Title

Comprehensive assessment of 11 de novo HiFi assemblers on complex eukaryotic genomes and metagenomes.

Authors

Yu, Wenjuan

Luo, Haohui

Yang, Jinbao

et al.

Publication Date

2024-03-01

DOI

10.1101/gr.278232.123

Peer reviewed

Resource

Comprehensive assessment of 11 de novo HiFi assemblers on complex eukaryotic genomes and metagenomes

Wenjuan Yu,[1,6] Haohui Luo,[1,6] Jinbao Yang,[1,5,6] Shengchen Zhang,[1,5,6] Heling Jiang,[1,6] Xianjia Zhao,[1,4] Xingqi Hui,[1,4] Da Sun,[1] Liang Li,[2] Xiu-qing Wei,[2] Stefano Lonardi,[3] and Weihua Pan[1]

[1]*Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China;* [2]*Fruit Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian 350002, China;* [3]*Department of Computer Science and Engineering, University of California, Riverside, California 92521, USA;* [4]*School of Agricultural Sciences, Zhengzhou University, Zhengzhou, Henan 450001, China;* [5]*College of Informatics, Huazhong Agricultural University, Wuhan 430070, China*

Pacific Biosciences (PacBio) HiFi sequencing technology generates long reads (>10 kbp) with very high accuracy (<0.01% sequencing error). Although several de novo assembly tools are available for HiFi reads, there are no comprehensive studies on the evaluation of these assemblers. We evaluated the performance of 11 de novo HiFi assemblers on (1) real data for three eukaryotic genomes; (2) 34 synthetic data sets with different ploidy, sequencing coverage levels, heterozygosity rates, and sequencing error rates; (3) one real metagenomic data set; and (4) five synthetic metagenomic data sets with different composition abundance and heterozygosity rates. The 11 assemblers were evaluated using quality assessment tool (QUAST) and benchmarking universal single-copy ortholog (BUSCO). We also used several additional criteria, namely, completion rate, single-copy completion rate, duplicated completion rate, average proportion of largest category, average distance difference, quality value, run-time, and memory utilization. Results show that hifiasm and hifiasm-meta should be the first choice for assembling eukaryotic genomes and metagenomes with HiFi data. We performed a comprehensive benchmarking study of commonly used assemblers on complex eukaryotic genomes and metagenomes. Our study will help the research community to choose the most appropriate assembler for their data and identify possible improvements in assembly algorithms.

[Supplemental material is available for this article.]

Advances in sequencing technology have been a driving force in molecular biology and genomics, in particular for de novo genome assembly (Alhakami et al. 2017; Sohn and Nam 2018; Sun et al. 2022b). Single-molecule sequencing (SMS) technologies currently on the market can generate long reads that can span most repetitive regions in eukaryotic genomes and thus have simplified the de novo assembly problem. A notable example of SMS is Pacific Biosciences (PacBio) HiFi technology that can provide reads >10 kbp with very high accuracy (<0.01% sequencing error). Oxford Nanopore Technologies (ONT) can generate even longer reads (up to a few mega base pairs), but the sequencing error rate can be as high as 3%. PacBio HiFi reads have enabled significant improvements in the assembly of the human genome (Wenger et al. 2019; Vollger et al. 2020), as well several other eukaryotic genomes (Jain et al. 2021; Song et al. 2021; Xue et al. 2021; Ríos-Touma et al. 2022; Sun et al. 2022a; Wang et al. 2022).

The problem of de novo assembly can be computationally challenging because of the high repetitive content of genomes, sequencing errors, nonuniform, or insufficient sequencing coverage and chimeric reads. In the literature, the problem is solved either using the overlap graph (Li et al. 2012), the de Bruijin graph (Miller et al. 2010; Li et al. 2012), or the string graph (Ben-Bassat and Chor 2014), depending on the nature and the number of the input reads. These methods also play an essential role in SMS assembly (Jain et al. 2021).

Assemblers for long SMS reads also use the overlap graph, the de Bruijin graph, and the string graph (or any combination thereof). For instance, Canu (Koren et al. 2017) integrates hybrid error correction PBcR (Koren et al. 2012), MinHash Alignment Process (MHAP), and some modules from the Celera Assembler (Berlin et al. 2015). miniasm (Li 2016) implements the overlap and layout steps in the overlap-layout-consensus assembly paradigm. The absence of the consensus step makes miniasm particularly fast, but its application is limited to relatively small genomes that are not very repetitive (Li 2016). HiCanu (Nurk et al. 2020) is a special version of Canu that leverages the high-quality of HiFi reads. FALCON (Chin et al. 2016) builds primary contigs via a string graph and then generates the haplotype-resolved assembly using phased reads. Shasta was designed for the assembly of human genome

[6]**These authors contributed equally to this work.**
Corresponding authors: panweihua@caas.cn, stelo@ucr.edu, weixiuqing47@foxmail.com
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.278232.123.

with Oxford Nanopore reads, but the developers have recently added a HiFi-mode (Shafin et al. 2020). Peregrine (Mills et al. 2020) takes advantage of the high accuracy of HiFi reads and uses sparse hierarchical minimizers to index reads, thereby reducing the high computational cost of all-against-all alignment step in the overlap-layout-consensus pipeline. hifiasm was specifically designed for HiFi reads: It uses a phased assembly graph to reconstruct the haplotype of diploid genomes (Cheng et al. 2021). hifiasm-meta represents a variant of hifiasm devised to perform the assembly of multiple genomes present within a metagenomic sample. This tool incorporates a novel read selection step and introduces innovative criteria aimed at safeguarding reads that originate from genomes with limited coverage (Feng et al. 2022). ABruijn (Lin et al. 2016), Flye (Kolmogorov et al. 2019), HiFlye, and metaFlye (Kolmogorov et al. 2020) are based on the de Bruijn graph. ABruijn combines de Bruijin and the overlap-layout-consensus approaches. Flye uses the repeat graph, which extends the de Bruijn graph so that it can deal with the errors in long reads. wtdbg2 also uses a combination of overlap graph and de Bruijin graph (Ruan and Li 2020): It uses a data structure called fuzzy-Bruijn graph to enable an efficient all-versus-all read alignment. MECAT and NECAT are assemblers for PacBio reads and ONT reads, respectively, that use a fast-scoring method to filter out spurious alignments (Xiao et al. 2017; Chen et al. 2021). Verkko is an improved Canu assembler that can assemble HiFi reads and ONT reads simultaneously and was used for the telomere-to-telomere assembly of the human genome (Rautiainen et al. 2023).

At the time of writing, users can choose from at least 25 assemblers for SMS reads (see Supplemental Table 1), depending on the type of reads they have (Chin et al. 2016; Li 2016; Lin et al. 2016; Kamath et al. 2017; Koren et al. 2017; Wick et al. 2017; Xiao et al. 2017; Nowoshilow et al. 2018; Du and Liang 2019; Kolmogorov et al. 2019, 2020; Mills et al. 2020; Nurk et al. 2020; Ruan and Li 2020; Shafin et al. 2020; Chen et al. 2021; Cheng et al. 2021; Luo et al. 2021; Feng et al. 2022; Rautiainen et al. 2023). However, choosing the "best assembler" for their data is a daunting proposition, because the performance of an assembler depends on organism ploidy, genome repetitive content, genome size, heterozygosity, and many other factors. Even if we focus only on HiFi assemblers, there is no comprehensive study that could guide users on the expected performance of these assemblers on large eukaryotic genomes. The only studies we could find were those of (1) Zhang et al. (2022a), who tested Flye, HiCanu, hifiasm, NECAT, and NextDenovo on HiFi and ONT reads, but only on brewer's yeast; and (2) Gavrielatos et al. (2021), who tested Canu, hifiasm, WENGAN (Di Genova et al. 2021), and HiCanu on Hifi and ONT reads, but only on the fruit fly and human haploid genome and in the context of hybrid assembly.

To address this shortcoming, here we report on a comprehensive assessment of the performance of 11 SMS assemblers for HiFi reads (nine for genome assembly and two for metagenomes). Our choice of these 11 assemblers was based on their (1) popularity (based on their usage/citations), (2) user friendliness (e.g., how easy it is to install and run them), (3) algorithmic novelty, and (4) the fact that they are currently actively maintained. The 11 HiFi assemblers we selected are HiCanu, hifiasm, HiFlye, hifiasm-meta, metaFlye, Peregrine, Shasta, Verkko, MECAT2, miniasm, and NextDenovo. We studied the performance of these assemblers under various conditions, including various sequencing coverage, heterozygosity, and ploidy (see Supplemental Table 2). The 11 assemblers were evaluated using quality assessment tool (QUAST/MetaQUAST) (Mikheenko et al. 2016, 2018) and benchmarking universal single-copy ortholog (BUSCO) (Simão et al. 2015). We also measured new quality metrics, namely, completion rate, single-copy completion rate, duplicated completion rate, average proportion of largest category, average distance difference, and quality value. We also recorded run-time and memory utilization.

Results

A comprehensive set of experiments were conducted on the 11 genome assemblers using both real and simulated data for various choices of ploidy. Four assemblers were selected for a deeper analysis on data sets produced for different choices of sequencing coverage and heterozygosity. Finally, four metagenome assemblers were tested using both real and simulated metagenomics samples for several choices of the sample composition, in terms of both species abundance and the similarity among the constitutive genomes.

Experiments on complex eukaryotic genomes

Experimental results on real data with varying ploidy

All the assemblers were tested on real HiFi reads for rice (homozygous diploid), potato (heterozygous diploid), and wax apple (autotetraploid). Detailed statistics on these data sets and the calculation methods of evaluation criteria are provided in the Methods section.

First, the assembly contiguity was assessed by QUAST. Figure 1A shows the cumulative total size of contigs and the number of contigs that have a size in the range encoded by the color in the legend, for the rice data set (top), the potato data set (middle), and the wax apple data set (bottom). Observe that on potato and wax apple data, HiCanu, hifiasm, HiFlye, and Peregrine produced longer, more contiguous assemblies. In particular, hifiasm produced a larger proportion of contigs >10 Mbp (blue area in Fig. 1A). Verkko produced a long assembly composed primarily of relatively short contigs. MECAT2 instead produced the shortest assemblies. In Figure 1C, we ordered the assembled contigs by size and computed the cumulative contig length for different thresholds of the NG value. Observe that on all three data sets, hifiasm achieved the best contiguity.

Second, the genome completeness was assessed by BUSCO. The set of conserved genes in *gramineae* (4896 genes in total) was used to assess the rice assemblies; the set of conserved genes in *cruciferae* (5878 genes) was used to assess the potato assemblies; and the set of *eukaryotic* conserved genes (5878 genes) was used to assess the wax apple. Figure 1B shows the BUSCO assessment results for rice (top), potato (middle), and wax apple (bottom). Observe that HiCanu, hifiasm, HiFlye, NextDenovo, Peregrine, and Verkko achieved >98% completeness. According to this metric, miniasm did not perform well on rice and potato.

Third, the assembly accuracy was assessed using the QV score. The QV scores for all the assemblers on the three data sets are shown in Figure 1D. Observe that HiCanu, hifiasm, NextDenovo, Shasta, and Peregrine produced high QV across the three data sets. In contrast, MECAT2 and miniasm produced low accuracy assemblies on all data sets. Shasta generated poor accuracy assemblies on the wax apple, whereas its performance on rice and potato was satisfactory. It is noteworthy that Verkko had a worse QV score than did HiCanu, despite the fact that it shares some of its codebase.

Fourth, CPU time and memory usage was collected. Figure 1, E and F, shows the run time and the memory usage for all

assemblers on the three data sets. Observe that Shasta and NextDenovo consumed the lowest amount of resources. Peregrine was the slowest on rice and the wax apple data sets. MECAT2 was also slow, especially on the potato and the wax apple data set. NextDenovo and HiCanu used the smallest amount of memory. miniasm and MECAT2 used a much higher amount of memory, and the usage increased on higher-ploidy data sets.

Overall, hifiasm, HiCanu, HiFlye, and Peregrine showed a clear advantage over the other assemblers in terms of contiguity, completeness, and accuracy. In this group, however, Peregrine required much higher computational resources, in particular related to CPU time. NextDenovo achieved good results on the rice data sets but did not perform equally well on the potato and wax apple genome.

