



Figure 3: 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 (short-read-first hybrid assembly), Flye (long-read-only assembly), Flye+Medaka (long-read-only assembly), Flye+Medaka+Pilon (long-read-first hybrid assembly), Tricycler (long-read-only assembly), Tricycler+Medaka (long-read-only assembly) and Tricycler+Pilon (long-read-first hybrid assembly).

For each genome and each assembly approach, we aligned the two independently assembled chromosomes to each other to determine the mean assembly identity (A) and the worst identity in 100-bp sliding windows (B). For long-read-only assembly, Tricycler consistently achieved higher accuracy than Flye (both before and after Medaka polishing). Tricycler+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 (transformed into z-scores on a per-genome basis). IDEEL identifies the proportion of predicted proteins which are $\geq 95\%$ the length of their best-matching known protein in a database.