Figure S1

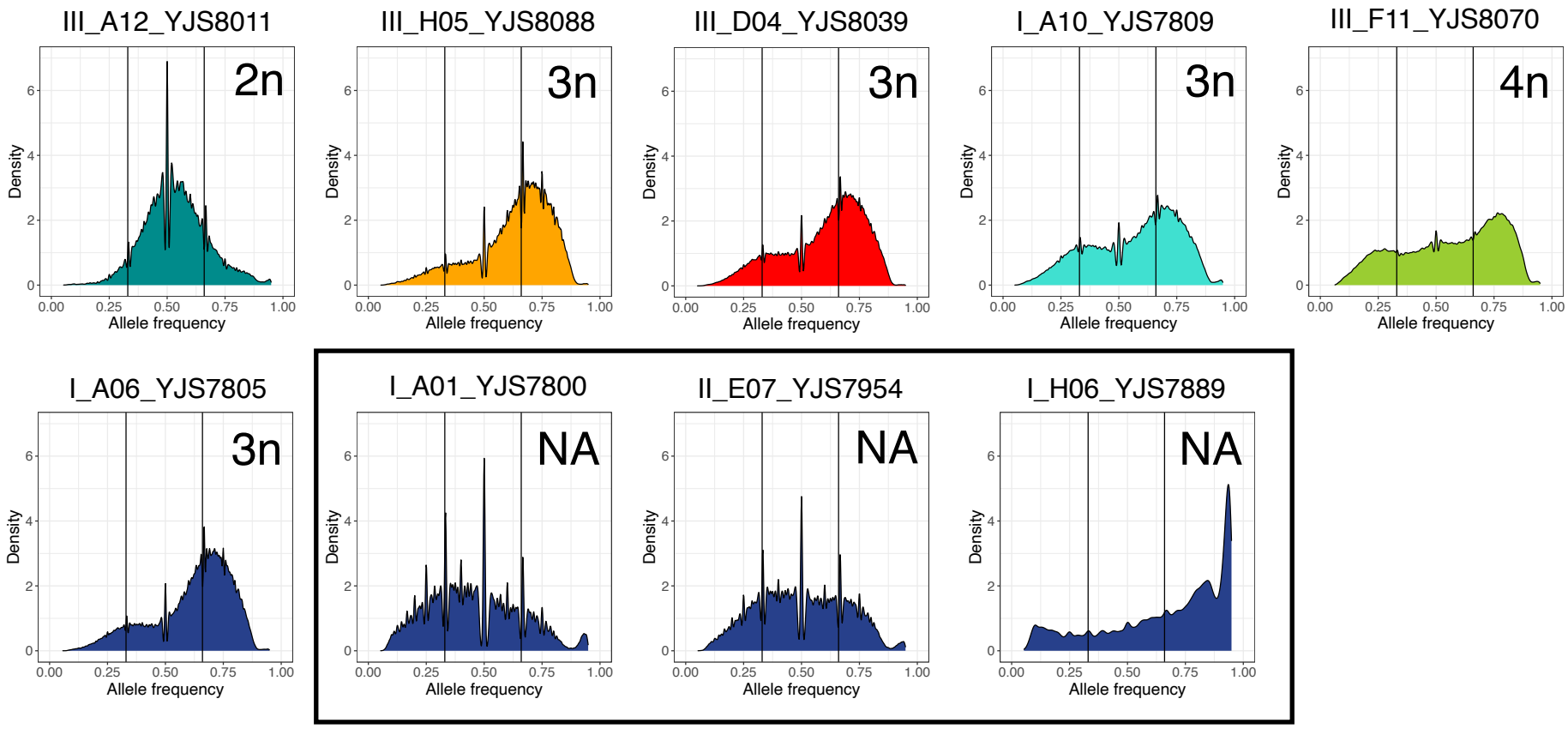
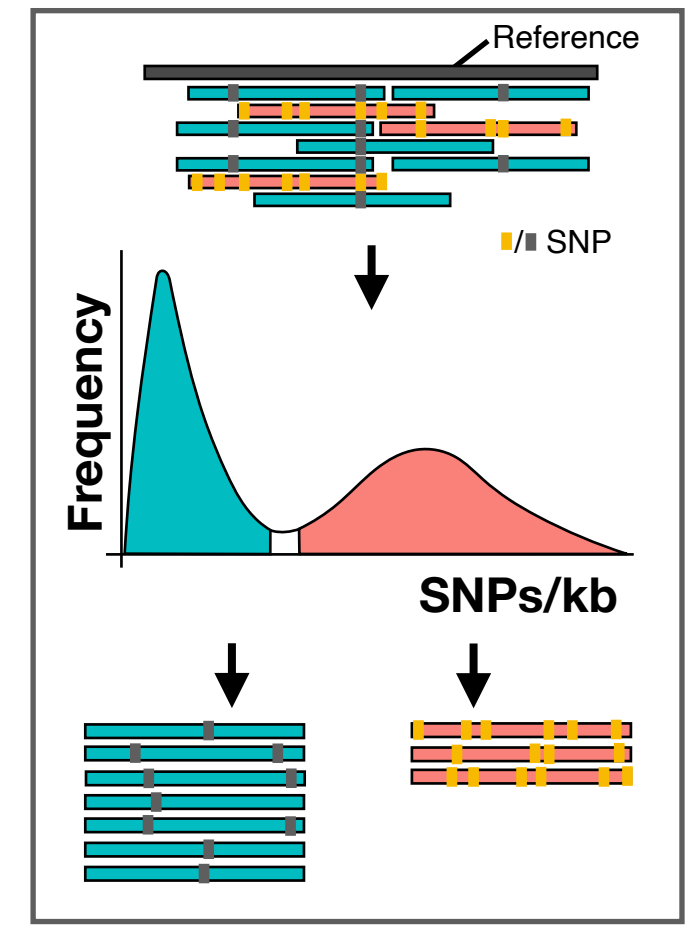


Figure S1: Allele frequencies and genome-wide coverage.
A Estimation of ploidy in 71 strains using short read sequencing data. Shown are examples of strains from each subpopulation. The level of ploidy varies between 2n (diploid) and 4n (tetraploid). For three individuals, ploidy could not be estimated based on 24,313 genome-wide distributed variants (framed in black). **B-C** Genome-wide coverage to detect potential aneuploidies. Coverage-based analysis did not show aneuploidies (segmental, chromosomal) that would explain the patterns of the three strains in (**A**), for which ploidy level failed to be determined by allele frequency. Shown are the strains I_A06_YJS7805 (**B**), for which ploidy could be determined (see Panel A), and II_E07_YJS7954 (**C**), where the usage of allelefrequencies was not sufficient to determine its ploidy.