Experimental results on synthetic data sets with varying ploidy

On the synthetic data set, a more comprehensive evaluation of contiguity, completeness, and accuracy was performed owing to the availability of the "ground truth" genome (i.e., one to four copies of the rice genome after introducing SNPs and structural variants; detailed statistics on these synthetic data sets are provided in the Methods). Figure 2 summarizes the experimental results of the nine HiFi assemblers on synthetic reads produced from the rice genome in haploid form (one copy of each chromosome), synthetic diploid (two copies), and synthetic tetraploid (four copies).

Figure 2A shows the cumulative total size of contigs and the number of contigs that have a size in the color-coded range, for the rice haploid data set (top), the rice diploid data set (middle), and the rice tetraploid data set (bottom). Observe that hifiasm, Peregrine, HiCanu, Verkko, and Shasta produced highly contiguous assemblies in all three data sets, with a high fraction of contigs >10 Mbp (Fig. 2A, left, blue subbars). HiFlye had the worst performance on all three synthetic data sets. A deeper analysis to explain HiFlye's poor performance was conducted in the subsection "HiFlye tested on synthetic data sets on varying sequencing error rates." miniasm and MECAT2 had an adequate performance only on the haploid data set, possibly because these assemblers were not designed to handle polyploid data sets. The results in Figure 2A were consistent with the NG curves in Figure 2C. In this latter figure, hifiasm, Peregrine, Verkko, and HiCanu produced a higher NG50 on all synthetic data sets. Supplemental Table 3 reports QUAST's evaluation of the assemblies, including the number and length of misassembled contigs, the duplication ratio (which measures redundant contigs), the fraction of the genome covered by the assembly, and the number of mismatches/indels in the assembly. Consistent with results above, hifiasm, Peregrine, Verkko, and HiCanu achieved genome fraction >99%, with the duplication ratio very close to 1.0 and a low number of mismatches/indels and misassemblies. HiFlye, miniasm, and MECAT2 did not perform as well.

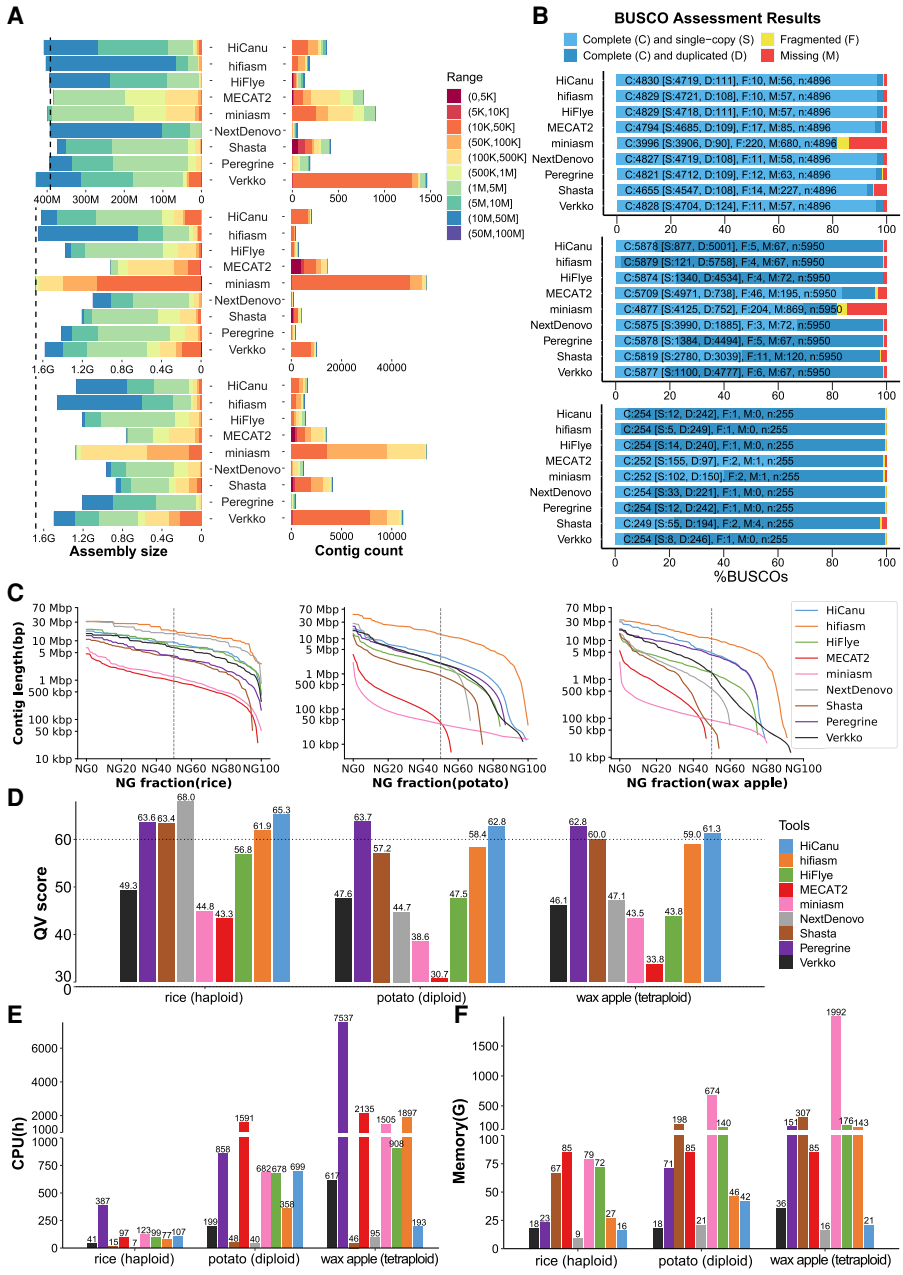

Figure 1. Summary of the performances of the selected genome assemblers on PacBio HiFi reads for rice (haploid), potato (diploid), and wax apple (tetraploid). (*A*) Cumulative total size of contigs (*left*) and number of contigs (*right*) that have a size in the range encoded by the color in the legend (*top*: rice; *middle*: potato; *bottom*: wax apple); the vertical dashed line indicates the expected genome size. (*B*) BUSCO completeness scores (*top*: rice; *middle*: potato; *bottom*: wax apple). (*C*) Contig length distribution for various choices of the NGx fraction threshold. (*D*) Quality value scores. (*E*) Running time analysis. (*F*) Memory usage analysis (legend on panel *D* also applies to panels *E*,*F*).

In Figure 2B, we used the three-completeness metrics described in the Methods section, namely, CR, SCR, and DCR, to evaluate the performance of the assemblers. HiCanu, Verkko, and hifiasm had the best performance according to these three metrics, achieving a completeness rate of ~99.9%. NextDenovo and MECAT2 performed well on haploid but were inadequate on diploid and tetraploid. HiFlye had the worst performance again.

Figure 2D shows the average proportion of the largest category score (APLC), whereas Figure 2E summarizes the average distance difference scores (ADF). Both accuracy metrics are defined in the Methods section. HiCanu, Peregrine, Verkko, and hifiasm performed well according to these criteria across varying ploidy, with APLC close to 1.0 and low values for ADF. On diploid and tetraploid data sets, other assemblers also had a small APLC. HiFlye and MECAT2 produced a high ADF, which is consistent with the large number of misassembled contigs detected by QUAST for these tools (Supplemental Table 3).

As expected, time and memory usages on synthetic data were similar to the usage on real data sets (Fig. 2F). Observe that Shasta and NextDenovo were the fastest on all synthetic data sets, whereas HiFlye, MECAT2, and Peregrine were the slowest. Also observe in Figure 2G that NextDenovo used the smallest amount of memory, whereas miniasm, Shasta, and MECAT2 used the largest amounts.

In summary, hifiasm, Peregrine, Verkko, HiCanu, and Shasta produced assemblies with higher contiguity, completeness, and accuracy than the other assemblers on synthetic HiFi reads data sets across varying ploidy. NextDenovo had a good performance only in the haploid data set. miniasm, NextDenovo, and MECAT2 failed on the diploid and tetraploid data sets. HiFlye failed on all data sets; a deeper analysis will be performed later.

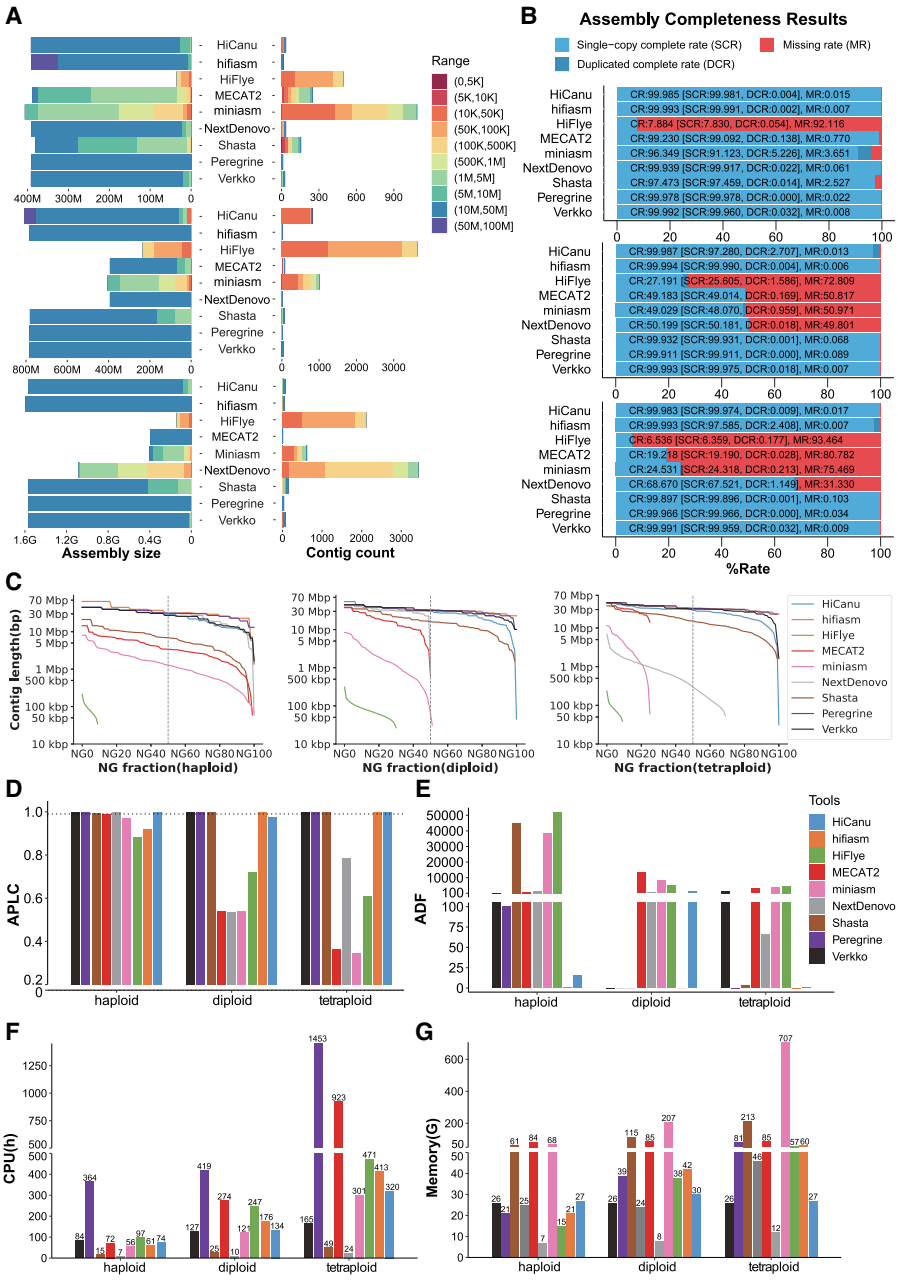

Figure 2. Summary of the performances of the selected genome assemblers on synthetic HiFi reads for rice in haploid form, synthetic diploid, and synthetic tetraploid. (*A*) Cumulative total size of contigs (*left*) and the number of contigs (*right*) that have a size in the range encoded by the color in the legend (*top*: haploid rice; *middle*: synthetic diploid rice; *bottom*: synthetic tetraploid rice); the vertical dashed line indicates the expected genome size. (*B*) Single complete rate, duplicated complete rate, and missing rate (*top*: haploid rice; *middle*: synthetic diploid rice; *bottom*: synthetic tetraploid rice). (*C*) Contig length distribution for various choices of the NGx fraction threshold. (*D*) Average proportion of largest category (APLC); the horizontal dashed line is APLC = 0.99. (*E*) Average distance difference (ADF). (*F*) Running time analysis. (*G*) Memory usage analysis (legend on panel *E* also applies to panels *F*,*G*).

