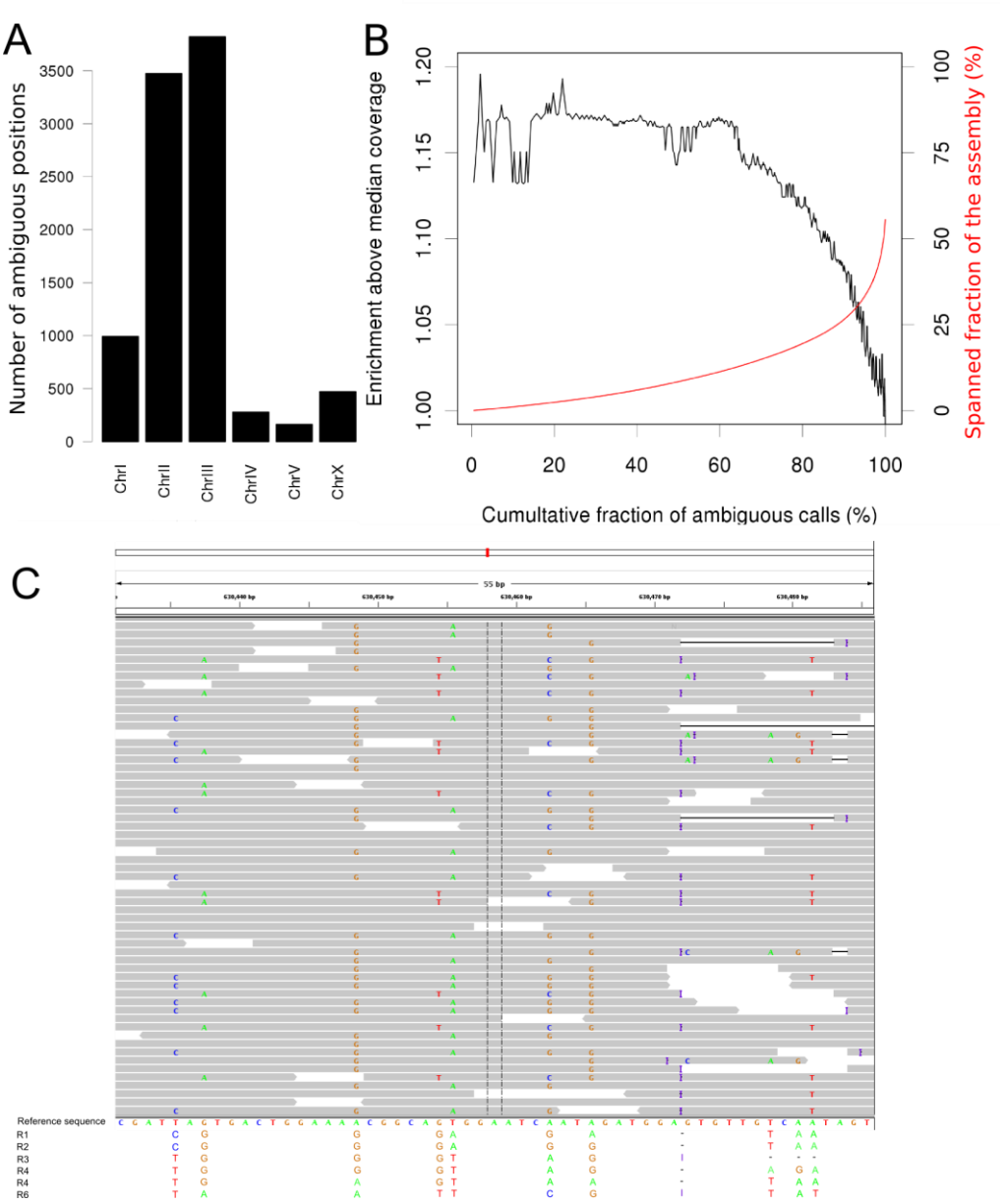

Figure S1 Heterozygosity in the *P. exspectus* assembly. **(A)** Ambiguous sites are present on autosomes and the X chromosome. Shown are the numbers of heterozygous calls per chromosome. *P. exspectatus.* Scaffolds were assigned to chromosomes by blasting *P. pacificus* genetic markers against the *P. exspectatus* assembly. **(B)** Ambiguous sites are associated with increased coverage. Non-overlapping windows of 10kb were ordered by decreasing number of ambiguous sites (x-axis) and the ratio of median coverage within the top X% above the genome-wide median coverage is plotted. At 60-70% of ambiguous sites, the ratio drops to again and reaches a value of one when approaching 100% of ambiguous sites. **(C)** Screenshot from the IGV browser