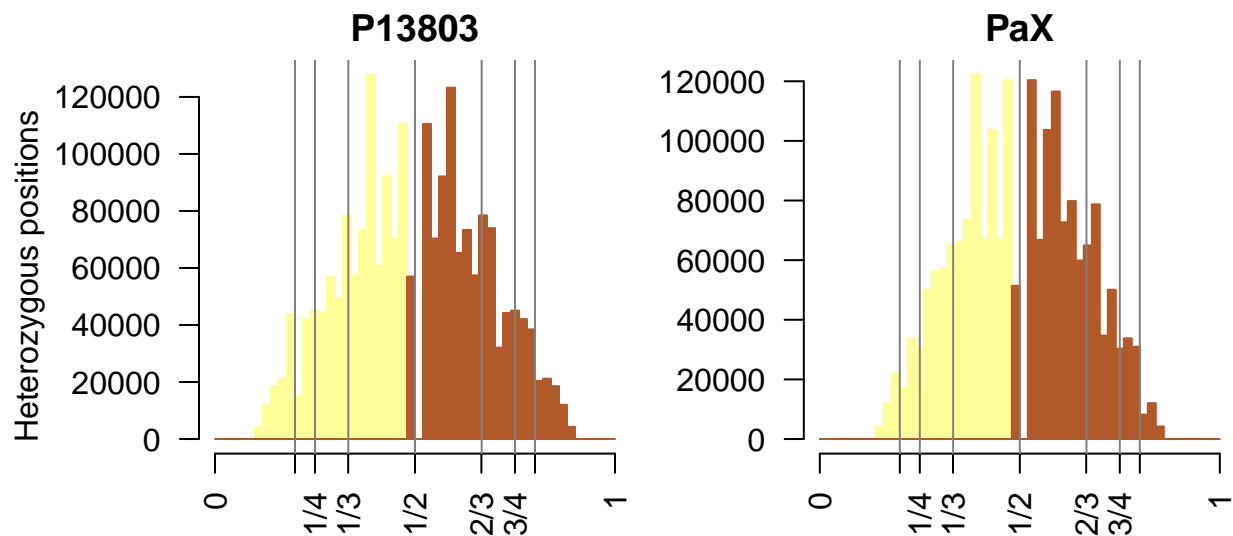
Supplemental item 4G. Histograms of allele balance for *P. ipomoeae*. The expectations for pentaploid (1/5, 4/5), tetraploid (1/4, 3/4), triploid (1/3, 2/3), and diploid (1/2) heterozygote sequenced allele frequency is indicated with vertical lines.

P13803   **PaX**

Supplemental item 4H. Histograms of allele balance for *P. andina.* The expectations for pentaploid (1/5, 4/5), tetraploid (1/4, 3/4), triploid (1/3, 2/3), and diploid (1/2) heterozygote sequenced allele frequency is indicated with vertical lines.

References

[1] MA DePristo, E Banks, R Poplin, KV Garimella, JR Maguire, C Hartl, AA Philippakis, G Del Angel, MA Rivas, M Hanna, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498, 2011.

[2] BJ Knaus and NJ Grünwald. vcfR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1):44–53, 2017.

[3] BJ Knaus and NJ Grünwald. Inferring variation in copy number using high throughput sequencing data in R. *Frontiers in Genetics*, 9:123, 2018.

[4] A McKenna, M Hanna, E Banks, A Sivachenko, K Cibulskis, A Kernytsky, K Garimella, D Altshuler, S Gabriel, M Daly, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303, 2010.

[5] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.