Experimental results on a human data set

We also performed a performance evaluation of HiCanu, hifiasm, HiFlye, Peregrine, Shasta, Verkko, MECAT2, miniasm, and NextDenovo on a human data set. On this large data set, miniasm did not complete owing to memory overflow, and Peregrine did not yield results after running it for 2 wk on 80 threads. We removed miniasm and Peregrine from the evaluation and added two new assemblers, namely, LJA (Bankevich et al. 2022) and rustmdbg (Ekim et al. 2021). Given the size of the human genome, when computing the CR, SCR, DCR, APLC, and ADF assembly scores, we sampled 300,000 unique *k*-mers from the reference genome uniformly at random 10 times and recorded the average.

A summary of the performances of these nine assemblers is reported in Figure 3. Observe in Figure 3B that HiCanu, hifiasm, LJA, and Verkko achieved a completeness close to 100%, whereas the

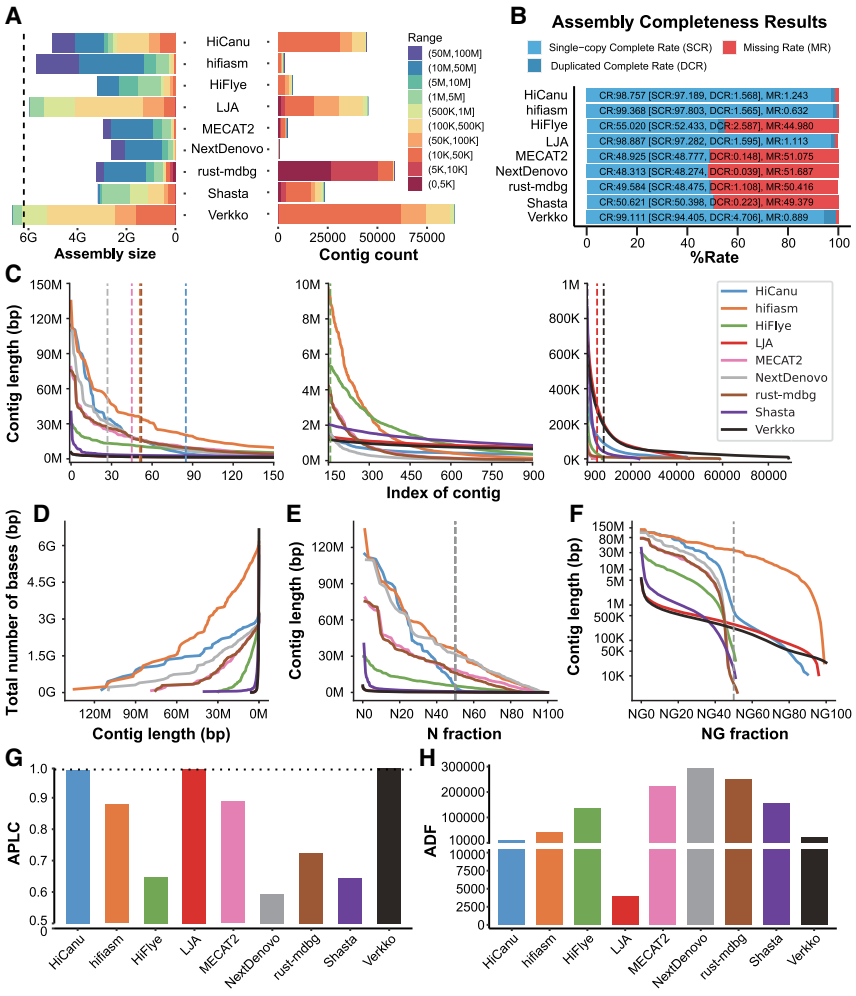

Figure 3. Summary of the performances of the selected genome assemblers on a human genome data set. (*A*) Cumulative total size of contigs (*left*) and the number of contigs (*right*) that have a size in the range encoded by the color in the legend; the vertical dashed line indicates the expected genome size. (*B*) Single-copy complete rate, duplicated complete rate, and missing rate. (*C*) Sorted contig length (*left*: longest 150; *middle*: index 150–900; *right*: index 900–80,000). (*D*) Cumulative assembly size as a function of the minimum contig length allowed in the assembly. (*E*) Nx length (the dashed line denotes the N50). (*F*) NGx length (the dashed line denotes the NG50). (*G*) Average proportion of largest category (APLC); the horizontal dashed line is APLC = 0.99. (*H*) Average distance difference (ADF).

other assemblers only achieved 50% completeness. The 50% completeness is likely because some of these assemblers generate consensus assemblies rather than haplotype-resolved assemblies. Also observe that Verkko's assembly contained a significant amount of redundant contigs compared with HiCanu, hifiasm, and LJA.

Among the four assemblers with high completeness, HiCanu, LJA, and Verkko achieved higher accuracy than did hifiasm (Fig. 3G,H). However, the contiguity (N50) of hifiasm (36.4 Mbp) was significantly higher than that of HiCanu (4.5 Mbp), LJA (299.0 kbp), and Verkko (190.5 kbp) (Fig. 3A,C–F). Overall, hifiasm produced the best performance on this data set.

Experimental results on synthetic data sets with varying sequencing coverage levels

An important quality of genome assembler is the ability to produce good assemblies even when the sequencing coverage is less than optimal. In this section, we tested HiCanu, Verkko, hifiasm, and HiFlye using synthetic data sets from the synthetic heterozygous diploid genome with sequencing depths of 10×, 20×, 30×, and 50×. Detailed statistics on the synthetic data sets are provided in the Methods section. Experimental results are summarized in Figure 4 and tabulated in Supplemental Table 4.

Several observations are in order. First, irrespective of the assemblers, there was a significant improvement on the assembly contiguity when the coverage increased from 10× to 20× (Fig. 4C–F). HiCanu's N50 increased from 362,820 bp to 30,836,250 bp; Verkko's N50 increased from 816,272 bp to 32,114,812 bp; and hifiasm increased from 6,168,309 bp to 32,623,532 bp. At the same time, the ADF of all assemblers decreased when the coverage increased from 10× to 20×. Verkko's ADF continued to decrease when coverage increased from 20× to 30×, but it was not the case for HiCanu and hifiasm. When the coverage increased from 10× to 20×, NA50 and NGA50 improved about 80 times for HiCanu, 40 times for Verkko, and five times for hifiasm. From 20× to 50×, the change of N50, NA50, and NGA50 were less pronounced.

Supplemental Table 4 and Figure 4B show that when the coverage increased from 10× to 20×, the assembly completeness for HiCanu and Verkko increased, whereas when the coverage increased from 20× to 50×, the assembly completeness did not change significantly. Also observe that the increase in coverage did not affect hifiasm's assembly completeness, which was stable over different coverages. HiFlye had better completeness with increasing coverage, but overall, HiFlye's assemblies were much worse than the other three assemblers.

The accuracy assessment was based on APLC, ADF, mismatches per 100 kbp, and number/length of misassembled contigs. The APLC for HiCanu increased with higher coverage. hifiasm was able to produce more contigs >50 Mbp (purple color in Fig. 4A) with 20× and 30× coverage; however, Supplemental Table 4 and the APLC indicate a misassembly on those long contigs. With 50× coverage, hifiasm corrected this mistake (Fig. 4G; Supplemental Table 4). With increased coverage Verkko produced an improvement in ADF (Fig. 4H).

In summary, HiCanu and Verkko's assemblies were more sensitive to the sequencing coverage, and they had an unsatisfactory performance at 10×. Instead, hifiasm had a more predictable and consistent performance across different choices of coverage, and its assemblies improved with increasing coverage.

Experimental results on synthetic data sets with varying heterozygosity rates

Another important quality of a genome assembler is the ability to deal with various levels of heterozygosity in diploid (or polyploid)

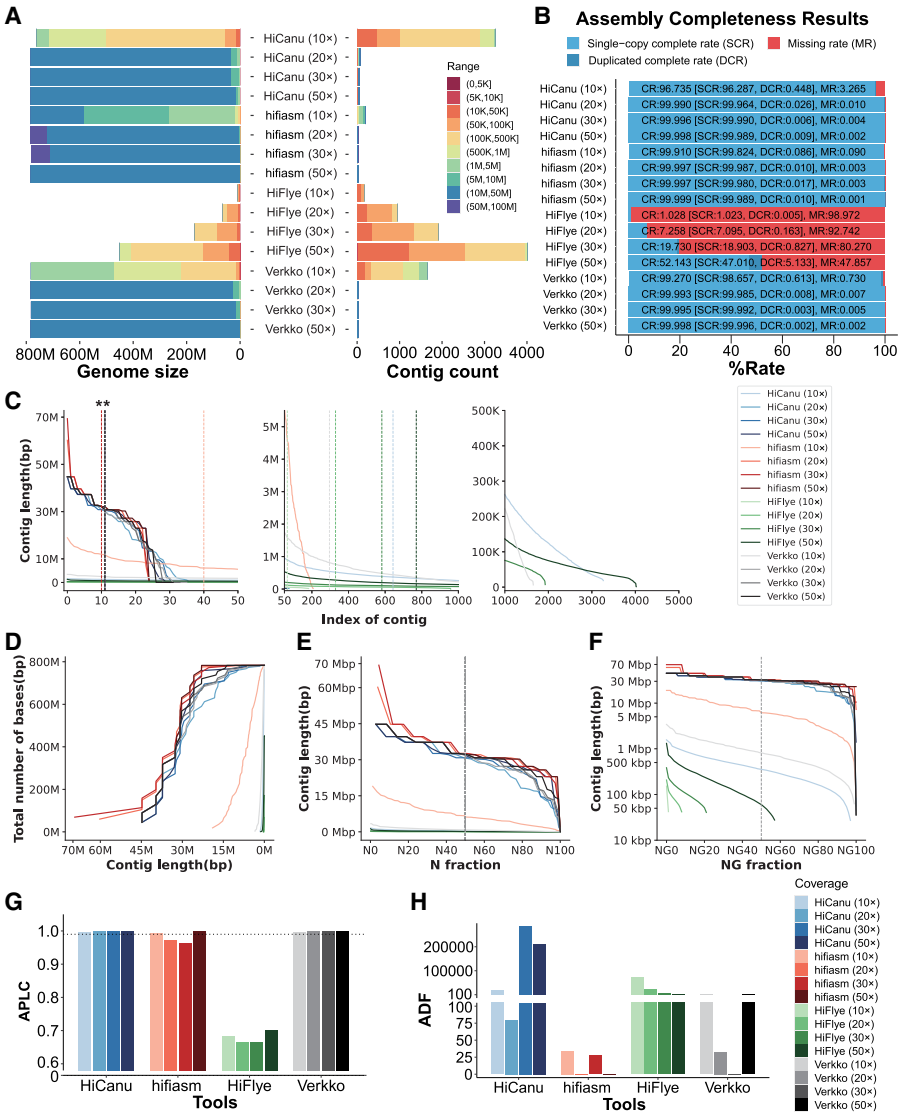

Figure 4. Summary of the performances of HiCanu, Verkko, hifiasm, and HiFlye on synthetic HiFi reads with coverage at 10×, 20×, 30×, and 50×. (*A*) Cumulative total size of contigs (*left*) and the number of contigs (*right*) that have a size in the range encoded by the color in the legend. (*B*) Single complete rate, duplicated complete rate, and missing rate. (*C*) Sorted contig length (*left*: longest 50; *middle*: index 50–1000; *right*: index 1000–5000). The horizontal lines represent the corresponding L50. (*) Overlapping data. (*D*) Cumulative assembly size as a function of the minimum contig length allowed in the assembly. (*E*) Nx length (the dashed line denotes the N50). (*F*) NGx length (the dashed line denotes the NG50). (*G*) Average proportion of largest category (APLC); the horizontal dashed line is APLC = 0.99. (*H*) Average distance difference (ADF).

genomes. In this section, we used the synthetic heterozygous diploid genome and generated synthetic reads at 20× coverage by varying the heterozygosity rates. Specifically, we used the SimSID script to introduce heterozygosity rates of 0.5%, 1.0%, 1.5%, and 2.5% (see Methods). Detailed statistics for the synthetic data sets are provided in the Methods section. Experimental results for HiCanu, Verkko, hifiasm, and HiFlye on these data sets are summarized in Figure 5 and tabulated in Supplemental Table 5.