Figure S2



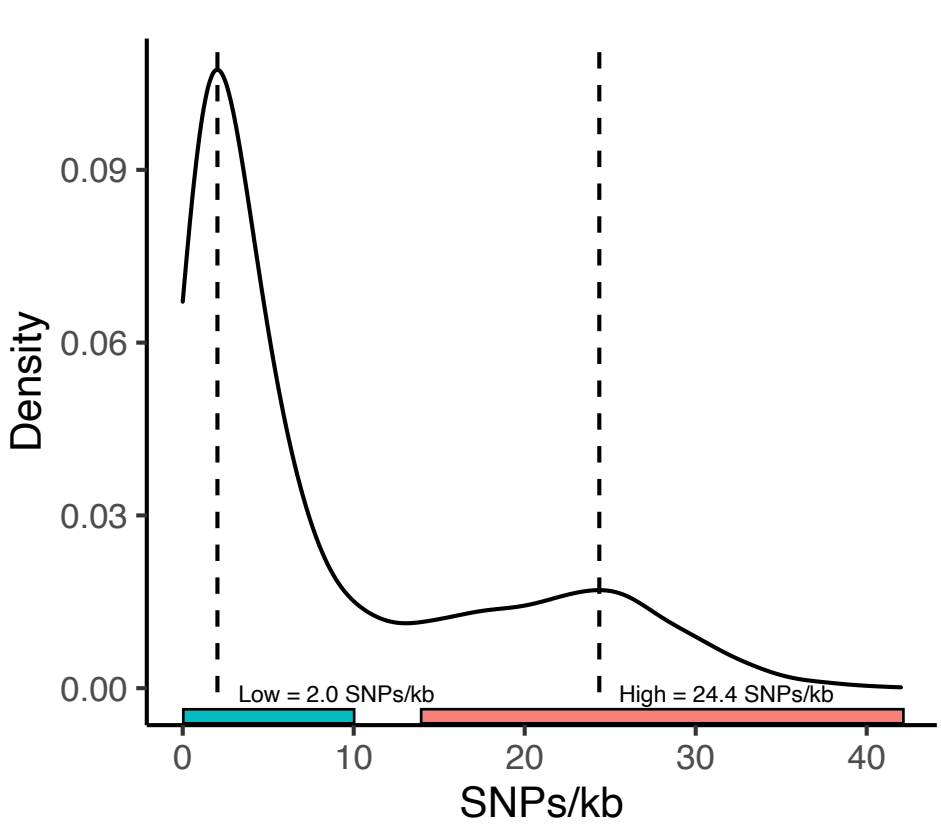


Figure S2: Separating sequencing reads based on intra-genomic variation to the reference genome.

A Long reads were first aligned to the reference genome *B. bruxellensis* (Fournier et al. 2017) and separated based on their variation (SNPs/kb). Respectively, reads with low genetic variation to the reference genome were clustered and defined as low-intra genomic variation, reads with high genetic variation to the reference genome were clustered and defined as high-intra genomic variation.

B Three subpopulations with high-intra genomic variation. Shown is the intra-genomic variation from strains of the subpopulations teq/EtOH, beer and wine 1. These strains harbour, besides low intra-genomic variation with an average of 2.0 SNPs/kb, a cluster of reads with high intra-genomic variation (average 24.4 SNPs/kb) to the reference genome.

