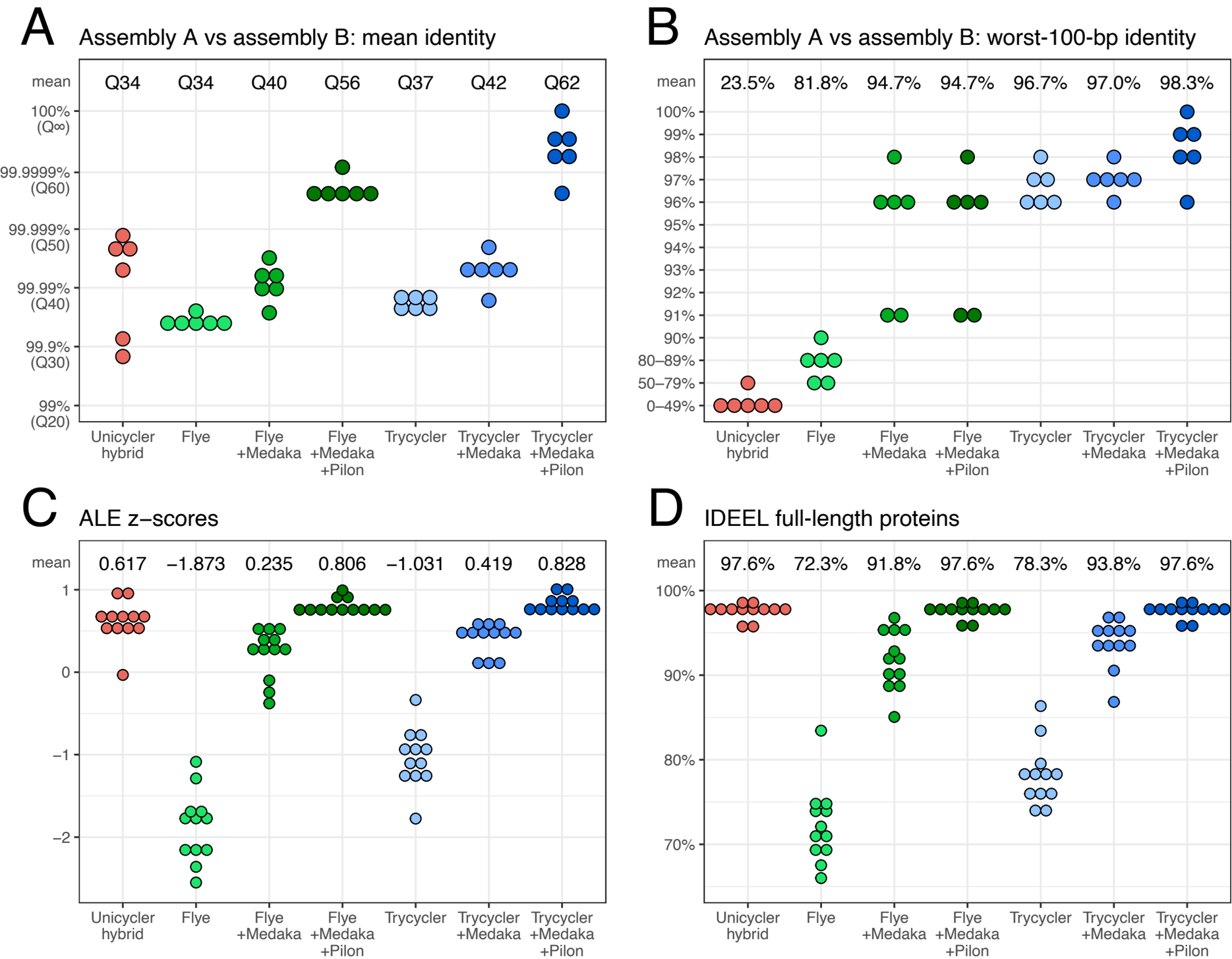

Figure 2: results for the real read tests. For six genomes, we produced two independent hybrid read sets from the same DNA extraction. The read sets were then assembled with Unicycler (hybrid assembly), Flye (long-read-only assembly), Flye+Medaka (long-read-only assembly), Flye+Medaka+Pilon (hybrid assembly), Trycycler (long-read-only assembly), Trycycler+Medaka (long-read-only assembly) and Trycycler+Pilon (hybrid assembly).

For each of the six genomes, we aligned the two independently assembled chromosomes to each other to determine the mean assembly identity (A) and the worst identity in a 100 bp sliding window (B). For long-read-only assembly, Trycycler consistently achieved higher accuracy than Flye (both before and after Medaka polishing). Trycycler+Medaka+Pilon achieved the highest accuracy and did better than alternative hybrid approaches (Unicycler and Flye+Medaka+Pilon).

We also assessed the accuracy of each of the 12 assembled chromosomes using ALE (C) and Ideel (D). ALE assigns a likelihood score to each assembly based on its concordance with the Illumina read set (for this plot we transformed the ALE scores into z-scores on a per-genome basis). IDEEL identifies the proportion of full-length protein-coding genes using a TrEMBL UniProt database for reference.