Several observations are in order. In terms of contiguity, Figure 5A shows that HiCanu's assembly contiguity degraded with higher levels of heterozygosity, whereas hifiasm's had a stable performance across all data sets. In fact, hifiasm had a better performance on all metrics compared with the other assemblers (Fig. 5C–F). HiFlye again produced incomplete, small assemblies on all synthetic data sets. In terms of completeness (CR, SCR, DCR), hifiasm, HiCanu, and Verkko had consistently good performance with varying heterozygosity rates (Fig. 5B).

With respect to assembly accuracy, there was no clear trend associated with increasing levels of heterozygosity (Fig. 5G, H). hifiasm's number of mismatches per 100 kbp assessed by QUAST was less than 0.1 in all data sets (Supplemental Table 5). However, HiCanu's APLC was better than that of hifiasm. hifiasm's lower APLC was caused by a single misassembly on a 70-Mbp contig (Supplemental Fig. 2). hifiasm produced the highest accuracy assemblies based on the ADF metric.

In summary, although HiCanu and HiFlye's performance degraded with increasing heterozygosity, Verkko and hifiasm had a consistently good performance across data sets. hifiasm, Verkko, and HiCanu performed well on low heterozygosity data sets. hifiasm performed well also on high heterozygosity data sets.

HiFlye tested on synthetic data sets on varying sequencing error rates

As reported previously, HiFlye produced high-quality assemblies on the three real data sets but unsatisfactory and inconsistent results on synthetic data sets. We hypothesized that HiFlye could be very sensitive to the sequencing error rate. To test this hypothesis, we started from (1) a real homozygous diploid rice genome, (2) a synthetic heterozygous diploid genome, and (3) synthetic autotetraploid genome (see Methods), and generated synthetic data sets with different error rates using PBsim associate with the error models learned from on real HiFi data sets. The list of data sets used for learning the error models are shown in the Methods section. HiFlye's results on these synthetic data sets are shown in Supplemental Table 6.

As hypothesized, the assembly results of HiFlye decreased with higher sequencing error rates. Lower sequencing error rates led to improved completeness and contiguity in the assembly results. Specifically, at a read accuracy of 99.8474%, N50s and NA50s were consistently <100 kbp for different ploidy. However, at a significantly higher accuracy of 99.99%, HiFlye encountered memory issues when handling diploid and tetraploid genomes. However, the reason behind this memory limitation is unclear.

Experimental results on metagenomic samples

For the metagenome assembly evaluation, we compared the performance of two metagenomic assemblers, namely, hifiasm-meta

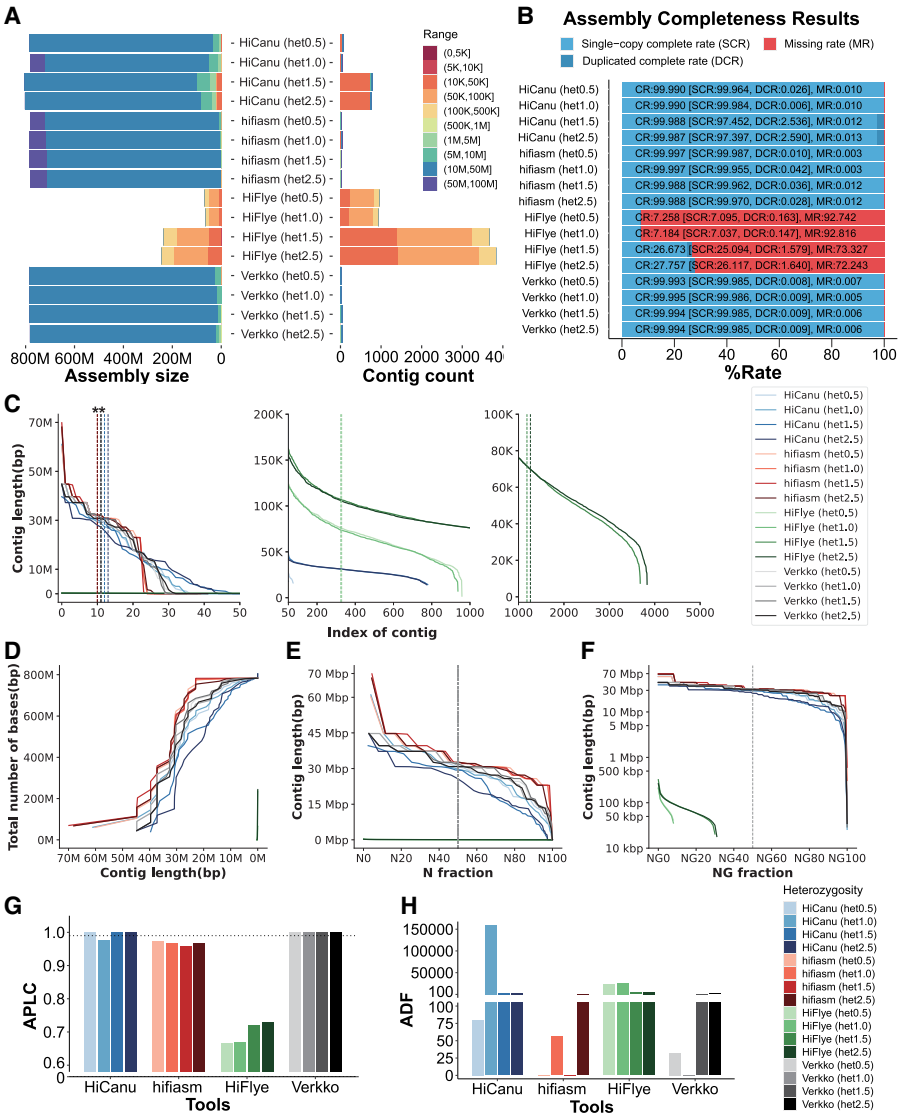

Figure 5. Summary of the performances of HiCanu, Verkko, hifiasm, and HiFlye on synthetic HiFi reads with heterozygosity rates of 0.5%, 1.0%, 1.5%, and 2.5%. (*A*) Cumulative total size of contigs (*left*) and the number of contigs (*right*) that have a size in the range encoded by the color in the legend. (*B*) Single complete rate, duplicated complete rate, and missing rate. (*C*) Sorted contig length (*left*: longest 50 contigs; *middle*: index 50–1000; *right*: index 1000–5000). The horizontal lines represent the corresponding N50. (*) Overlapping data. (*D*) Cumulative assembly size as a function of the minimum contig length allowed in the assembly. (*E*) Nx length (the dashed line denotes the N50). (*F*) NGx length (the dashed line denotes the NG50). (*G*) Average proportion of largest category (APLC); the horizontal dashed line is APLC = 0.99. (*H*) Average distance difference (ADF).

(Feng et al. 2022) and metaFlye (Kolmogorov et al. 2020), against two general assemblers, namely, HiCanu (Nurk et al. 2020) and NextDenovo.

On real metagenomic samples, the quality of assemblies was evaluated in three steps. First, the completeness and contamination of each assembled contig were assessed using CheckM (Parks et al. 2015), using marker genes that are specific to each species inferred lineage within a reference genome tree. Second, CheckM and the Genome Taxonomy Database Toolkit (GTDB-Tk) (Chaumeil et al. 2020, 2022) were used to identify and classify the assembled bacterial contigs in a reference genome tree; assemblies were compared using the numbers of identified contigs at different taxonomic levels. Third, the numbers of conserved 16s ribosomal RNA (rRNA) and 16s rRNA clusters were used to evaluate the completeness of the metagenomic assemblies; bacterial and archaeal ribosomal RNA predictor (Barrnap) was used to detect the 16s rRNA sequences, and VSEARCH (Rognes et al. 2016) was used for clustering 16s rRNAs (using a minimum of 97% identity).

On synthetic data, the quality of assemblies was evaluated with MetaQUAST (Mikheenko et al. 2016) by comparing them against the "ground truth" microbial genomes.

Experimental results of real metagenomic data

As mentioned above, the sheep gut metagenome data set (Bickhart et al. 2022) was used to test the four assemblers. A summary of the statistics of this data set is provided in the Methods section. Experimental results are illustrated in Figure 6 and tabulated in Supplemental Table 7. Observe in Supplemental Table 7 that hifiasm-meta produced the assembly with the largest total size, the largest number of contigs, the largest number of contigs >1 Mbp, and the largest number of circular contigs >1 Mbp. In addition, hifiasm-meta's assembly contained twice or more 16s rRNAs compared with the other assemblies, and it identified the largest number of rRNA clusters. Observe in Figure 6A that hifiasm-meta produced the highest number of high-quality contigs, that is, contigs with a CheckM completeness rate >90% and a CheckM contamination rate <10%. HiCanu produced the second highest number of contigs with a completeness rate >90%. metaFlye generated a large number of contigs but with comparatively lower quality. In terms of contamination, metaFlye worked best with an average contamination rate <1% (Fig. 6B). Observe in Figure 6C that the cumulative length of hifiasm-meta assembly was about two times larger than the second largest one. metaFlye, HiCanu, and NextDenovo generated a similar count of long contigs (>1 Mbp). Although NextDenovo produced fewer short contigs, its N50 was significantly higher than the other assemblers. To find out if the high N50 was solely owing to the small number of short contigs, we sorted the contigs from the longest to shortest and drew the contig length distribution in Figure 6D. Observe that the curve for NextDenovo is always below those of the other assemblers, which confirms our hypothesis. hifiasm-meta's curve is always above those of other assemblers, whereas metaFlye is in "second place."

To further compare the assemblies, we used CheckM in conjunction with GTDB-Tk to perform the taxonomic identification

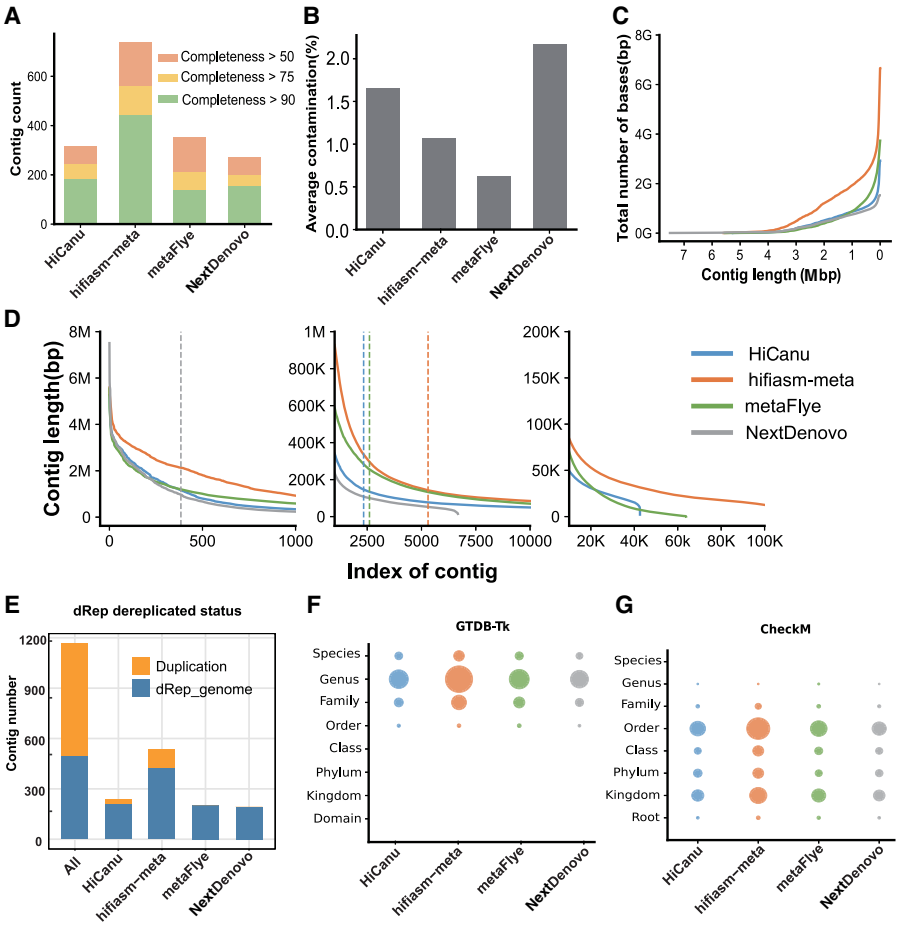

Figure 6. Summary of the performances of HiCanu, hifiasm-meta, metaFlye, and NextDenovo on the sheep gut metagenome data set. (*A*) Number of assembled contigs (contamination < 10%) for different levels of completeness; completeness was calculated by CheckM. (*B*) Average contamination rate calculated by CheckM. (*C*) Cumulative assembly size as a function of the minimum contig length allowed in the assembly. (*D*) Sorted contig length (*left*: longest 1000; *middle*: index 2500–10,000; *right*: index 20,000–100,000). (*E*) Duplication analysis. The orange bars indicate the number of duplicate contigs removed by the tool dRep; the blue bars represent the nonredundant contigs. (*F*) Taxonomic classification of contigs >1 Mbp with GTDB-Tk; the size of each circle represents the number of contigs identified at each taxonomic level. (*G*) Taxonomic classification of contigs >1 Mbp with CheckM; the size of each circle represents the number of contigs identified at each taxonomic level.

and classification for the assembled contigs >1 Mbp. Although the CheckM tends to classify contigs to higher taxonomic levels (mostly from order to kingdom) whereas GTDB-Tk tends to classify to lower ones (mostly species to family), they both indicate that hifiasm-meta is capable of assembling the largest number of species (Fig. 6F,G). Details of this experiment are shown in Supplemental Figure 3.