showing realigned *P. exspectatus* reads. One position shows three different genotypes, which cannot be explained by remaining heterozygosity, indicating that reads must originate from highly similar paralogous regions. The alignments suggest that at least six different regions (R1-R6) are required to explain the observed pattern.

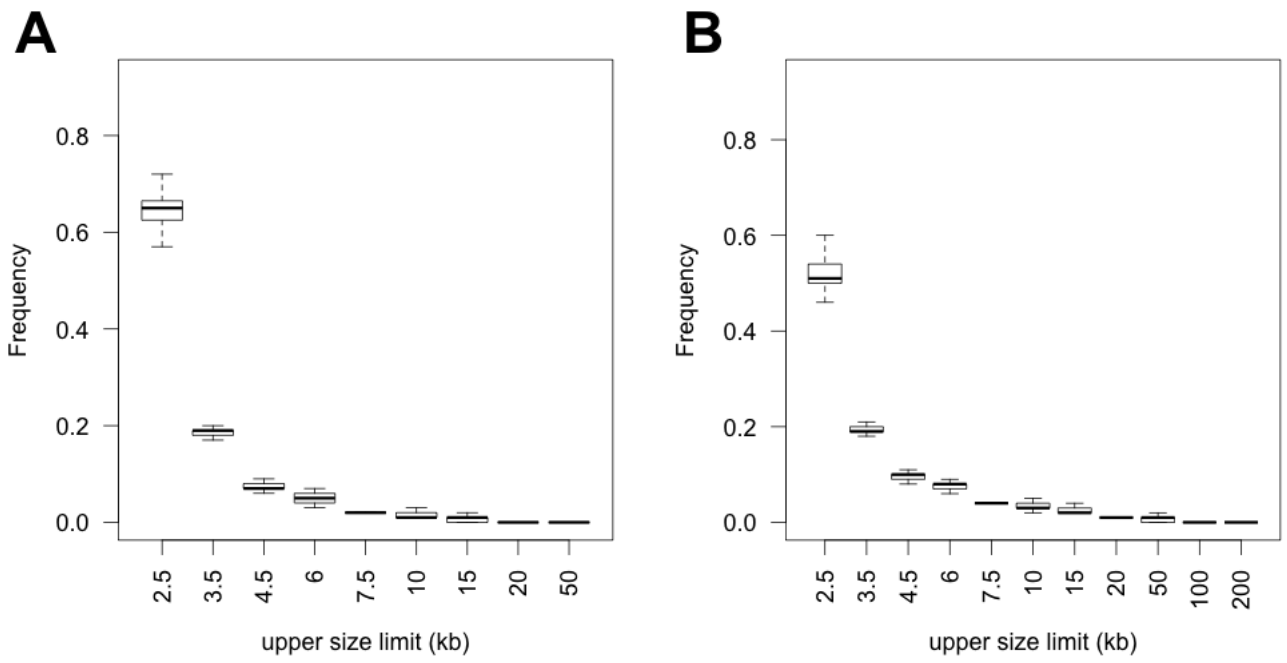

Figure S2 Size distribution of detected duplications and deletions. Duplications and deletions (≥ 2kb) were detected by comparing the read coverage in each of the 103 strains to the coverage of the resequenced reference strain (PS312) using the program cnv-seq (*P*<0.01). The boxplots show median and interquartile range of the size distribution of duplications **(A)** and deletions **(B)** across all 103 strains.

