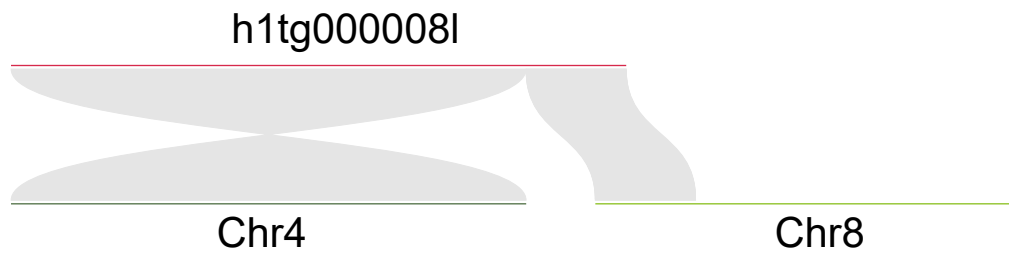
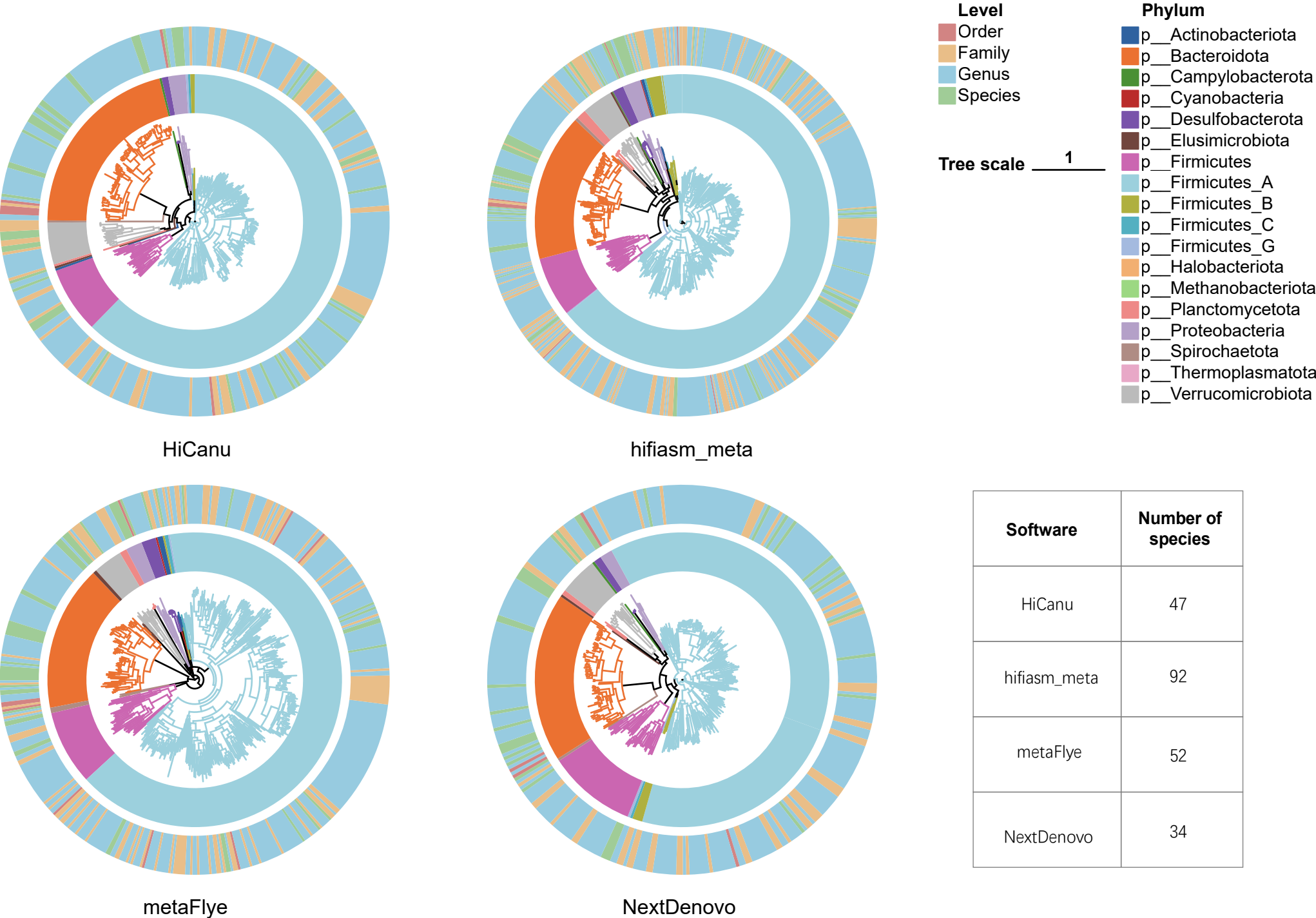