Observe in Figure 6A that the number of high-quality contigs obtained by hifiasm-meta is significantly higher than other assemblers. To investigate whether other assemblers produced contigs missed by hifiasm-meta on real metagenomic data, (1) we merged all the contigs produced by all the assemblers, (2) we selected high-quality contigs, namely, contigs with contamination ≤ 10 and completeness ≥ 75 using CheckM, and (3) we removed duplicated contigs using dRep v3.4.3 (Olm et al. 2017). Figure 6E shows that there were a total of 1168 high-quality contigs across all assemblies, which reduced to 497 after deduplication. In contrast, hifiasm-meta yielded 536 high-quality contigs, which reduced to 428 after deduplication. Although hifiasm-meta missed ~14% of the contigs in the merged assembly, we still recommend the use of hifiasm-meta for metagenomic assembly of HiFi data sets.

Experimental results of synthetic metagenomic data

Next, we tested the four assemblers on the synthetic data set described in the Methods section. Completeness, contiguity, and accuracy of the assemblers were evaluated using MetaQUAST and the five additional criteria described in the Methods section. Figure 7 illustrates the experimental results. Observe in Figure 7B that all assemblers achieved a low single-copy completeness rate (SCR) and duplicated completeness rate (DCR; <50%), which may be because some unique *k*-mers are missing in the synthetic reads (Supplemental Fig. 4) and the nonuniform abundances of species in the sample. hifiasm-meta produced the most contiguous and complete assembly among all the tools, followed by metaFlye and HiCanu. HiCanu and hifiasm generated assemblies with better accuracy than other two assemblers, including higher average proportions of the largest category (APLC), lower average distance differences (ADF), lower misassembled rates, lower numbers of mismatches per 100 kbp, and lower numbers of indels per 100 kbp (Fig. 7A–J; Supplemental Table 8).

Comparisons on different taxonomic levels

To compare the performance of the two metagenomic assemblers (hifiasm-meta and metaFlye), we tested them on several synthetic data sets (described in the Methods section) for various choices of the sequence similarities.

Experimental results in Figure 8B and Supplemental Table 9 show that assemblies produced by hifiasm-meta had higher completeness (i.e., the genome fraction computed by MetaQUAST) than that of metaFlye's assemblies. Also observe that the completeness of metaFlye's assemblies declines faster than hifiasm-meta's, as the similarity of genomes increases. However, hifiasm-meta assemblies had a genome fraction >100%, which indicated more redundant sequences than those of metaFlye. This redundancy was also reflected in the completeness criteria based on unique *k*-mer (Fig. 8G).

Figure 8, A and C–F, shows that hifiasm-meta's assemblies are more accurate than those of metaFlye. As the sequence similarity of genomes increases, the accuracies of hifiasm-meta and metaFlye both decline sharply in terms of ADF, the number of indels per 100 kbp, and the number of mismatches per 100 kbp. However, for both hifiasm-meta and metaFlye, APLC does not change significantly at different taxonomic levels. The low misassembled ratio for hifiasm-meta is partially because hifiasm-meta produced a much higher number of contigs compared with

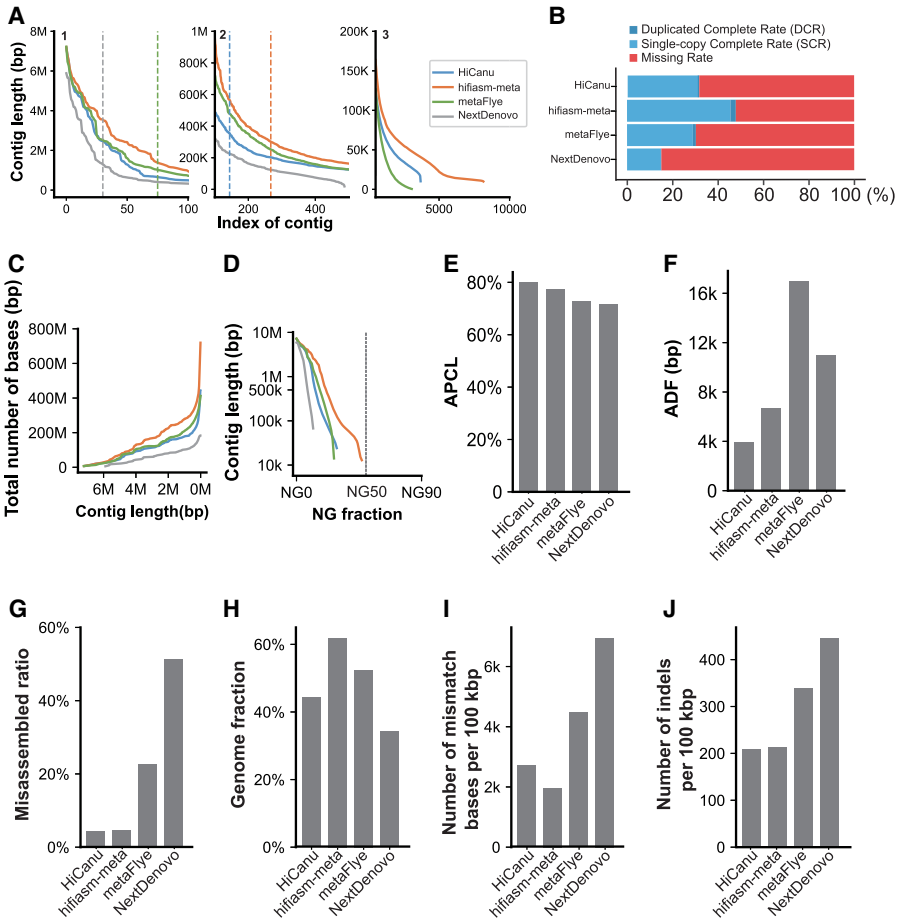

Figure 7. Summary of the performances of HiCanu, hifiasm-meta, metaFlye, and NextDenovo on a synthetic data set. (*A*) Sorted contig length (*left*: index 1–100; *middle*: index 1–500; *right*: index 1–10,000); the dashed line represents the N50 contig length. (*B*) Single-copy completeness rate (SCR) and duplicated completeness rate (DCR). (*C*) Cumulative assembly size as a function of the minimum contig length allowed in the assembly. (*D*) NGx length (the dashed line denotes the NG50). (*E*) Average proportion of largest category (APLC). (*F*) Average distance difference (ADF). (*G*) Misassembled contig rate computed by MetaQUAST. (*H*) Genome fraction computed by MetaQUAST. (*I*) Number of mismatches per 100 kbp computed by MetaQUAST. (*J*) Number of indels per 100 kbp computed by MetaQUAST.

metaFlye. In addition, those contigs were short, and short contigs are more likely to be assembled correctly. The absolute number of misassembled contigs and misassembly count is shown in Figure 8, H and I. Observe that the absolute numbers of misassembled contigs and the misassembly counts are comparable at the phylum and class levels, but they are significantly different on family level. The total length of correct contigs are in Figure 8J.

Figure 9 shows the contiguity results at different taxonomic levels for the two assemblers. Observe that at the phylum and class levels, hifiasm-meta and metaFlye showed comparable performance in terms of contig length. However, because hifiasm-meta retained more short-contigs, it achieved a higher total length compared with metaFlye, resulting in higher completeness. At the family and genus levels, hifiasm-meta produced assemblies with a significantly better contiguity compared with metaFlye (Fig. 9A–M).

Discussion

As mentioned above, choosing the "best" assembler to assemble a new genome is a daunting proposition, because the performance of an assembler often depends on the genome ploidy, repetitive content, size, and heterozygosity, as well as many other factors. To the best of our knowledge, there are very few comprehensive studies that can guide users on the choice of the most appropriate assembler for their data. Even if we focus only on HiFi data, we are not aware of any comprehensive study that compares the performance of all modern HiFi assemblers on large eukaryotic genomes. As mentioned in the introduction, studies (Gavrielatos et al. 2021; Zhang et al. 2022a) are exclusively focused on the yeast, *Drosophila*, and human genomes.

In this paper, we addressed this shortcoming by performing an extensive set of experiments to assess the performance of the most popular genome assemblers for HiFi reads. Several real and synthetic data sets for complex eukaryotic genomes and metagenomes were used to test the contiguity, completeness, and accuracy of 11 assemblers. We hope that our data sets (which include HiFi reads for a newly sequenced tetraploid, the wax apple) will become a benchmark for future assembler development. Five novel *k*-mer-based criteria were introduced to help assess genome assemblies' quality.

Our concluding remarks are in order. Overall, hifiasm performed consistently well across all experiments with varying ploidy, coverage, and heterozygosity. hifiasm was also the least sensitive to the sequencing coverage, as long as it was higher than 20×. It clearly ranked first in the overall performance.

HiCanu also produced good assemblies, albeit not as good as hifiasm. Overall, it ranked second among all the assemblers we tested. On real data, HiCanu produced assemblies of a quality similar to that of hifiasm. Although HiCanu's accuracy was as high as hifiasm, HiCanu's contiguity was often lower than that of hifiasm. Similar results were obtained on synthetic data sets.

The performance of Verkko was not as good as HiCanu despite the fact that they share a similar codebase. Verkko produced assemblies with lower N50, lower largest contig, and lower QV. We speculate that the difference is owing to the different data structure used by Verkko (de Bruijn graph) compared with HiCanu (overlap graph).

Unfortunately, HiFlye failed on all simulated data sets. Our experiments seem to indicate that HiFlye is very sensitive to the sequencing error rate of the input reads. This mediocre performance seems to contradict the results of Zhang et al. (2022a), which show that HiFlye outperformed HiCanu, hifiasm, NextDenovo, and Flye on HiFi reads for baker's yeast. Three factors could explain the difference between our experimental results and those of Zhang et al. (2022a). First, HiFlye might work better on the yeast genome because of its relatively small size and small amounts of repetitive

