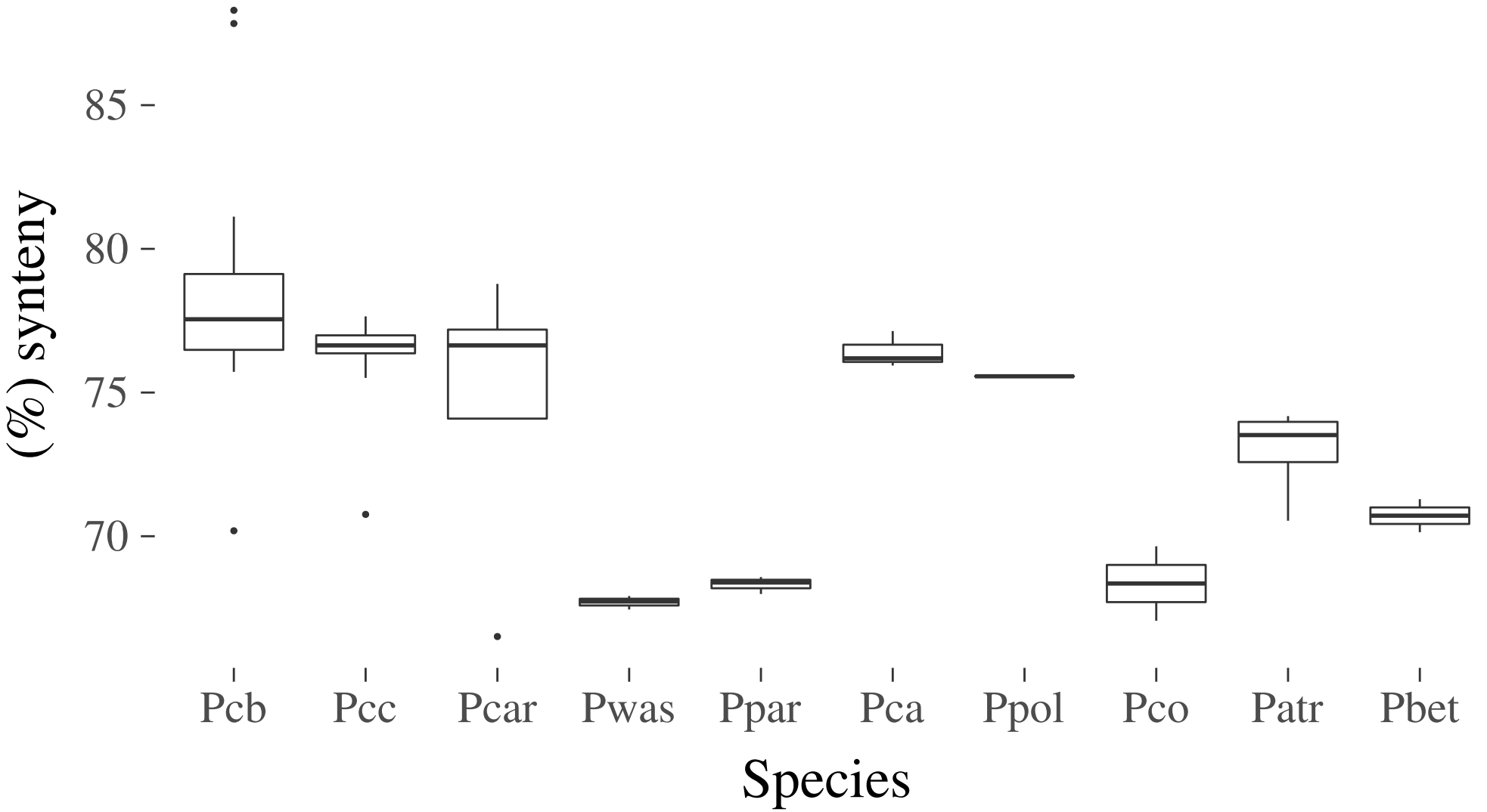


Figure S1 – Overall synteny between Pcb 1692 genome and 99 SRE strains: The boxes illustrate, for each species-label, the overall syntenic percentage of all respective strains compared to Pcb 1692 genome. The percentages found during the synteny analysis are represented in the y axis, while the x axis represents species labels. Species abbreviations follow the same pattern utilized in Figure 3.