Figure S3

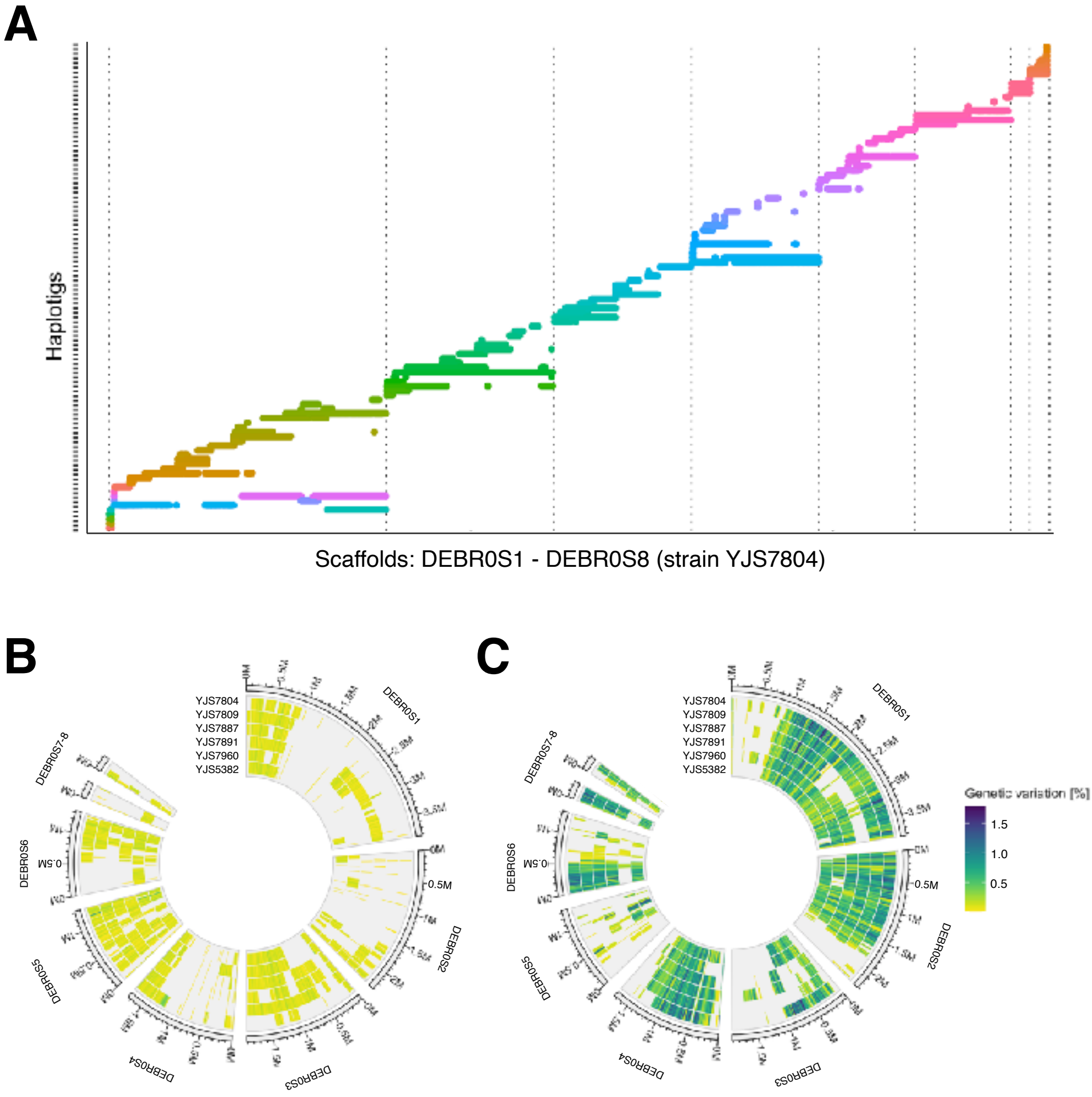


Figure S3: Phasing the polyploid wine 2 subpopulation with low intra-genomic variation.

A Separation of haplotypes. The program nPhase (Abou Saada et al. 2021) separated the chromosomes into haplotypes, which in most cases, resolves the chromosomes into two more haplotigs at a given region.

B-C Intra-genomic variation. The separation of regions underlying two (**B**) or three (or more; **C**) haplotypes correspond to different levels of intra-genomic variation. Regions with two haplotypes have on average intra-genomic variation of 0.09%, while this increases to 0.54% regions with three haplotypes.

Figure S4

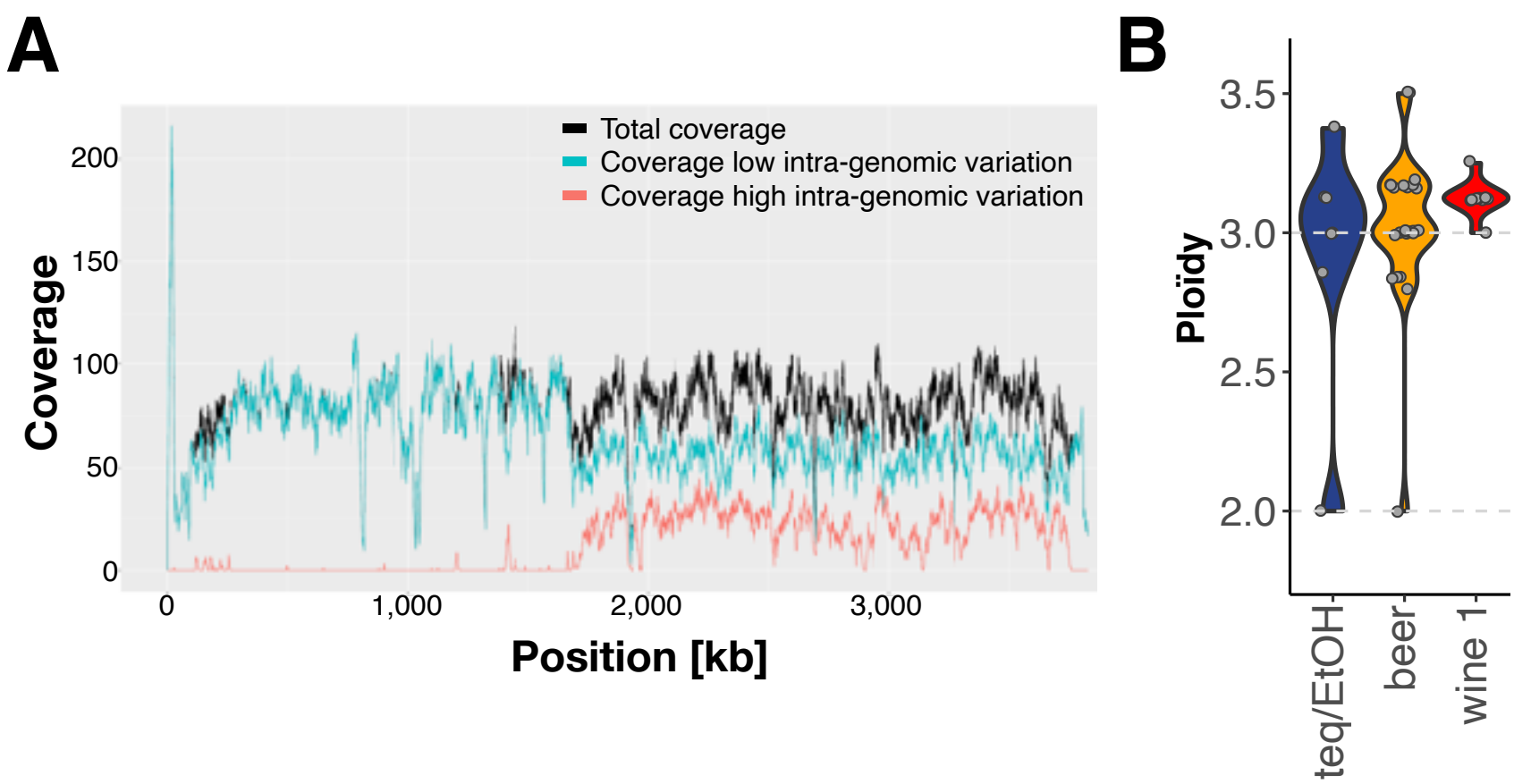


Figure S4: Coverage analysis and ploidy determination using long read sequencing data.

A Example of chromosome 1 (DEBR0S1) of a beer strain where long reads were aligned to the reference genome of *B. bruxellensis* (Fournier et al. 2017). First, reads have been separated based on the number of SNPs/kb, respectively into clusters of low or high intra-genomic variation to the reference genome (see: Material & Methods), and then compared to the total coverage at a given site.