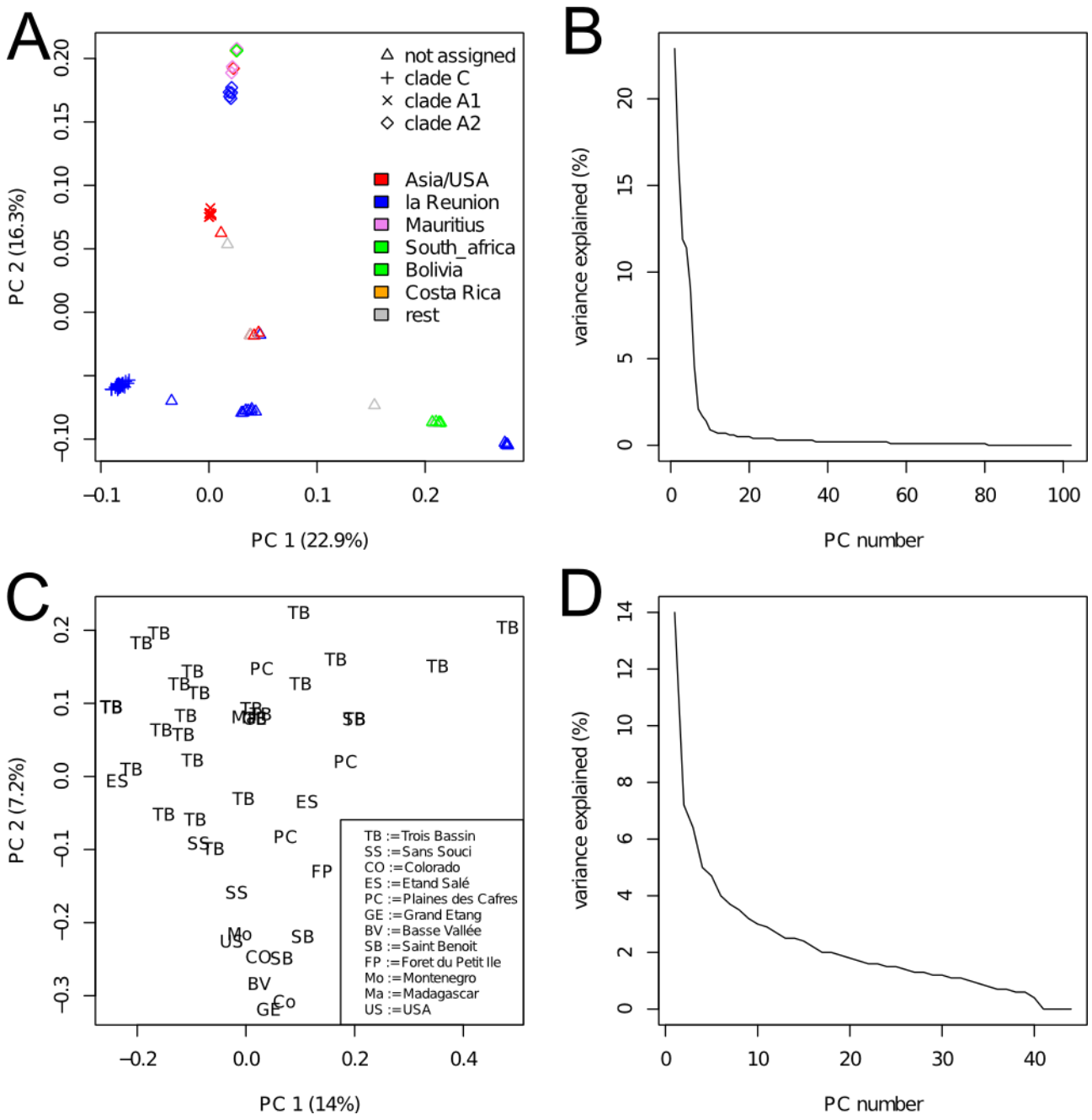

Figure S3 Principal component analysis of SNV data. In order to reduce bias by short range LD, principal component analysis was performed by EIGENSOFT 3.0, using one biallelic SNV with 5-95% allele frequency per 50kb window. **(A)** First two principal components (PC) for all strains. Both PCs were significant (*P<0.001*) according to Tracy-Widom statistics. **(B)** Variance, explained by the individual principal components. **(C)** First two PCs for strains from clade C (*P<0.001*, according to Tracy-Widom statistics). The first two PCs reveal a separation