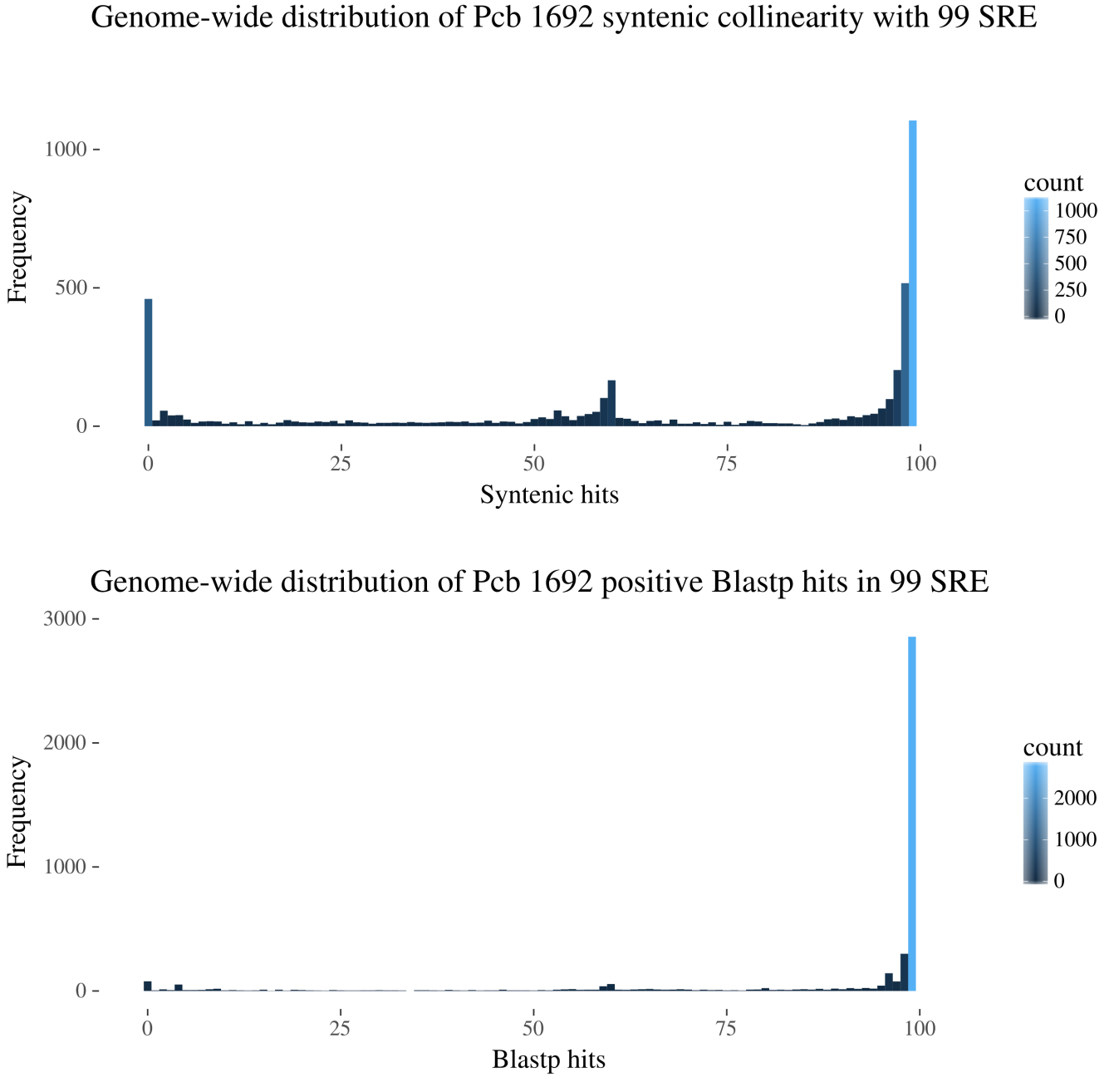


Figure S2 - Frequency of conservation of PcbPr1 genes in SRE in comparison to genome-wide conservation of Pcb 1692 protein-coding genes.
Blastp parameters used: evalue threshold: 1.