B With the coverage of the reads bearing low or high intra-genomic variation to the reference genome (**A**), the ploidy was estimated for strains from the three subpopulations teq/EtOH, beer and wine 1. The reads containing high intra-genomic variation contributed on average to a third of the total coverage at each site, reflecting a triploid state (3n) for these strains. Ploidy was converted from ratios (see Material & Methods).

Figure S5

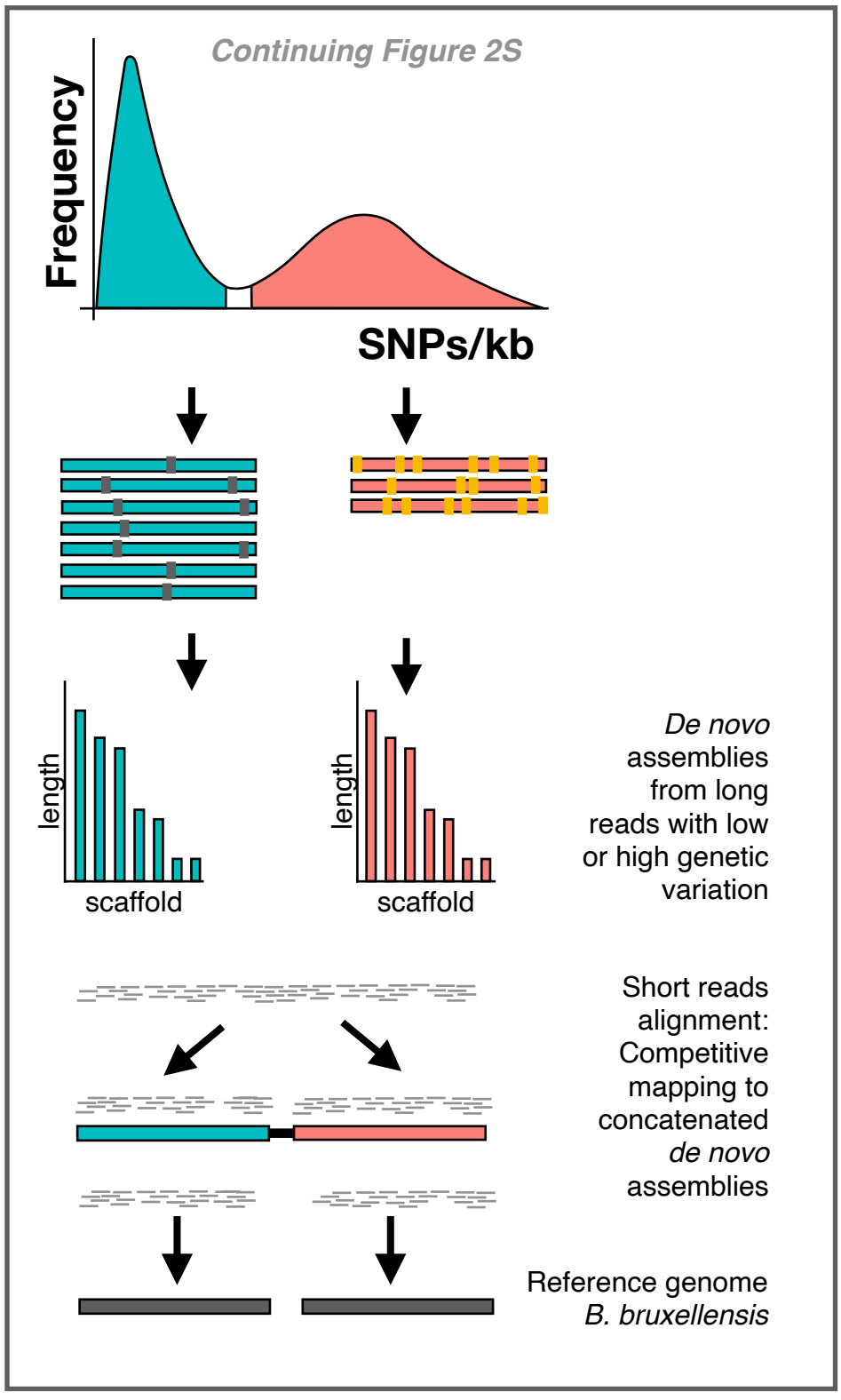
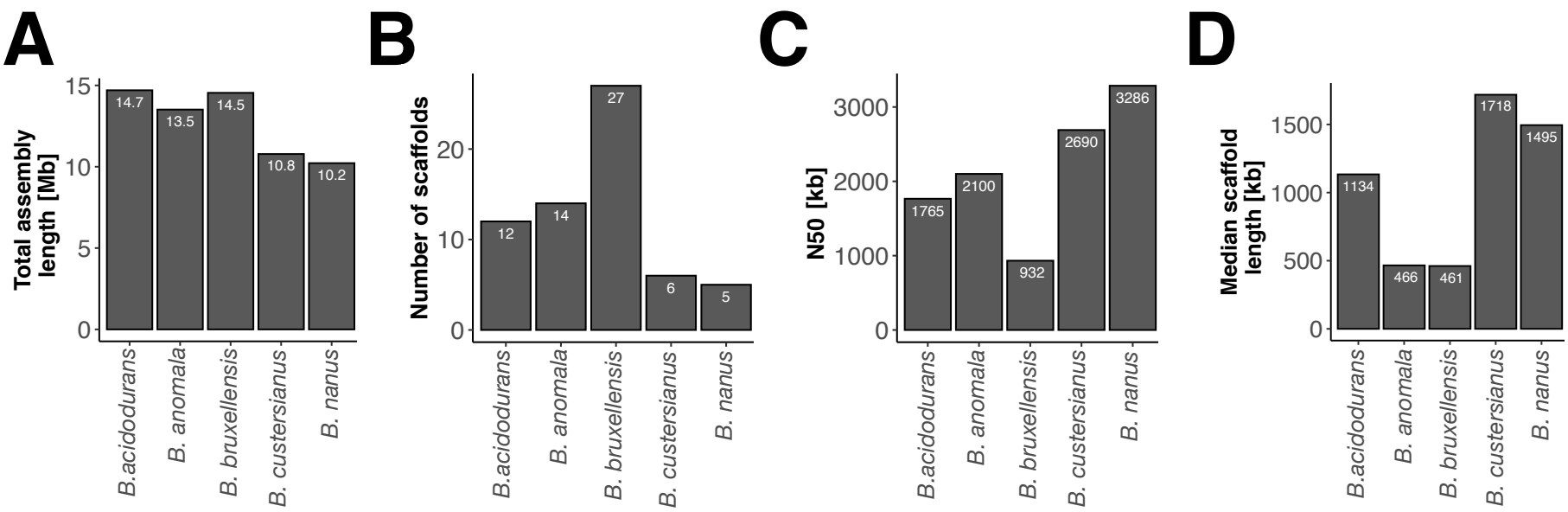