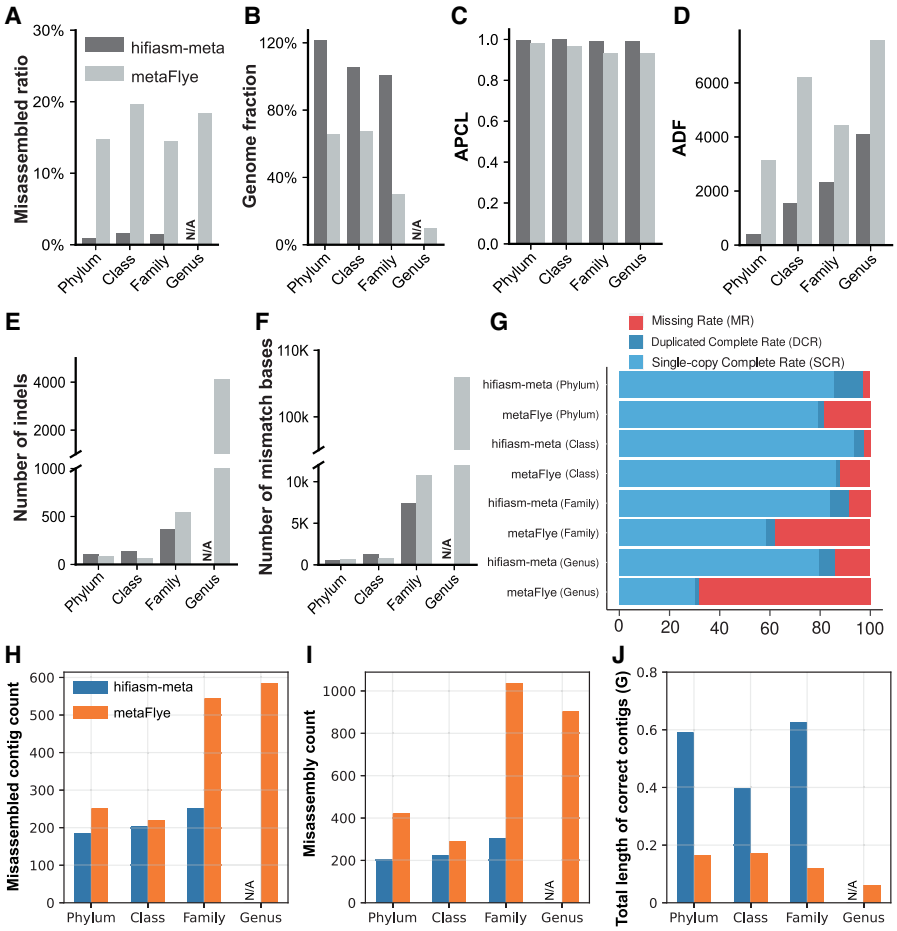

Figure 8. Comparison of hifiasm (dark gray) and metaFlye (light gray) on data set at the phylum, class, family, and genus levels. (*A*) Misassembled contig rate from MetaQUAST. (*B*) Genome fraction from MetaQUAST. (*C*) Average proportion of largest category (APLC). (*D*) Average distance difference (ADF). (*E*) Number of indels per 100 kbp from MetaQUAST. (*F*) Number of mismatches per 100 kbp from MetaQUAST. (*G*) Unique *k*-mer-based complete rates (including single complete rate, duplicated complete rate, and missing rate). At the genus level, MetaQUAST was not able to generate evaluation results for the hifiasm-meta, leading to the missing bars in the figures. (*H*) Misassembled contig count in different taxonomic levels. (*I*) Misassembly (a concept defined by QUAST representing the breakpoint in the alignment between assembly of reference) count in different taxonomic levels. (*J*) Total length of correct contigs in different taxonomic levels. At the genus level, owing to the ultra-high computational consumption, QUAST result of hifiasm-meta cannot be obtained within limited time and thus is represented as "N/A."

content. In our experiments, HiFlye worked better on the rice genome than more complex genomes such as the potato and wax apple. Second, the yeast HiFi data used by Zhang et al. (2022a) might have had a lower sequencing error rate compared with ours. Our experiments show that HiFlye is extremely sensitive to sequencing errors. Third, Zhang et al. (2022a) used a reference yeast genome that is not the same strain as the one sequenced with PacBio HiFi, which could have led to evaluation inaccuracies.

Shasta and Peregrine had good performance across all ploidy on real data sets. Their main limitation is the high resource consumption both in time and memory occupation. NextDenovo and MECAT2 only showed good performance in haploid data sets. On real data sets, miniasm produced genome assemblies of a size similar to that of hifiasm but had lower contiguity and lower accuracy, mainly with respect to BUSCO and QV.

hifiasm-meta performed consistently well on real metagenomic data and synthetic metagenome data sets at different taxonomic levels. NextDenovo had the poorest performance in terms of genome size, contiguity, and accuracy (e.g., highest contamination rate) on the real metagenomic data sets. Although HiCanu is not specifically designed for metagenome assembly, it performed comparably well to metaFlye, particularly in synthetic data sets, achieving higher accuracy and completeness. metaFlye outperformed HiCanu and NextDenovo in real metagenomic data and still showed a significant disparity in terms of completeness and contiguity compared with hifiasm-meta. Overall, none of the assembled genomes achieved a satisfactory NG50 on synthetic data sets.

In the past 20 yr, breakthrough algorithmic advances in genome assembly have enabled scientists to assemble larger and more complex eukaryotic genomes. Despite these advances, our experiments show that there is still space for improving genome assemblers. For instance, our experimental results show that none of the assemblers were able to generate chromosome-level (or telomere-to-telomere) assemblies solely from HiFi reads. More efforts are needed to improve the assemblies' contiguity, as well as the assembly quality in low-coverage and repetitive regions (e.g., centromere and ribosomal DNA).

The problem of assembling genomes from metagenomic samples is even more difficult. Our experiments clearly highlighted the challenges of assembling many genomes simultaneously. For instance, hifiasm-meta often generated redundant assemblies, whereas metaFlye produced assemblies with low completeness. No metagenome assembler performed well on microbial metagenomic data sets with low coverage, leading to low completeness and uneven abundances. New methods are needed for improving the assembly quality of low-coverage genomes.

Finally, we believe that assembly evaluation methods are currently insufficient. BUSCO is a great tool, but it falls short when evaluating the completeness of eukaryotic assemblies when those species contain few conserved genes. QUAST is very useful but cannot be used to evaluate the assembly quality for repetitive regions owing to unreliable sequence alignments. For metagenomic assembly evaluations, CheckM can provide misleading outcomes. For example, it can give a good score to an assembly that contains a mix of sequences from closely related genomes. We believe that new evaluation methods are needed for more accurately evaluating genome assemblies.

In this study, we performed a comprehensive benchmarking of 11 assemblers on eukaryotic genomes and metagenomes. On eukaryotic genomes we measured contiguity, completeness, and accuracy across varying ploidy, coverage levels, and heterozygosities. On

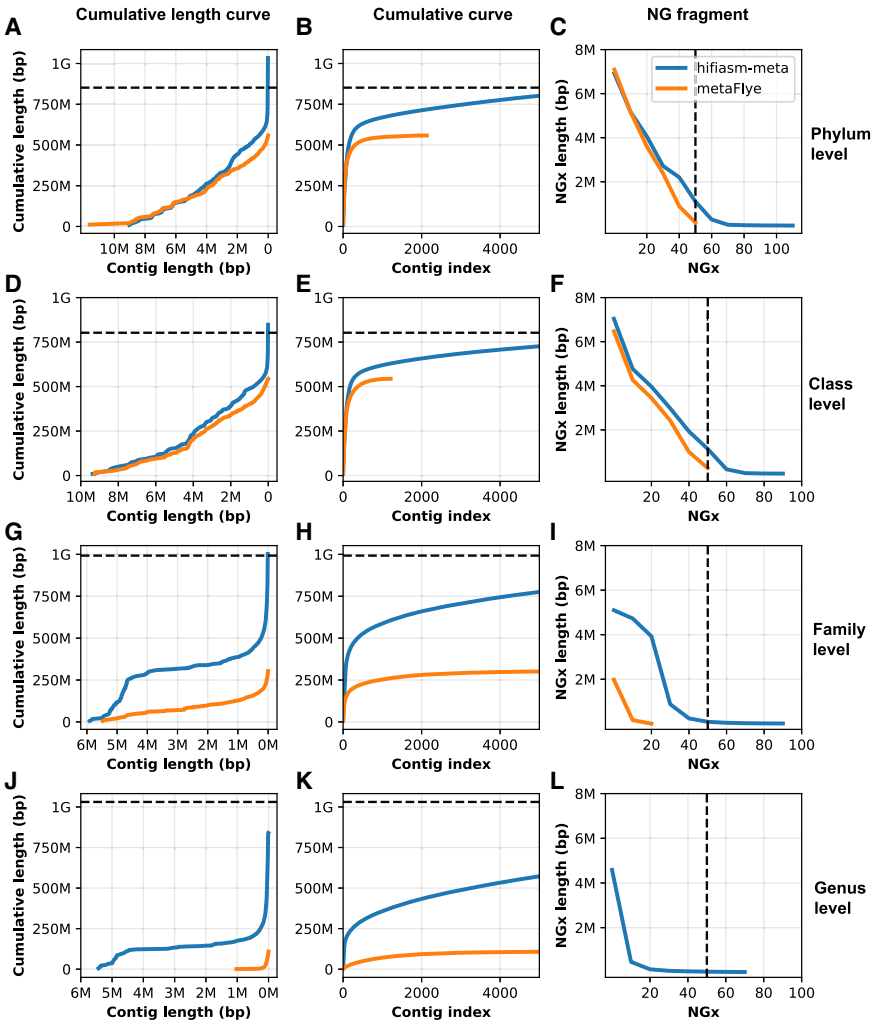

Figure 9. Comparison of hifiasm-meta and metaFlye's contiguity at different taxonomic levels. (*A*–*C*) Phylum level, (*D*–*F*) class level, (*G*–*I*) family level, and (*J*–*L*) genus level. (*A*,*D*,*G*,*J*) The *x*-axis is the contig length (sorted in decreasing order), and the *y*-axis is the cumulative contig lengths. (*B*,*E*,*H*,*K*) The *x*-axis represents the contig index (sorted in decreasing order), and the *y*-axis is the cumulative contig length. (*C*,*F*,*I*,*L*) Values for NGx; the dashed vertical line indicates the NG50.

metagenomes, we measured contiguity, completeness, and accuracy across different composition profiles and different taxonomic levels. Our experiments clearly show that hifiasm and hifiasm-meta should be the first choice for assembling eukaryotic genomes and metagenomes with HiFi data.

Methods

Data sets

Real data sets with varying ploidy

We tested the performances of the assemblers on the real HiFi reads from three eukaryotic genomes: (1) the homozygous diploid rice (*Oryza sativa*) genome ZS97 (Song et al. 2021), (2) the heterozygous diploid potato (*Solanum tuberosum*) genome (NCBI BioProject [https://www.ncbi.nlm.nih.gov/bioproject/] accession numbers PRJNA686812 and PRJNA573826) (Zhou et al. 2020), and (3) the autotetraploid wax apple (*Syzygium samarangense*) genome. The HiFi reads for rice and potato were downloaded from NCBI, whereas those for the wax apple were generated as part of this study (see Data access). The "Tub" variety of wax apple was selected for sequencing: Young leaves were collected from an individual tree planted in the field of Fujian Academy of Agricultural Sciences (Fujian Province, China) under the voucher number GPLWFJGSS0058. Genomic DNA was isolated using the Qiagen plant genomic DNA kit according to the standard PacBio operating procedure. Then, the genomic DNA was sheared by g-TUBE (Covaris), resulting in 6- to 20-kbp fragments, and sequenced using a PacBio sequel II instrument using the HiFi protocol, generating 39× coverage. Two of the authors of this paper have previously generated PacBio CLR, ONT, and Hi-C data for the wax apple for another sequencing project (Wei et al. 2023). With the addition of the HiFi reads, the wax apple data set is now the most comprehensive data set for polyploid genomes. The statistics of these data sets and their accession numbers are shown in Supplemental Table 10.

Synthetic data sets with varying ploidy

To perform a comprehensive evaluation of the performance of the HiFi assemblers, a gap-free assembly of the homozygous diploid rice genome ZS97 was used to produce (1) a synthetic heterozygous diploid genome and (2) a synthetic autotetraploid genome. First, we generated two sets of synthetic chromosomes (four in the case of the autotetraploid) by adding SNPs and structural variations (insertions and deletions) to the ZS97 genomes (Song et al. 2021) using a custom script (https://github.com/sc-zhang/ALLHiC_Evaluate_Data_Generators/blob/main/sim_snp_indel.py). To generate realistic synthetic genomes, we tuned the parameters of sim_snp_indel.py until the reads generated by wigsim (https://github.com/lh3/wgsim) from the synthetic genomes produced a GenomeScope2 *k*-mer distribution that matched the *k*-mer distribution of the heterozygous diploid big berry manzanita (Huang et al. 2022) and the autotetraploid potato (Bao et al. 2022), respectively (Supplemental Fig. 1; Bao et al. 2022; Huang et al. 2022). Following this procedure, the synthetic heterozygous diploid genome was generated using a heterozygosity rate of 2.6%, whereas the autotetraploid genome was generated using a heterozygosity rate of 1.0%. The proportion of SNPs, insertions, and deletions were 1:1:1. PBsim v1.0.4 (Ono et al. 2013) was used to generate synthetic HiFi reads for (1) the ZS97 genome (homozygous diploid), (2) the synthetic heterozygous diploid genome, and (3) the synthetic autotetraploid genome. Detailed statistics for these data sets are shown in Supplemental Table 11.

Synthetic diploid human data set

A synthetic diploid human genome was produced by merging the data from two real haploid human genomes as follows. First, HiFi reads

from CHM1 (Vollger et al. 2023) and CHM13 (Jarvis et al. 2022) were sampled at the same sequencing depth (32×). Second, the two sets of HiFi reads were merged to obtain a synthetic diploid human data set. The telomere-to-telomere level assembly of CHM13 and the contigs (>1 Mbp) of CHM1 were used as the "ground-truth" genome.