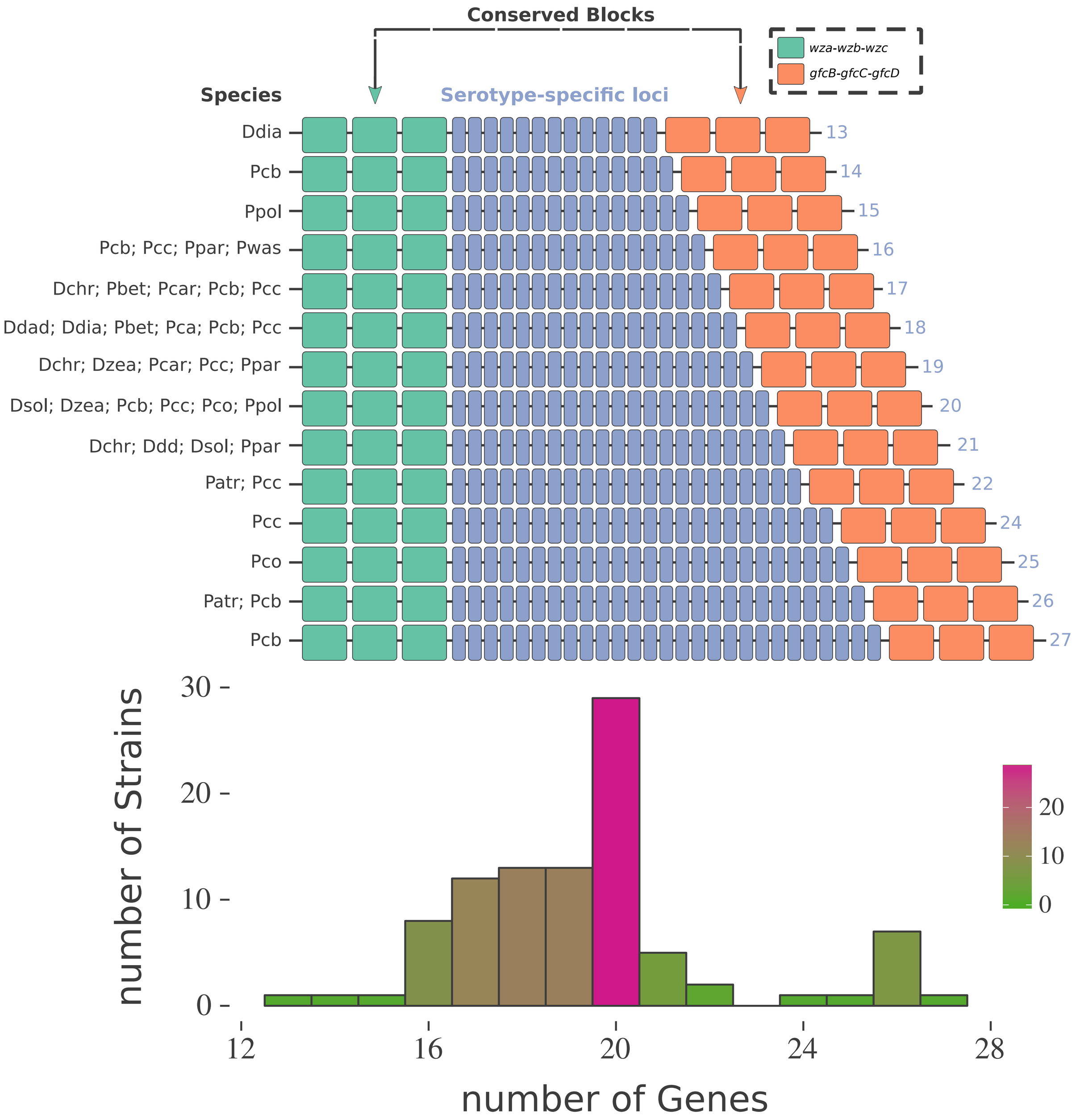


Figure S3 – Representativeness of serotype-specific block sizes within GFC region in SRE genomes: In the scheme on the top, each row contains a different architecture size of serotype-specific blocks found in SRE. These rows are labeled: on the left by 4-letter species abbreviation as in Table S7; on the right by the respective number of serotype-specific genes in the block. The histogram on the bottom depicts the frequency by which each serotype-specific block-size is represented across 100 SRE strains.