Figure S5: Competitive mapping for comparative genomic analysis (see also Figure 2S).
In order to perform a comparative analysis on the different genomic copies, first independent *de novo* assemblies were constructed from reads with low or high intra-genomic variation to the reference genome (Fournier et al. 2017). *De novo* assemblies from three strains of the different subpopulations were then concatenated and used as reference sequences for the other strains of the same subpopulation (see: Material & Methods). Short sequencing reads were separated based on a comparative mapping approach using the concatenated *de novo* genome assemblies. Finally, the separated short sequencing reads were aligned back to the reference genome of *B. bruxellensis*, to perform comparative analyses.

Figure S6



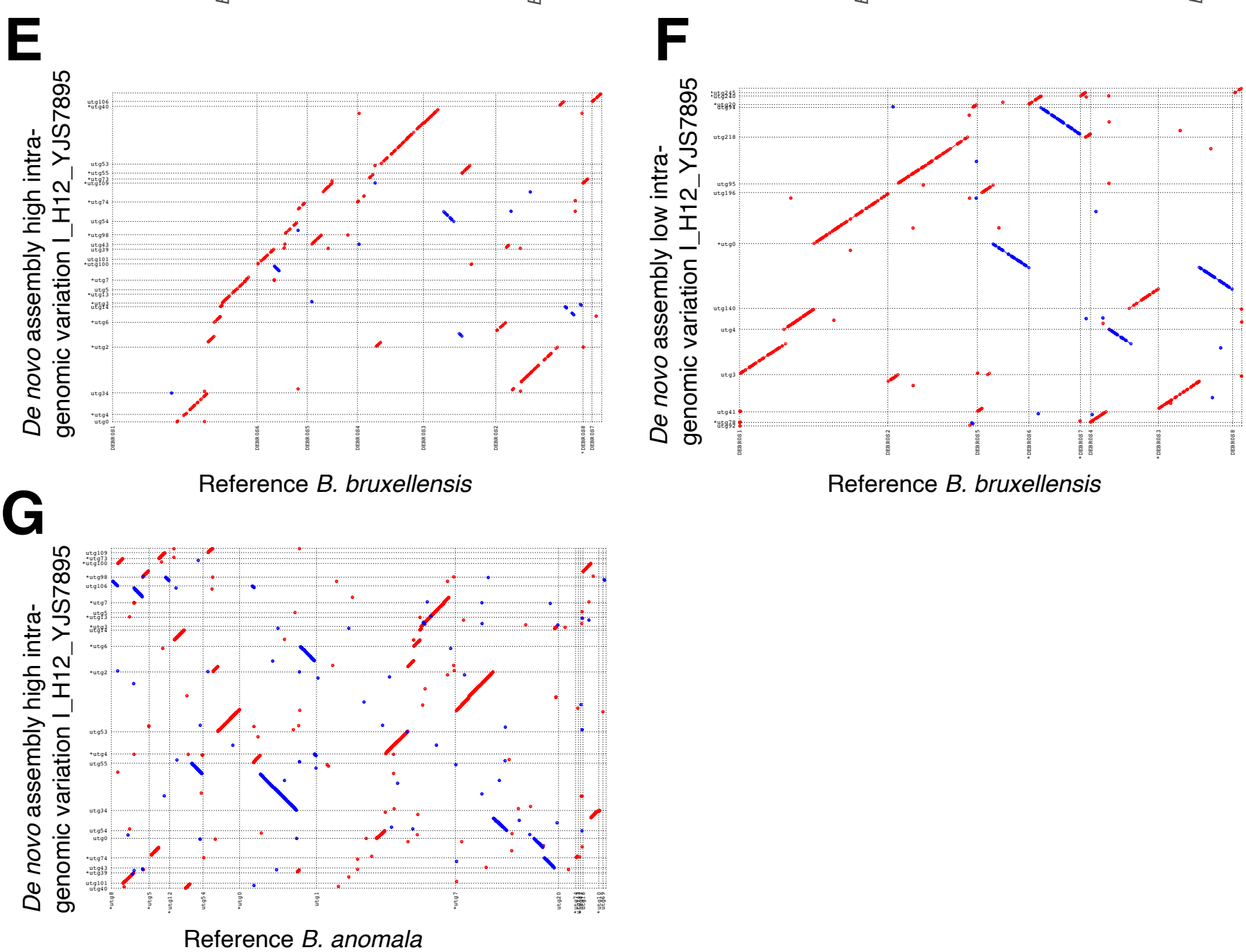


Figure S6: *De novo* genome assembly statistics for four sister species of *B. bruxellensis* and *B. bruxellensis*.

The assembly statistics (**A-D**) are in accordance with those from Roach & Borneman (2020).

E-F Collinearity plots comparing synteny between *de novo* assemblies. Both assemblies, respectively performed using reads with high or low intra-genomic variation reveal a good synteny, albeit rearrangements with the reference genome *B. bruxellensis* (Fournier et al 2017) can be seen (MUMmer parameters: --mum -l 200).

G Collinearity plots comparing synteny between the *de novo* assembly from reads with high intra-genomic variation and *B. anomala*. Collinearity is disrupted by many small syntenic elements, which only appear as parameters were loosen (MUMmer parameters: -- mum -l20 -c 30 -b 100).