between strains sampled from Trois Bassins as opposed to most other locations. **(D)** Variance, explained by the individual PCs for clade C.

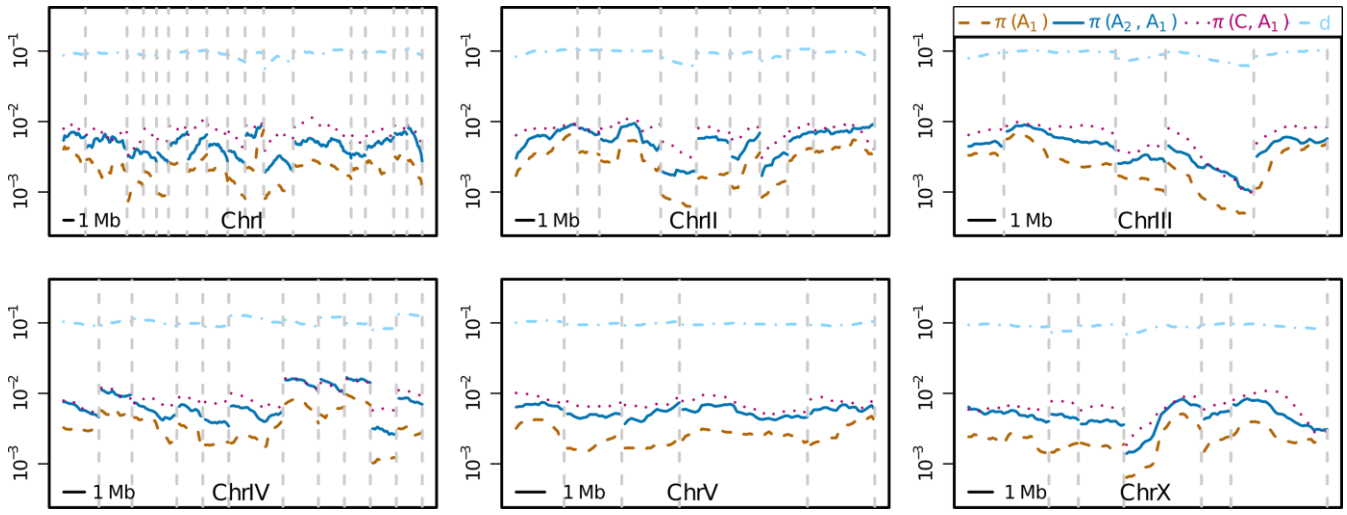

Figure S4 Clade A_1 diversity. Divergence (d) between *P. pacificus* and *P. exspectatus*, nucleotide diversity π within clade A_1 and between A_1 and clades C and A_2 are shown. π values represent averages over 1Mb windows. Supercontigs (>1Mb) with markers on the genetic map were concatenated to visualize the chromosomal distribution. The dashed lines denote supercontig boundaries with unknown physical distance.