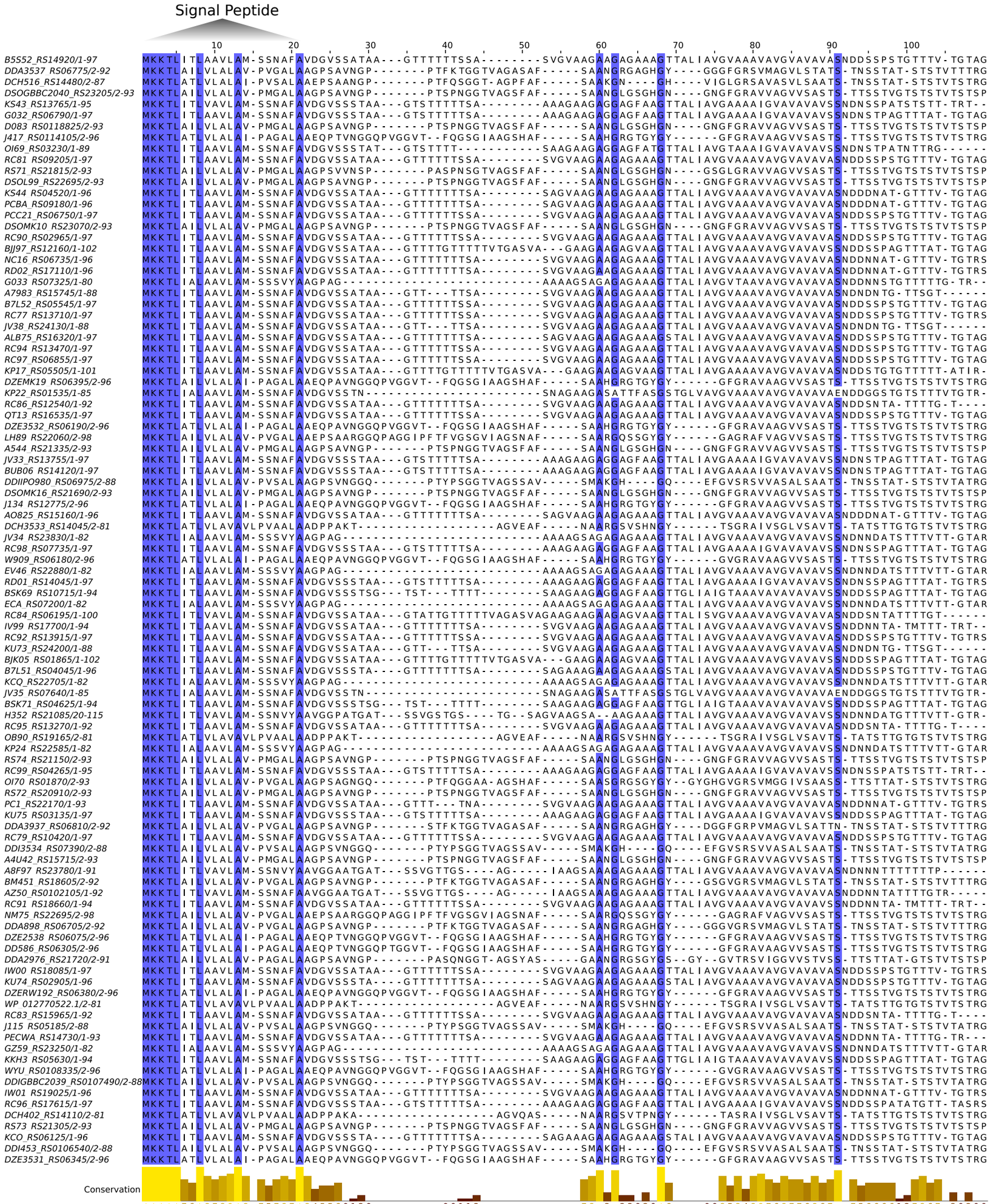


Figure S4 – YjbF-upstream sequence alignment: All proteins encoded by genes upstream of yjbF/gfcB in 100 SRE strains were aligned. Columns highlighted in purple represent conserved residues.