Figure S7

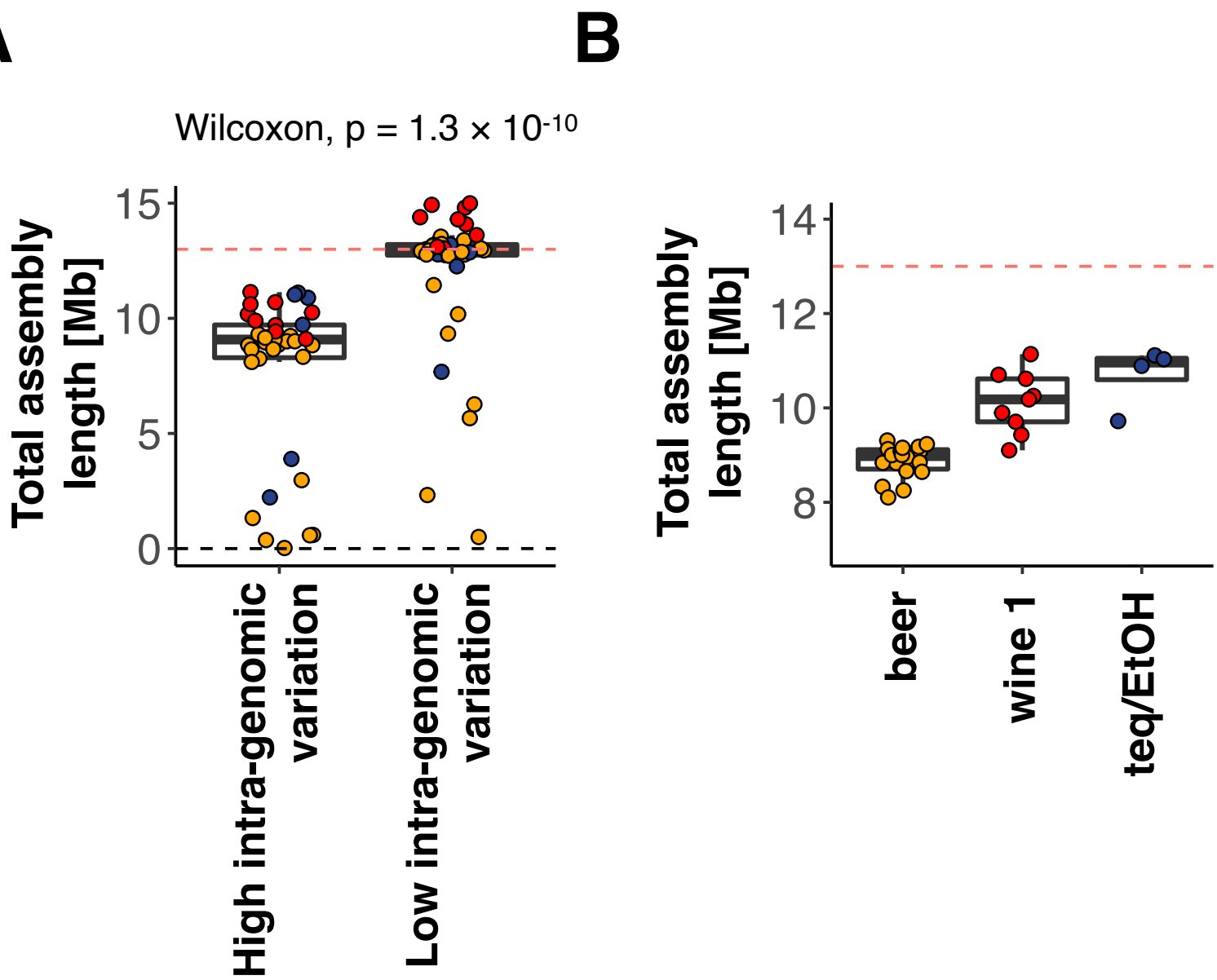


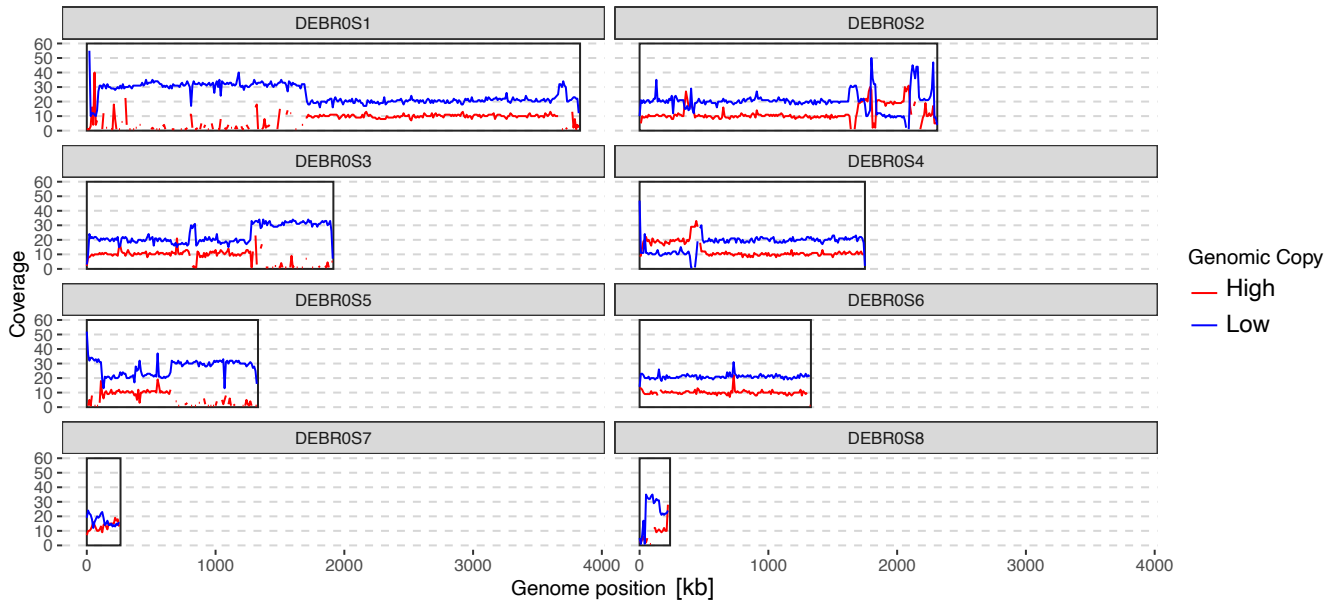
Figure S7: *De novo* genome assemblies.

A Genome assemblies were prepared for the reads harbouring either low or high-intra genomic variation to the reference. The difference in total assembly size for the *de novo* assemblies of the high intra-genomic variation is 9.08 Mb, significantly different to those based on low intra-genomic variation with 12.92 Mb (median; p value = 1.3×10^{-10}).