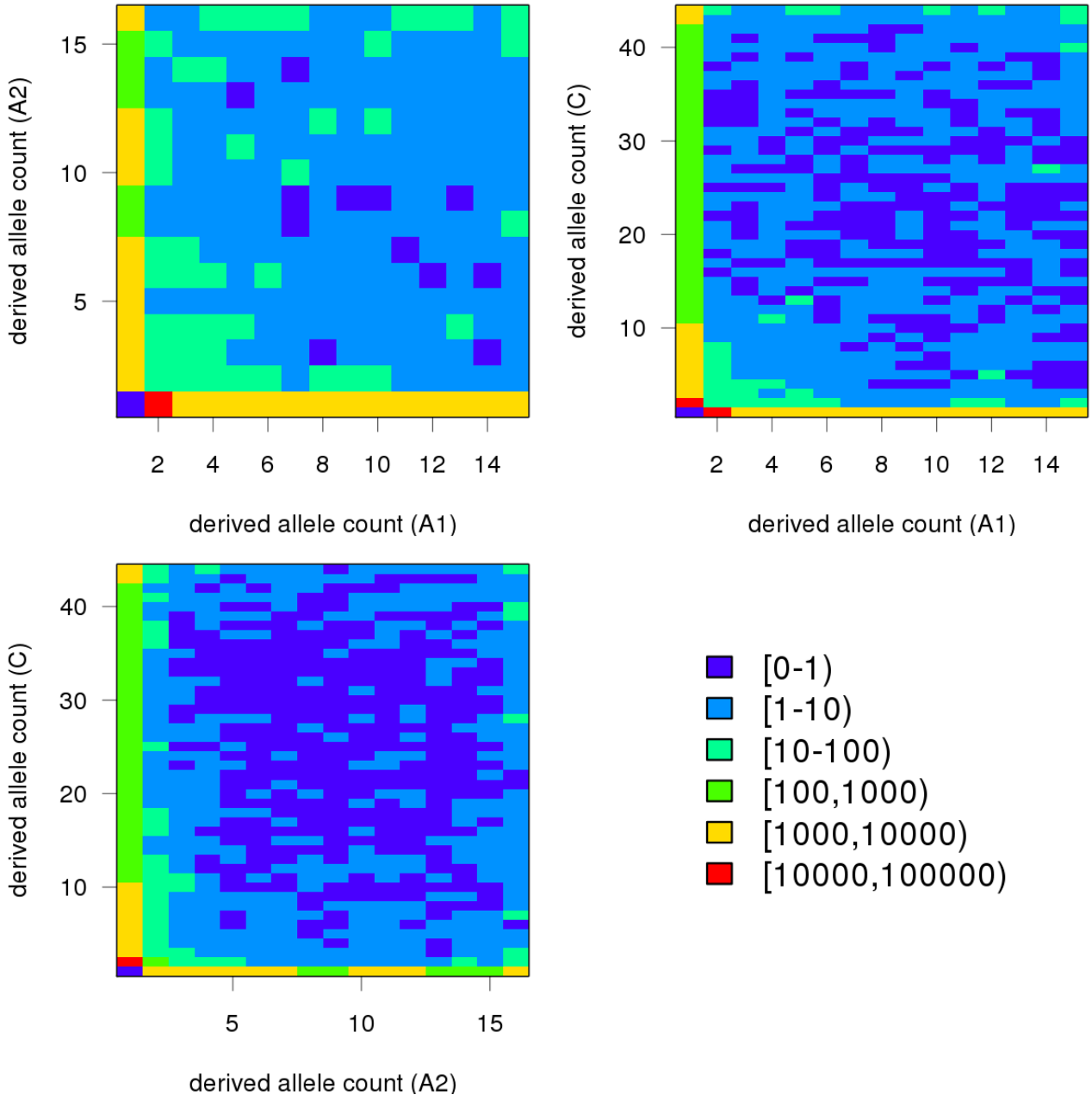

Figure S5 Joint site frequency spectra (SFS) for clades A1, A2, and C. Total numbers of derived allele counts are shown for all three pairwise comparisons of clades. The vast majority of variation is located along the x and y-axis indicating that most variation is clade-specific. Only a small percentage of total variation is shared between clades and may therefore represent ancestral diversity, introgression, or convergent evolution.