Synthetic data sets with varying heterozygosity rates, coverage levels, and sequencing error rates

To generate synthetic data sets with varying heterozygosity rates, we generate four diploid genomes with the heterozygosity rates 0.5%, 1.0%, 1.5%, and 2.5%. The proportion of SNPs, insertions, and deletions was 2:1:1 (parameter is recommended by Zhang et al. 2019). HiFi reads were simulated by PBsim v1.0.4. Detailed statistics for these data sets are shown in Supplemental Table 12.

To generate the synthetic data sets with varying read coverages, we used PBsim v1.0.4 on the synthetic heterozygous diploid genome with different coverages by changing the parameter "--depth". Detailed statistics for these data sets are shown in Supplemental Table 13.

To generate the synthetic data sets with varying sequencing error rates, we used PBsim v1.0.4 to simulate HiFi reads for the ZS97 genome and the two synthetic genomes in with coverage at 20× and the sequencing error models trained on the real HiFi data sets shown in Supplemental Table 14. Detailed statistics of these synthetic data sets are shown in Supplemental Table 15.

Real and synthetic metagenome data sets

To evaluate the performance of metagenome assemblers, we used a real sheep gut data set (NCBI BioProject accession number PRJNA595610) (Bickhart et al. 2022) and created five synthetic data sets. Detailed statistics of the sheep gut data set are shown in Supplemental Table 16.

To build the first synthetic data sets, we selected 382 bacterial genomes from the GTDB database covering 90 genera of Bacteroidetes, Actinobacteria, Firmicutes, Proteobacteria, and Fusobacteria phyla and merged them into a metagenome by assigning different abundances to them. The abundance profile was obtained by sampling coverage values from a real chicken gut metagenome assembly (Zhang et al. 2022b). PBsim v1.0.4 was used to simulate the HiFi reads with the error model learned on the real sheep gut data set. Detailed statistics for this synthetic data set are shown in Supplemental Table 17.

The other four synthetic data sets were built for different taxonomic levels such as phylum, class, family, and genus. For example, to build the phylum data set, 200 genomes from the same phylum were selected, and abundances and the simulated HiFi reads were generated in the same way as the first synthetic data sets. The synthetic data sets of class, family, and genus were also generated in the same way as phylum. Detailed statistics for these synthetic data sets and references information are shown in Supplemental Table 18 (Mash distance of references), Supplemental Table 19, and Supplemental_Metagenome_Reference.xlsx (references list).

Assembler performance evaluation

Our comprehensive assessment of the assemblers was performed by evaluating the completeness, contiguity, and accuracy of the assembled genomes for real and synthetic data sets.

Contiguity

QUAST v5.0.2 was used to assess the contiguity of the assemblies. We recorded N50, L50, and the length of the longest contig. We recall that the N50 is defined as the length for which the set of all contigs of that length or longer covers at least half of the assembled genome. L50 is defined as the count of the smallest number of contigs whose total length makes up half of the assembled genome. NG50 is defined as the length for which the set of all contigs of that length or longer covers at least 50% of the length of the actual genome. LG50 is defined as the count of the smallest number of contigs whose total length makes up half of the actual genome.

Completeness

BUSCO was used to assess genome completeness on all assemblies produced on real data. BUSCO measures the fraction of highly conserved genes that are present in the assembly (full length or fragmented).

On simulated data, three additional criteria were used to measure completeness, namely, *completeness rate* (CR), *single-copy completeness rate* (SCR), and *duplicated completeness rate* (DCR). CR, SCR, and DCR are based on *k*-mer analysis and are explained next with the help of Figure 10. In all our experiments, we used $k = 21$, which is the value typically used for eukaryotic genomes (Rhie et al. 2020). We call the set of *k*-mers that are unique in the reference genome $S_{RefG\_unikmers}$ and the set of all *k*-mers in the genome assembly S_{CtgG_kmers} (see Fig. 10). Only nonoverlapping unique *k*-mers on reference were used in the calculation of these completeness criteria and other metrics. If the assembly is complete, then all these

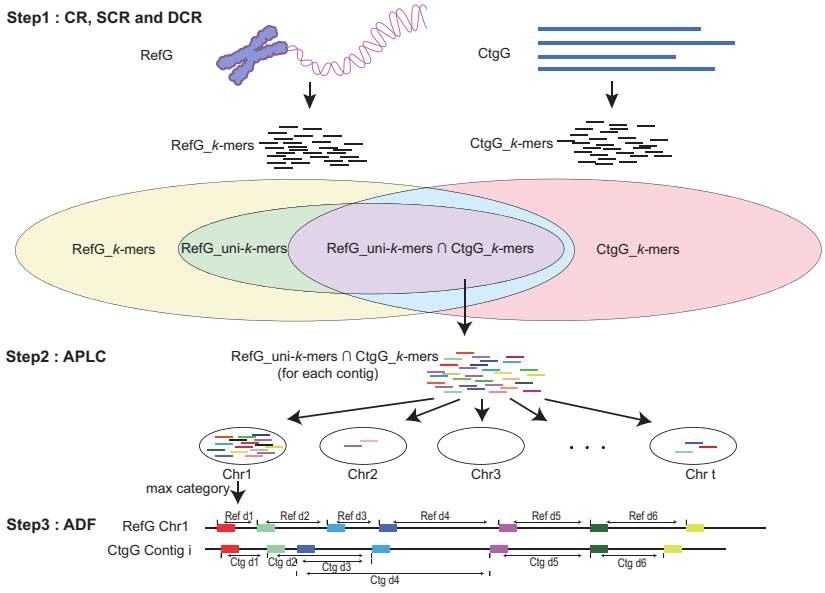

Figure 10. Completeness and accuracy of assembled genomes on synthetic data are evaluated based on a *k*-mer analysis.

unique *k*-mers in the reference are expected to appear in the assembly. In general, however, $S_{RefG\_unikmers}$ is a subset of S_{CtgG_kmers}. Thus, we define the *completeness rate* (CR) as follows:

$$CR = \frac{|S_{CtgG\_kmers} \cap S_{RefG\_unikmers}|}{|S_{RefG\_unikmers}|}.$$

CR ranges between zero and one, where one indicates a complete assembly.

Observe that the set $S_{CtgG\_kmers} \cap S_{RefG\_unikmers}$ contains single-copy *k*-mers and duplicated *k*-mers in the assembly. Thus, we can define the *single-copy completeness rate* (SCR) as follows:

$$SCR = \frac{|S_{CtgG\_SC\_kmers} \cap S_{RefG\_unikmers}|}{|S_{RefG\_unikmers}|},$$

where $S_{CtgG\_SC\_kmers}$ is the set of single copy *k*-mers in the assembly. Similarly, the *duplicated completeness* rate (DCR) is defined as

$$DCR = \frac{|S_{CtgG\_DC\_kmers} \cap S_{RefG\_unikmers}|}{|S_{RefG\_unikmers}|},$$

where $S_{CtgG\_DC\_kmers}$ is the set of duplicated *k*-mers in the assembly. Obviously, $CR = SCR + DCR$.

Observe that the *completeness rate* (CR) is similar to one of the quality measure in Quast-lg. Both criteria evaluate the completeness using the proportion of unique *k*-mers in the reference that also appear in the assembly. However, in our method, we use one *k*-mer as the representative in a set of overlapping unique *k*-mers, whereas Quast-lg uses all of them. Our method avoids overestimating or underestimating the effect of assembly errors if they are related to a number of overlapped unique *k*-mers far from the average.

Accuracy

To evaluate the assembly accuracy, we use the consensus quality value (QV), the N-rate, the APLC, and the average distance difference (ADF).

The *consensus quality value* (QV) was defined by Rhie et al. (2020), and it measures the frequency of consensus errors in the assembly. QV is defined as follows:

$$QV = -10 \log_{10}\left(1 - \left(\frac{K_{shared}}{K_{total}}\right)^{\frac{1}{k}}\right),$$

where the K_{total} is the total number of *k*-mers found in an assembly, and K_{shared} is the number of shared *k*-mers between the assembly and the reads. Observe that $P = \left(\frac{K_{shared}}{K_{total}}\right)^{\frac{1}{k}}$ represents the probability that a base in the assembly is correct, and $-10 \log_{10}(1-P)$ can be interpreted as a Phred quality score (Ewing et al. 1998). For instance, QV > 60 indicates an excellent assembly at the base level (e.g., QV = 60 translates to an accuracy of 99.9999%).

The *N-rate* is the proportion of ambiguous bases (N's) in the assembly, as reported by QUAST. The lower is the N-rate, the better the assembly.

Now observe that a contig misassembly can be detected if that contig contains unique *k*-mers from two or more chromosomes. We proposed a measure called the *average proportion of the largest category* (APLC), which captures the misassembly rate for each assembled contigs. First, we define $PChr_{a,c}$, which is the proportion of unique *k*-mers for Chromosome *a* that appear in contig *c*, as follows:

$$PChr_{a,c} = \frac{|(S_{Ctg\_i\_kmers} \cap S_{RefG\_unikmers}) \cap S_{RefG\_chr\_a\_unikmer}|}{|S_{Ctg\_c\_kmers} \cap S_{RefG\_unikmers}|},$$

where *a in* $[1, t]$, *t* is the number of chromosomes, $S_{RefG\_chr\_a\_unikmer}$ is the set of unique *k*-mers in Chromosome *a*, and $S_{Ctg\_c\_kmers}$ is the set of *k*-mers in contig *c*. Next, we define $PChr_c$ as the largest value of $PChr_{a,c}$ over all the t chromosomes for contig *c*, as follows:

$$PChr_c = max_{a \in [1,t]} PChr_{a,c}.$$

Finally, we can define the *average proportion of the largest category* (APLC) as the average value PLC_c over all the contigs, defined as follows:

$$APLC = \sum_{i=1}^{n} PLC_c / n,$$

where *n* is the number of the contigs in the assembly.

The final measure of accuracy is based on the distance between pairs of unique adjacent *k*-mer, which is expected to be the same in the reference genome and the assembled contigs. First, unique *k*-mers are sorted by their position in the reference genome, and then the distance of those *k*-mers in the assembly is calculated. The difference DF_c for a contig *c*, is defined as follows:

$$DF_c = \sum_{i=1}^{m-1} |(pos_{Ref\_kmer_{i+1}} - pos_{Ref\_kmer_i}) - (pos_{Ctg_c\_kmer_{i+1}} - pos_{Ctg_c\_kmer_i})|,$$

where *m* is the number of unique *k*-mer, $pos_{Ref\_kmer_i}$ is the position of the *i*th *k*-mer in the reference genome, and $pos_{Ctg_c\_kmer_i}$ is the position of the *i*th *k*-mer in assembled contig *c*. Finally, we can define the *average distance difference* (ADF) as follows:

$$ADF = \sum_{j=1}^{n} DF_j / n,$$

where *n* is the number of the contigs in the assembly. The smaller the value of ADF, the more accurate the assembly is.

In **Step 1**, the CR, SCR, and DCR are computed from the shared *k*-mers that are unique in the reference and the *k*-mers in the assembled contigs (represented in purple). In **Step 2**, the APLC is computed from the *k*-mers in common between the unique *k*-mers in the reference genome and *k*-mers in specific contigs. In **Step 3**, the ADF is computed from pairs of adjacent unique *k*-mers.

Runtime and memory usage

Time and memory usage were recorded for all experiments in this study. All assemblers were run on an Inspur Cluster Engine Linux cluster at Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences. The cluster has six main nodes, each of which has 80 CPUs and 3 TB of memory. The memory usage of every assembler was looking up "maximum resident set size (kbytes)" using the command "/usr/bin/time –v."

Data access

The wax apple sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA928838. All generated simulated sequencing data sets in this study are available in http://ftp.agis.org.cn:8888/~panweihua/benchmark/, and these data sets are listed in Supplemental Table 20. The scripts that implement the five quality criteria and evaluate the genome

assemblies in this study can be obtained from GitHub (https://github.com/rookieluohh/benchmark) and from the Supplemental Material (Supplemental Custom Scripts and Supplemental Five Criteria, respectively). The versions and running commands of the assemblers are listed in Supplemental Table 21.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Yongyao Li from the Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, for his technical support in data storage and maintenance. We also thank the funding support for this project. This work was supported by the National Natural Science Foundation of China (grant no. 32100501); the Shenzhen Science and Technology Program (grant no. RCBS20210609103819020); the Innovation Program of Chinese Academy of Agricultural Sciences; the Science and Technology Innovation Team of Fujian Academy of Agricultural Sciences (CXTD2021008-2); and the U.S. National Science Foundation ("Improving de novo Genome Assembly using Optical Maps"; NSF 1814359).