B The total assembly sizes of strains from different subpopulations (only high intra-genomic variation). Lowest total assembly length was determined for the subpopulation beer (8.84 Mb), followed by wine 1 (10.18 Mb), and teq/EtOH (10.34 Mb). Red dotted line represents the assembly length of the *B. bruxellensis* reference genome assembly (Fournier et al. 2017; 13 Mb).

Figure S8

A



B

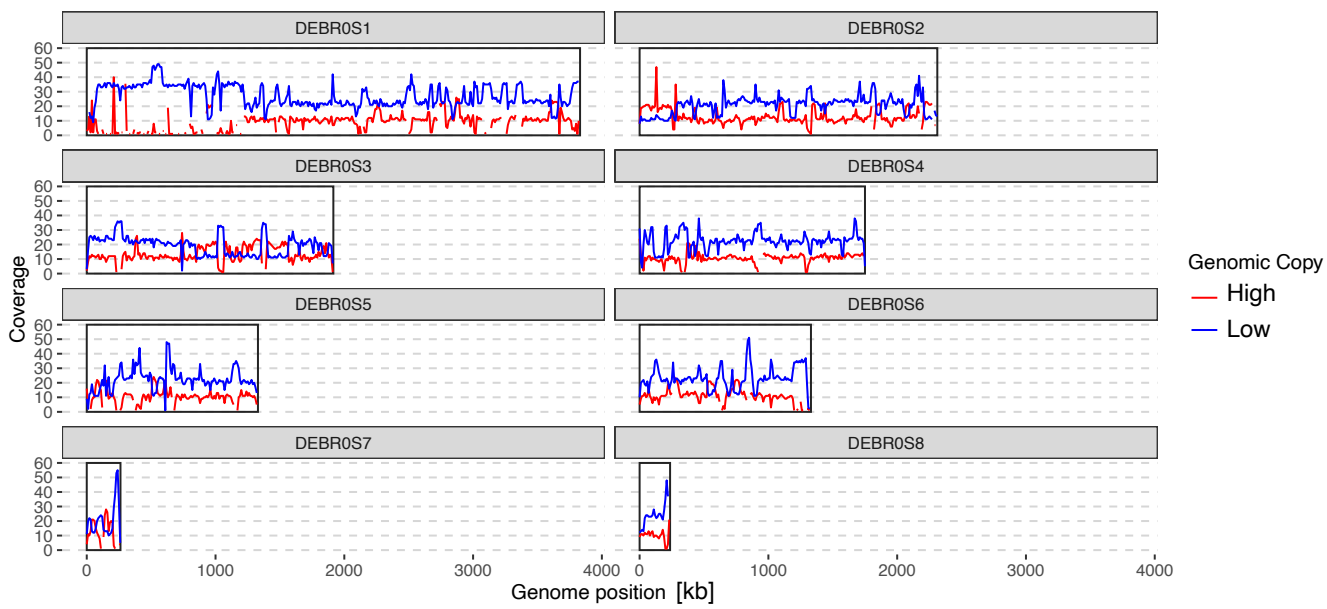


Figure S8: Reciprocal shifts in coverage underlying LOH events.
A-B Example of coverage plots, calculated using 10 kb windows along the scaffolds for two strains from the subpopulations beer and wine 1. Plotted are the strains IH12_YJS7895 (**A**; beer) and III_D04_YJS8039 (**B**; wine 1), which were initially aligned to subpopulation-specific reference genomes (concatenated *de novo* assemblies) to separate reads with low or high intra-genomic variation, and then aligned back to the reference genome *B. bruxellensis* (Fournier *et al.*, 2017). In red (High), the reads from the high intra-genomic copy are shown, in blue (Low), the reads from the low intra-genomic copy. Example (A): The average coverage of the High genomic copy is 10x (=haploid), while the Low genomic copy is 20x (=diploid). The total coverage (sum of High + Low) is 30x. This coverage of Low vs. High is not consistent across the genome, where shifts in coverage show that additional regions of a genomic copy have been acquired or were lost.