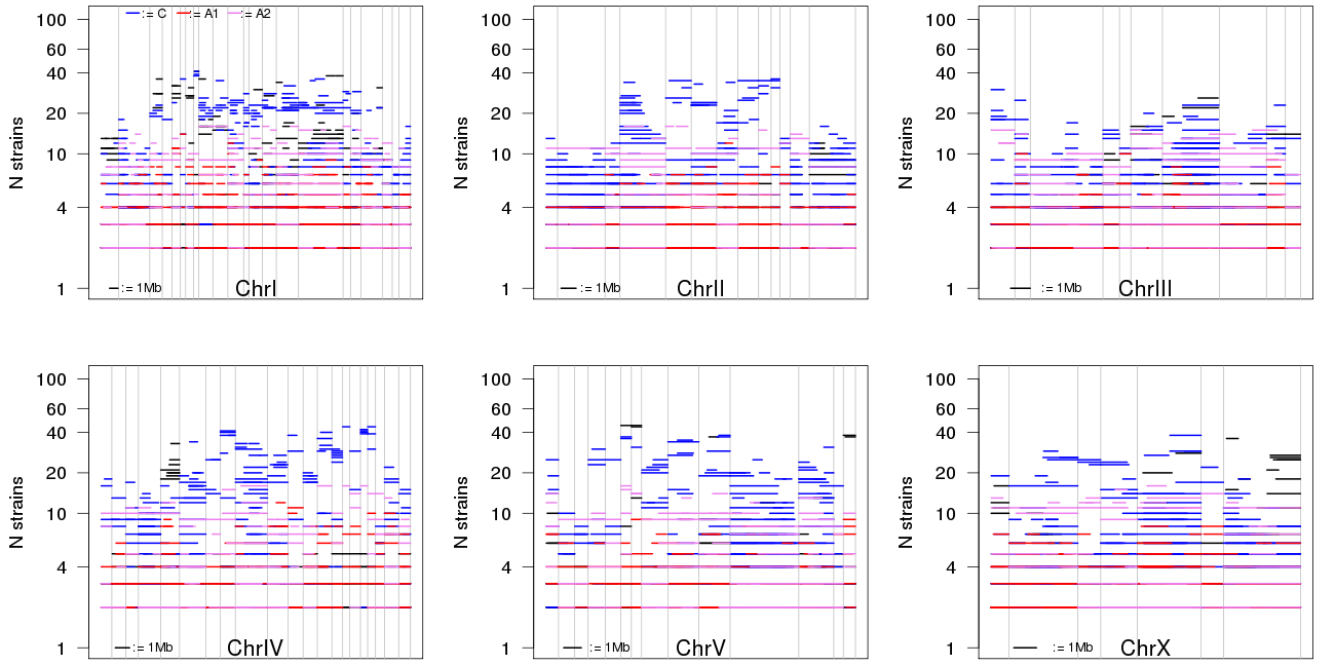

Figure S6 Length and frequency distribution of shared haplotype blocks. The x-axis denotes the relative positions of shared haplotypes separated by supercountig boundaries. All supercontigs with markers that have been mapped to the genetic mapped are shown. The haplotypes are colored with respect to a clade if all strains sharing that haplotype are member of that clade or black otherwise.