Author contributions: W.Y. designed the experiments and wrote the initial draft of the manuscript. H.L. proposed the five criteria and performed some genome assemblies. J.Y. designed the experiments for the metagenomic study and performed the metagenomic assembly. S.Z. prepared the simulated data sets. H.J. supported H.L. for eukaryote genome assembly. X.Z. performed genome assembly and the outcome assessment. X.H. and D.S. performed metagenomic assembly evaluation. L.L. prepared the samples for the wax apple and conducted data quality control. X.W. provided the wax apple materials. S.L. edited the manuscript and supervised the project. W.P. supervised the project and edited the manuscript.

References

Alhakami H, Mirebrahim H, Lonardi S. 2017. A comparative evaluation of genome assembly reconciliation tools. *Genome Biol* **18:** 93. doi:10.1186/s13059-017-1213-3

Bankevich A, Bzikadze AV, Kolmogorov M, Antipov D, Pevzner PA. 2022. Multiplex de Bruijn graphs enable genome assembly from long, high-fidelity reads. *Nat Biotechnol* **40:** 1075–1081. doi:10.1038/s41587-022-01220-6

Bao Z, Li C, Li G, Wang P, Peng Z, Cheng L, Li H, Zhang Z, Li Y, Huang W, et al. 2022. Genome architecture and tetrasomic inheritance of autotetraploid potato. *Mol Plant* **15:** 1211–1226. doi:10.1016/j.molp.2022.06.009

Ben-Bassat I, Chor B. 2014. String graph construction using incremental hashing. *Bioinformatics* **30:** 3515–3523. doi:10.1093/bioinformatics/btu578

Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM. 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol* **33:** 623–630. doi:10.1038/nbt.3238

Bickhart DM, Kolmogorov M, Tseng E, Portik DM, Korobeynikov A, Tolstoganov I, Uritskiy G, Liachko I, Sullivan ST, Shin SB, et al. 2022. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat Biotechnol* **40:** 711–719. doi:10.1038/s41587-021-01130-z

Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36:** 1925–1927. doi:10.1093/bioinformatics/btz848

Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38:** 5315–5316. doi:10.1093/bioinformatics/btac672

Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, Wang Y-X, Xing J-F, Huang Z-J, Wang D-P, et al. 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun* **12:** 60. doi:10.1038/s41467-020-20236-7

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18:** 170–175. doi:10.1038/s41592-020-01056-5

Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13:** 1050–1054. doi:10.1038/nmeth.4035

Di Genova A, Buena-Atienza E, Ossowski S, Sagot M-F. 2021. Efficient hybrid de novo assembly of human genomes with WENGAN. *Nat Biotechnol* **39:** 422–430. doi:10.1038/s41587-020-00747-w

Du H, Liang C. 2019. Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads. *Nat Commun* **10:** 5360. doi:10.1038/s41467-019-13355-3

Ekim B, Berger B, Chikhi R. 2021. Minimizer-space de Bruijn graphs: whole-genome assembly of long reads in minutes on a personal computer. *Cell Syst* **12:** 958–968.e956. doi:10.1016/j.cels.2021.08.009

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. *Genome Res* **8:** 175–185. doi:10.1101/gr.8.3.175

Feng X, Cheng H, Portik D, Li H. 2022. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nat Methods* **19:** 671–674. doi:10.1038/s41592-022-01478-3

Gavrielatos M, Kyriakidis K, Spandidos DA, Michalopoulos I. 2021. Benchmarking of next and third generation sequencing technologies and their associated algorithms for de *novo* genome assembly. *Mol Med Rep* **23:** 251. doi:10.3892/mmr.2021.11890

Huang Y, Escalona M, Morrison G, Marimuthu MPA, Nguyen O, Toffelmier E, Shaffer HB, Litt A. 2022. Reference genome assembly of the big berry manzanita (*Arctostaphylos glauca*). *J Hered* **113:** 188–196. doi:10.1093/jhered/esab071

Jain R, Habermann BH, Mignot T. 2021. Complete genome assembly of *Myxococcus xanthus* strain DZ2 using long high-fidelity (HiFi) reads generated with PacBio technology. *Microbiol Resour Announc* **10:** e0053021. doi:10.1128/mra.00530-21

Jarvis ED, Formenti G, Rhie A, Guarracino A, Yang C, Wood J, Tracey A, Thibaud-Nissen F, Vollger MR, Porubsky D, et al. 2022. Semi-automated assembly of high-quality diploid human reference genomes. *Nature* **611:** 519–531. doi:10.1038/s41586-022-05325-5

Kamath GM, Shomorony I, Xia F, Courtade TA, Tse DN. 2017. HINGE: long-read assembly achieves optimal repeat resolution. *Genome Res* **27:** 747–756. doi:10.1101/gr.216465.116

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37:** 540–546. doi:10.1038/s41587-019-0072-8

Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, et al. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* **17:** 1103–1110. doi:10.1038/s41592-020-00971-x

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, et al. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* **30:** 693–700. doi:10.1038/nbt.2280

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* **27:** 722–736. doi:10.1101/gr.215087.116

Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32:** 2103–2110. doi:10.1093/bioinformatics/btw152

Li Z, Chen Y, Mu D, Yuan J, Shi Y, Zhang H, Gan J, Li N, Hu X, Liu B, et al. 2012. Comparison of the two major classes of assembly algorithms: overlap–layout–consensus and de-bruijn-graph. *Brief Funct Genomics* **11:** 25–37. doi:10.1093/bfgp/elr035

Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. 2016. Assembly of long error-prone reads using de Bruijn graphs. *Proc Natl Acad Sci* **113:** E8396–E8405. doi:10.1073/pnas.1604560113

Luo X, Kang X, Schönhuth A. 2021. phasebook: haplotype-aware de novo assembly of diploid genomes from long reads. *Genome Biol* **22:** 299. doi:10.1186/s13059-021-02512-x

Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* **32:** 1088–1090. doi:10.1093/bioinformatics/btv697

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34:** i142–i150. doi:10.1093/bioinformatics/bty266

Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* **95:** 315–327. doi:10.1016/j.ygeno.2010.03.001

Mills C, Muruganujan A, Ebert D, Marconett CN, Lewinger JP, Thomas PD, Mi H. 2020. PEREGRINE: a genome-wide prediction of enhancer to gene relationships supported by experimental evidence. *PLoS One* **15:** e0243791. doi:10.1371/journal.pone.0243791

Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AWC, Pippel M, Winkler S, Hastie AR, Young G, Roscito JG, et al. 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554:** 50–55. doi:10.1038/nature25458

Nurk S, Walenz BP, Rhie A, Vollger MR, Logsdon GA, Grothe R, Miga KH, Eichler EE, Phillippy AM, Koren S. 2020. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res* **30:** 1291–1305. doi:10.1101/gr.263566.120

Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11:** 2864–2868. doi:10.1038/ismej.2017.126

Ono Y, Asai K, Hamada M. 2013. PBSIM: PacBio reads simulator—toward accurate genome assembly. *Bioinformatics* **29:** 119–121. doi:10.1093/bioinformatics/bts649

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25:** 1043–1055. doi:10.1101/gr.186072.114

Rautiainen M, Nurk S, Walenz BP, Logsdon GA, Porubsky D, Rhie A, Eichler EE, Phillippy AM, Koren S. 2023. Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nat Biotechnol* **41:** 1474–1482. doi:10.1038/s41587-023-01662-6

Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21:** 245. doi:10.1186/s13059-020-02134-9

Ríos-Touma B, Holzenthal RW, Rázuri-Gonzales E, Heckenhauer J, Pauls SU, Storer CG, Frandsen PB. 2022. De novo genome assembly and annotation of an Andean caddisfly, *Atopsyche davidsoni* Sykora, 1991, a model for genome research of high-elevation adaptations. *Genome Biol Evol* **14:** evab286. doi:10.1093/gbe/evab286

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4:** e2584. doi:10.7717/peerj.2584

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17:** 155–158. doi:10.1038/s41592-019-0669-3

Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K, Maurer N, Koren S, et al. 2020. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol* **38:** 1044–1053. doi:10.1038/s41587-020-0503-6

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31:** 3210–3212. doi:10.1093/bioinformatics/btv351

Sohn J, Nam J-W. 2018. The present and future of de novo whole-genome assembly. *Brief Bioinformatics* **19:** 23–40. doi:10.1093/bib/bbw096

Song J-M, Xie W-Z, Wang S, Guo Y-X, Koo D-H, Kudrna D, Gong C, Huang Y, Feng J-W, Zhang W, et al. 2021. Two gap-free reference genomes and a global view of the centromere architecture in rice. *Mol Plant* **14:** 1757–1767. doi:10.1016/j.molp.2021.06.018

Sun H, Jiao W-B, Krause K, Campoy JA, Goel M, Folz-Donahue K, Kukat C, Huettel B, Schneeberger K. 2022a. Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nat Genet* **54:** 342–348. doi:10.1038/s41588-022-01015-0

Sun Y, Shang L, Zhu Q-H, Fan L, Guo L. 2022b. Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci* **27:** 391–401. doi:10.1016/j.tplants.2021.10.006

Vollger MR, Logsdon GA, Audano PA, Sulovari A, Porubsky D, Peluso P, Wenger AM, Concepcion GT, Kronenberg ZN, Munson KM, et al. 2020. Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads. *Ann Hum Genet* **84:** 125–140. doi:10.1111/ahg.12364

Vollger MR, Dishuck PC, Harvey WT, DeWitt WS, Guitart X, Goldberg ME, Rozanski AN, Lucas J, Asri M, Abel HJ, et al. 2023. Increased mutation and gene conversion within human segmental duplications. *Nature* **617:** 325–334. doi:10.1038/s41586-023-05895-y

Wang B, Yang X, Jia Y, Xu Y, Jia P, Dang N, Wang S, Xu T, Zhao X, Gao S, et al. 2022. High-quality *Arabidopsis thaliana* genome assembly with nanopore and HiFi long reads. *Genomics Proteomics Bioinformatics* **20:** 4–13. doi:10.1016/j.gpb.2021.08.003

Wei X, Chen M, Zhang X, Wang Y, Li L, Xu L, Wang H, Jiang M, Wang C, Zeng L, et al. 2023. The haplotype-resolved autotetraploid genome assembly provides insights into the genomic evolution and fruit divergence in wax apple (*Syzygium samarangense* (Blume) Merr. and Perry). *Hortic Res* **10:** uhad214. doi:10.1093/hr/uhad214

Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* **37:** 1155–1162. doi:10.1038/s41587-019-0217-9

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13:** e1005595. doi:10.1371/journal.pcbi.1005595

Xiao C-L, Chen Y, Xie S-Q, Chen K-N, Wang Y, Han Y, Luo F, Xie Z. 2017. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat Methods* **14:** 1072–1074. doi:10.1038/nmeth.4432

Xue L, Gao Y, Wu M, Tian T, Fan H, Huang Y, Huang Z, Li D, Xu L. 2021. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. *Genome Biol* **22:** 203. doi:10.1186/s13059-021-02430-y

Zhang X, Zhang S, Zhao Q, Ming R, Tang H. 2019. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat Plants* **5:** 833–845. doi:10.1093/bib/bbac146

Zhang X, Liu C-G, Yang S-H, Wang X, Bai F-W, Wang Z. 2022a. Benchmarking of long-read sequencing, assemblers and polishers for yeast genome. *Nat Plants* **23:** bbac146. doi:10.1038/s41477-019-0487-8

Zhang Y, Jiang F, Yang B, Wang S, Wang H, Wang A, Xu D, Fan W. 2022b. Improved microbial genomes and gene catalog of the chicken gut from metagenomic sequencing of high-fidelity long reads. *GigaScience* **11:** giac116. doi:10.1093/gigascience/giac116

Zhou Q, Tang D, Huang W, Yang Z, Zhang Y, Hamilton JP, Visser RGF, Bachem CWB, Robin Buell C, Zhang Z, et al. 2020. Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat Genet* **52:** 1018–1023. doi:10.1038/s41588-020-0699-x

Received June 29, 2023; accepted in revised form January 23, 2024.

Comprehensive assessment of 11 de novo HiFi assemblers on complex eukaryotic genomes and metagenomes

Wenjuan Yu, Haohui Luo, Jinbao Yang, et al.

Genome Res. published online March 1, 2024
Access the most recent version at doi:10.1101/gr.278232.123