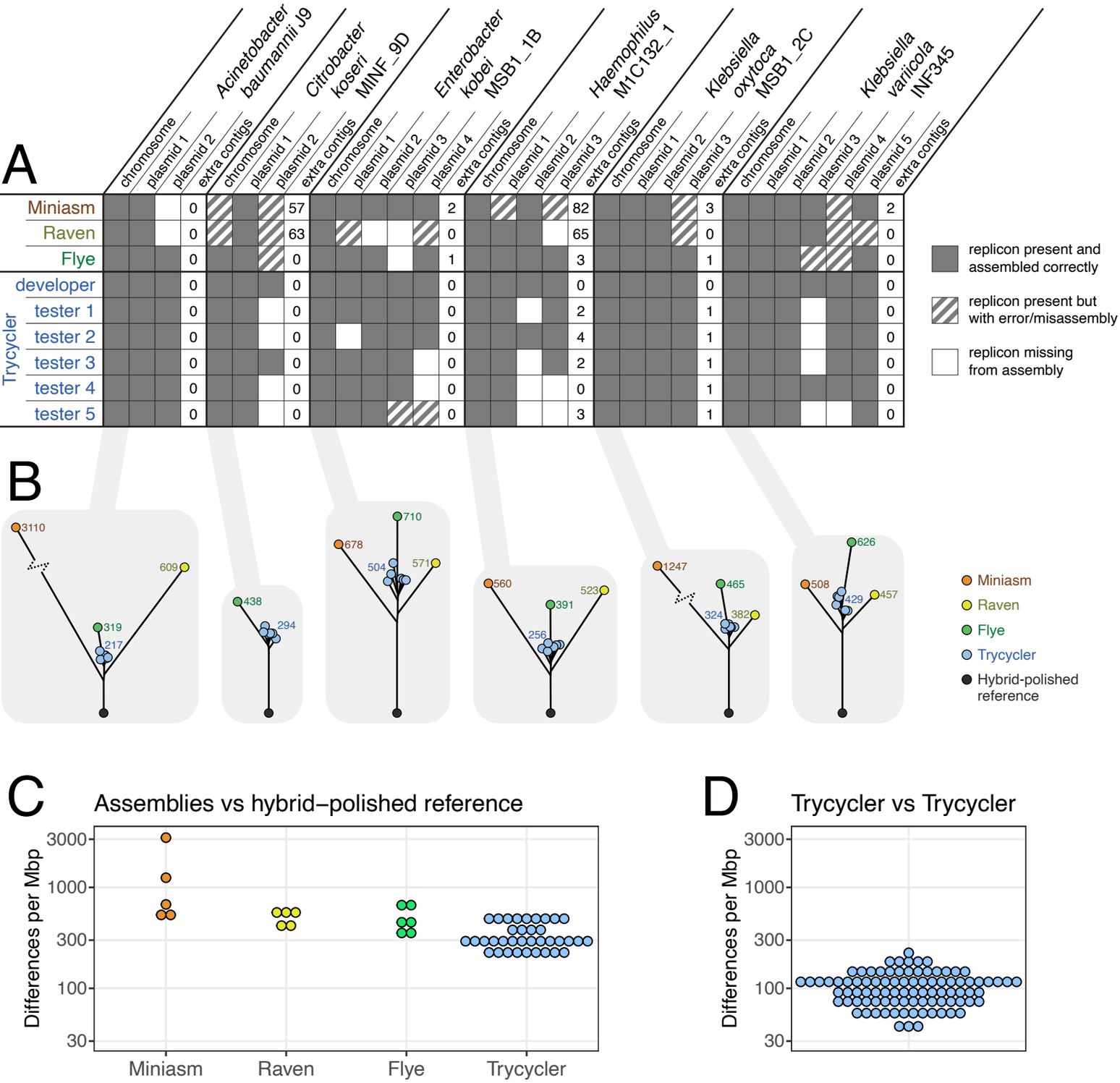

Figure 4: results for the multi-user test which assessed the consistency of Trycycler assemblies when run by different users. Results include assemblies from three different long-read assemblers (Miniasm/Minipolish, Raven and Flye, all automated and deterministic for a given set of reads and parameters, i.e. independent of user) and Trycycler assemblies from six different users (the developer of Trycycler and five testers).

A: Presence/absence matrix for the replicons in the test genomes. Each replicon was classified as either present in the assembly, absent from the assembly, or present but with an error/misassembly (see Table S3 for more detail). The number of additional contigs (e.g. spurious or contaminant sequences) is also indicated for each assembly. All Trycycler assemblies contained an accurate chromosome, and only one Trycycler assembly contained misassemblies. However, in many cases the Trycycler testers excluded a true plasmid (most commonly a small plasmid) or included an additional plasmid (most commonly constructed from cross-barcode contaminating reads).

B: Neighbour-joining trees of all available assemblies for each of the chromosomes, based on pairwise alignment distances. Hybrid-polished (Medaka+Pilon) versions of the developer's Trycycler assemblies were included as reference sequences. The values indicate the number of single-bp differences per Mbp between each assembly and the polished reference (values for Trycycler are the mean of all six Trycycler assemblies). For each genome, the Trycycler assemblies cluster tightly and are closer to the polished reference than those from other long-read assemblers.

C: Differences between each assembled chromosome and the hybrid-polished reference. Values are single-bp differences per Mbp of sequence. Trycycler assemblies contain fewer differences, on average, compared to the single-assembler assemblies.

D: Pairwise differences between Trycycler assemblies of each chromosome. Values are single-bp differences per Mbp of sequence, and there are 90 values (6 genomes × 15 unique pairwise combinations per genome).