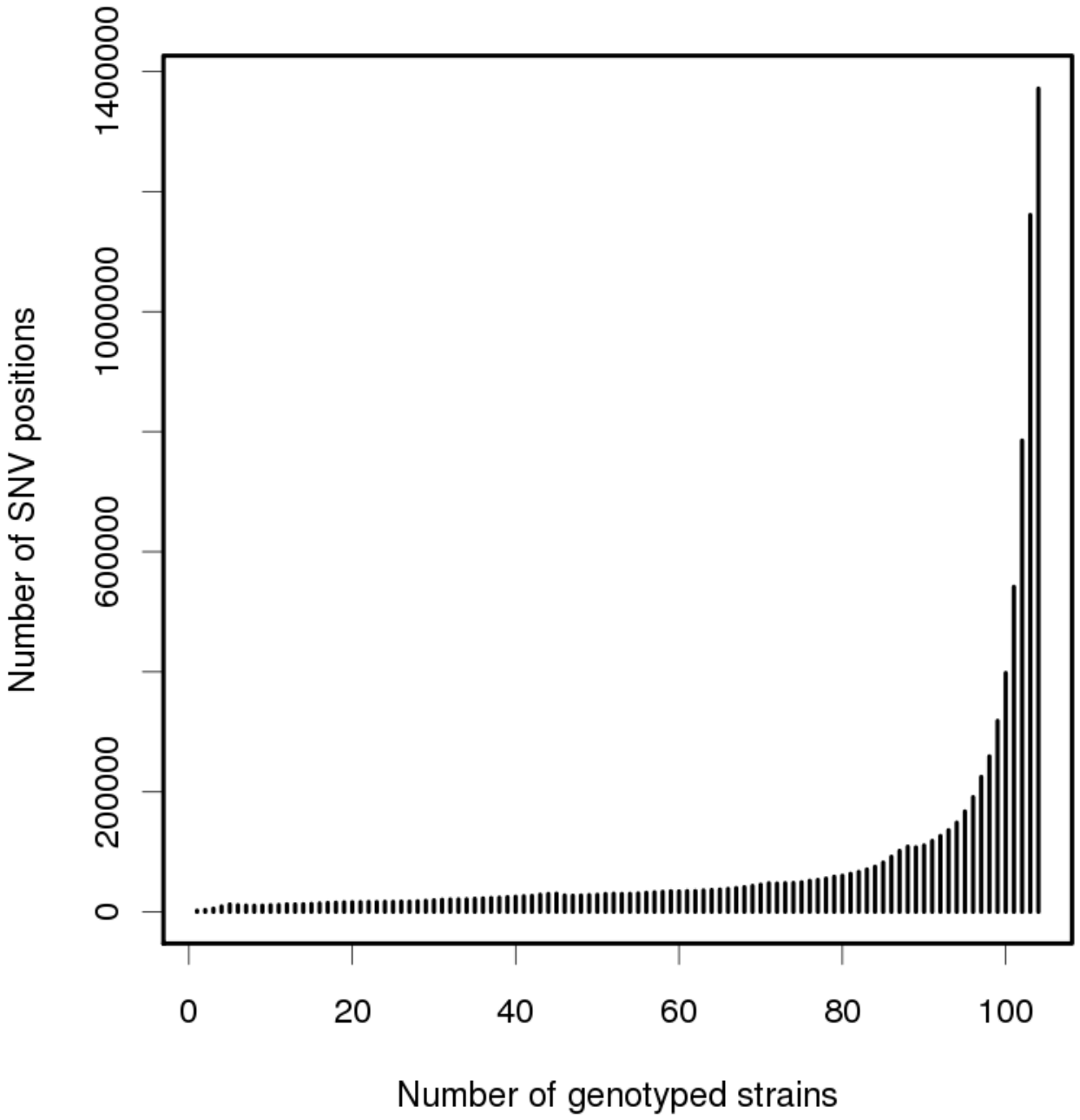

Figure S7 Number of SNV positions that could be genotyped in X strains. For all detected variable sites, we tested in how many strains these sites could be genotyped reliably (coverage ≥ 2, samtools quality score ≥ 20, no signal of heterozygosity in any of the strains). Around 1.4 million positions could reliably be genotyped in all 104 strains. For most population genetic analysis was done using only those 1.4 mio SNV positions.

Table S1 Numbers about sequenced strains

Available for download as an Excel file at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.159855/-/DC1