Figure S9

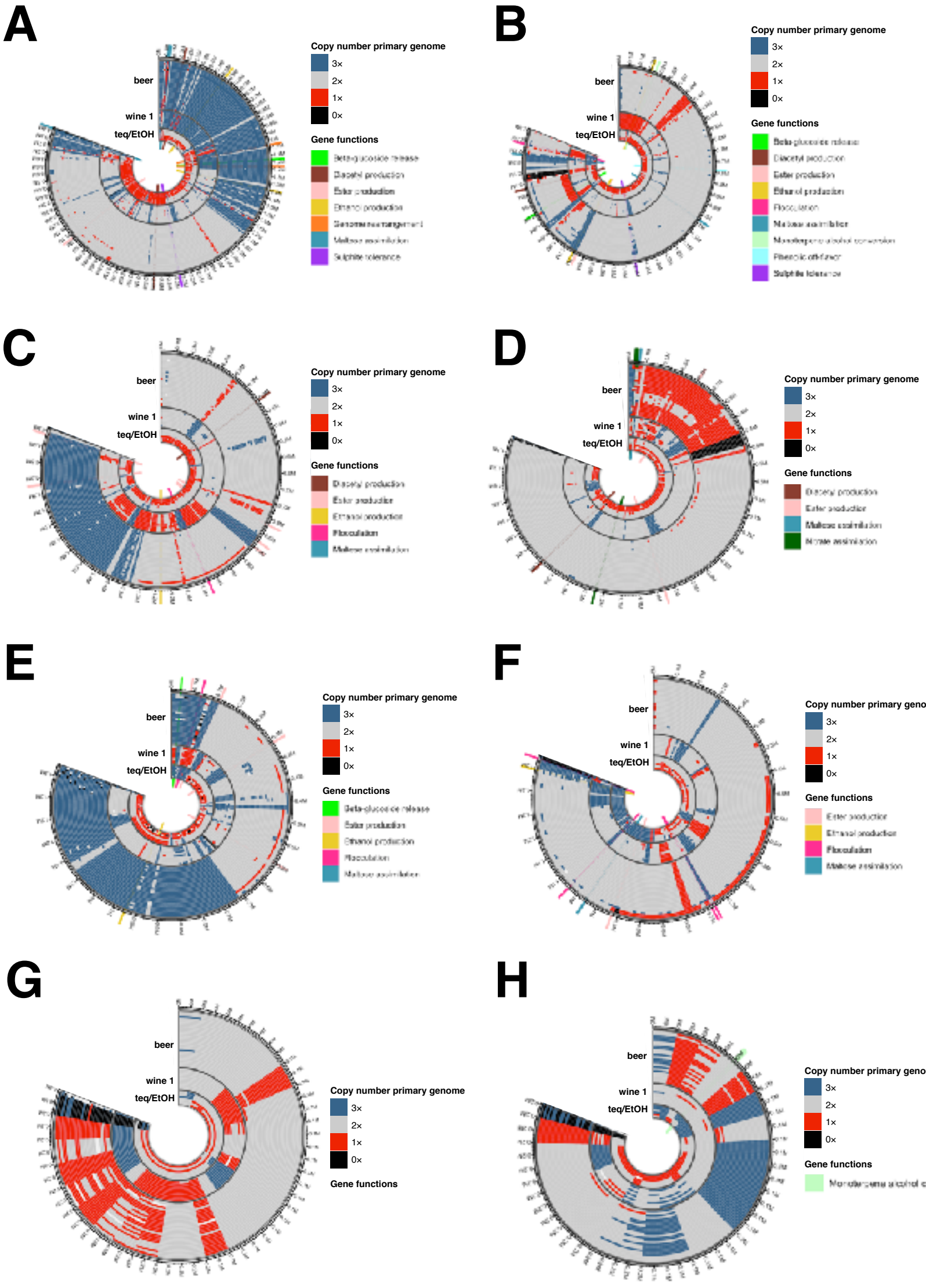


Figure legend, see next page.

Figure S9

Figure S9: Candidate approach to test for functional enrichment in polyploids.
A-H A set of 65 candidate genes were used to check if regions, underlying genomic modifications in the polyploid strains from the subpopulations teq/EtOH, beer and wine 1, are particularly enriched in genes underlying different functions. The subset of genes is from Colomer et al. (2020). Plots are showing the number of haplotypes associated with low or high intra-genomic variation in 10 kb windows.

Figure S10

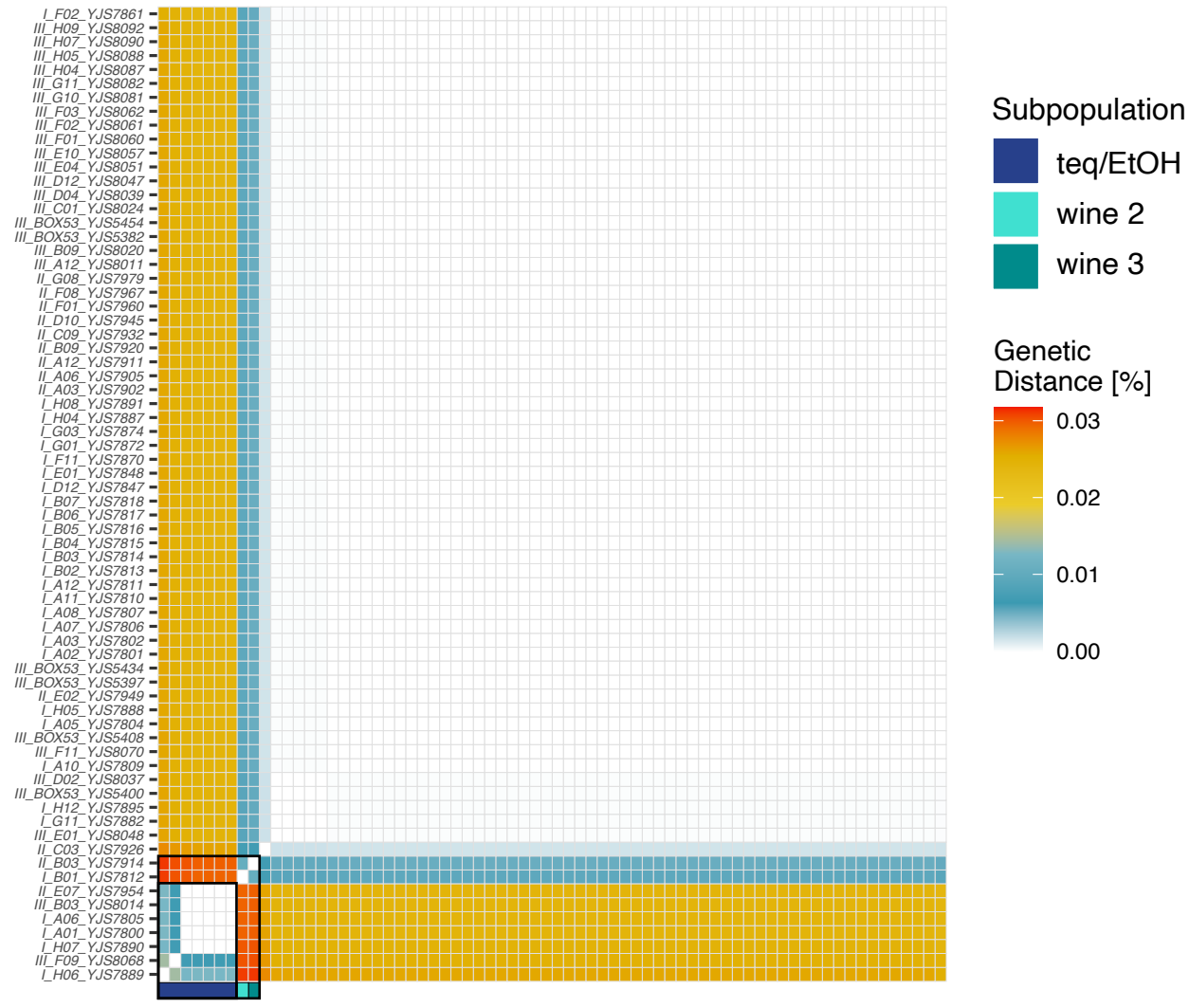


Figure S10: Pairwise genetic distance.

We concatenated gene sequences (*ATP6, ATP8, COX2, COX3, NAD1, NAD3, NAD4, NAD4L*) to calculate pairwise genetic distance between 70 strains (III_G10_YJS8081 is missing since mitochondrial assembly could not be generated). Cluster analysis reveals that the teq/EtOH subpopulation is highly diverged to other strains.