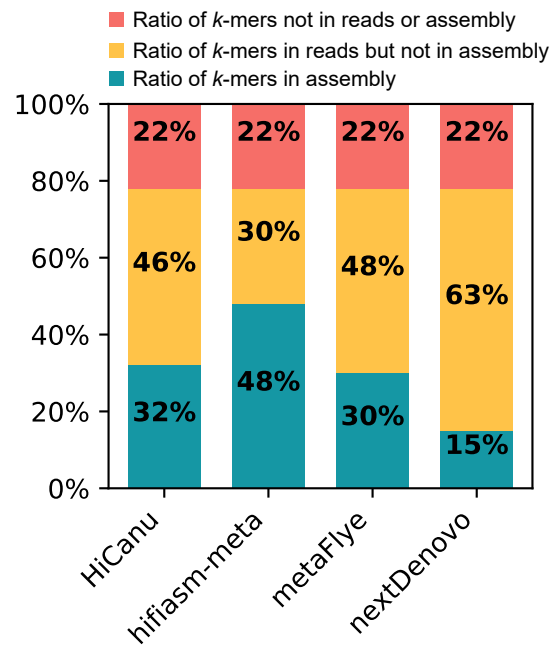
Supplemental Figure 1. Distribution of *k*-mer frequencies of real heterozygous diploid genome and autotetraploid genome and synthetic genomes. A. Distribution of *k*-mer frequencies of big berry manzanita (1) (heterozygous diploid genome); B. Distribution of *k*-mer frequencies of the synthetic heterozygous diploid genome; C. Distribution of *k*-mer frequencies of the potato genome (2) (autotetraploid); D. Distribution of *k*-mer frequencies of the synthetic autotetraploid genome.



Supplemental Figure 2. Investigating the reason for the low APLC of hifiasm . Contig h1tg0000081 was mis-assembled as shown in this alignment.



Supplemental Figure 3. Phylogenetic tree on contigs longer than 1Mbp from the assemblies of four assemblers. Phylogenetic tree is constructed by GTDB-Tk. Each leaf of the tree represents one contig. We determined the taxonomic level and domain in which the contigs are classified in the tree, with the assigned level being the minimum level that GTDB-Tk can assign the sequence to, either a family or genus. The outer circles represent the taxonomic level of the corresponding contig species, with the levels we counted being order, family, genus, and species as shown in the legend. The inner circles are colored according to the names of the domains in which the contigs are located in the tree, and the colors of each domain are shown in the legend. The table in the figure shows the number of contigs that each assembly can assign to species. The higher the number of contigs that can be assigned to species in theory, the better the assembly results.



Supplemental Figure 4. Unique *k*-mer completeness of reads and assemblies. In each bar plot, blue part represents the fraction of reference unique *k*-mers appearing in the assembly; yellow part represents the fraction of reference unique *k*-mers appearing in the reads but not in assembly; red part represents the fraction of reference unique *k*-mers appearing not in reads.