



## Article

<https://doi.org/10.1038/s41587-025-02971-8>

# Troubleshooting common errors in assemblies of long-read metagenomes

Received: 21 April 2025

Accepted: 25 November 2025

Published online: 02 January 2026

Check for updates

Florian Trigodet<sup>1,2</sup>, Rohan Sachdeva<sup>1</sup> , Jillian F. Banfield<sup>1</sup> & A. Murat Eren<sup>1</sup>

Assessing the accuracy of long-read assemblies, especially from complex environmental metagenomes that include underrepresented organisms, is challenging. Here we benchmark four state-of-the-art long-read assembly software programs, HiCanu, hifiasm-meta, metaFlye and metaMDBG, on 21 PacBio HiFi metagenomes spanning mock communities, gut microbiomes and ocean samples. By quantifying read clipping events, in which long reads are systematically split during mapping to maximize the agreement with assembled contigs, we identify where assemblies diverge from their source reads. Our analyses reveal that long-read metagenome assemblies can include >40 errors per 100 million base pairs of assembled contigs, including multi-domain chimeras, prematurely circularized sequences, haplotyping errors, excessive repeats and phantom sequences. We provide an open-source tool and a reproducible workflow for rigorous evaluation of assembly errors, charting a path toward more reliable genome recovery from long-read metagenomes.

Second-generation sequencing technologies have enabled the reconstruction of microbial genomes directly from environmental 'metagenomes' without cultivation<sup>1</sup>, a strategy that substantially enhanced our understanding of microbial diversity and function<sup>2</sup>. Yet, the relatively short sequencing reads produced by second-generation sequencing posed substantial limitations on assembly<sup>3,4</sup>, a critical computational step in genome recovery workflows in which reads are stitched together to rebuild contiguous segments of DNA (contigs) before binning, and often led to highly fragmented and sometimes highly contaminated genomes from metagenomes<sup>5–7</sup>. The emergence of third-generation sequencing technologies, such as those implemented by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), advanced genome-resolved metagenomics through ultra-long<sup>8</sup>, increasingly accurate reads that are longer than the length of common repeats in many bacterial and archaeal genomes<sup>9</sup>, providing solutions for complex

genomic puzzles with unprecedented precision<sup>10</sup>. These new opportunities rely on improved long-read assembly algorithms, an area of active research with multiple successful software packages that can assemble long reads into complete chromosomes<sup>11–14</sup>. However, assessing the accuracy of the assembled long reads from metagenomes poses challenges, especially for complex environments with substantial diversity that often include poorly studied organisms.

The benchmarking of assembly algorithms typically relies on a few key principles, including the use of mock or simulated datasets, evaluation of contig length distributions, assessment of unassembled read fractions, genes, functions and comparisons with other assemblers<sup>15,16</sup>. K-mer-based solutions<sup>17</sup> and short-read alignment statistics have been effectively used to unveil structural variants as well as misassemblies of eukaryotic genomes<sup>18–22</sup> including the human genome<sup>23–30</sup>. While the compendium of strategies offers critical tools for algorithm

<sup>1</sup>Helmholtz Institute for Functional Marine Biodiversity, Oldenburg, Germany. <sup>2</sup>Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany. <sup>3</sup>Innovative Genomics Institute, University of California Berkeley, Berkeley, CA, USA. <sup>4</sup>Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA. <sup>5</sup>Department of Environmental Science Policy and Management, University of California, Berkeley, CA, USA. <sup>6</sup>Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>7</sup>Biomedicine Discovery Institute, Monash University, Melbourne, Victoria, Australia. <sup>8</sup>Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany. <sup>9</sup>Marine Biological Laboratory, Woods Hole, MA, USA. <sup>10</sup>Max Planck Institute for Marine Microbiology, Bremen, Germany. e-mail: [jbanfield@berkeley.edu](mailto:jbanfield@berkeley.edu); [meren@hifmb.de](mailto:meren@hifmb.de)

development, the developers of metagenomic assemblers rarely take into account how individual reads align to assemblies and quantify the rate, origins and impact of mismatches between original reads and final contigs. Here we assessed the extent of agreement between individual high-fidelity PacBio reads and assembly results reported by four state-of-the-art long-read assemblers, and observed a wide range of issues, including haplotyping errors, chimerism, premature circularization and regions of contigs that are not supported by any of the input sequences. Assemblies of ONT long reads show similar errors<sup>31</sup>, yet here we limit our benchmarks to PacBio given the relatively higher base accuracy of PacBio reads (~99.95%, or 5 errors per 10 kilobase pair (kb)) compared with the ONT reads (~99%, or 100 errors per 10 kb). Overall, our survey offers reproducible means to identify long-read assembly errors and insights into their downstream implications for researchers who develop or use long-read assembly algorithms to consider.

## Results

Our study benchmarks four state-of-the-art long-read assembly software programs: (1) HiCanu v2.2 (ref. 14), (2) hifiasm-meta v0.3 (ref. 11), (3) metaFlye v2.9.5 (ref. 13) and (4) metaMDBG v1 (ref. 12), based on their assembly performance of 21 PacBio HiFi metagenomes (Supplementary Table 1). We include HiCanu in our analysis despite the fact that it was originally developed for genome assembly as it has recently been benchmarked against hifiasm-meta and metaFlye<sup>11</sup>. Of the 21 metagenomes we include in our benchmarks, 13 represent those that were also used by the authors of at least one assembly algorithm to evaluate performance, and correspond to mock communities (Zymo-HiFi D6331 and ATCC MSA-1003) and anaerobic digesters (AD2W1, ADW20 and ADW40), as well as gut samples from humans (Human O1, Human O2, Human V1 and Human V2), chickens and sheep (Chicken, SheepA and SheepB). The remaining eight are novel HiFi metagenomes from the surface ocean (sample names start with HADS)<sup>32</sup>, a key biome that was not included in previous benchmarks and represents a complex ecosystem with relatively little genomic representation (Supplementary Table 1).

Investigating the agreement between individual long reads and assemblies they support requires the alignment of reads to resulting contigs. For this task, our study uses minimap2, a high-performance mapping software designed to align long reads to reference sequences<sup>33</sup>. While short-read alignment software such as Bowtie2 (ref. 34) or BWA<sup>35</sup> is typically used to perform end-to-end alignments, minimap2 allows premature ends in read alignments and can map the remainder of the read to another reference location. This so-called read clipping procedure is a critical tool to identify locations in the final assembly that is poorly supported by individual long reads. To gain quantitative insights into assembly artifacts, we developed 'anvi-script-find-misassemblies'<sup>36</sup>, a script that comprehensively summarizes locations and frequencies of read clipping events, in which long reads are split systematically by the mapping software to maximize agreement between reads and assemblies, in addition to regions in contigs that are not supported by individual long reads.

### Assembly errors are common to all long-read assemblers

Instances in which the vast majority of long reads are clipped at the same nucleotide position are strong evidence that the final assembled sequence contains an error. However, quantifying the number of errors in a given assembly is not as straightforward as counting the clipping events as a single assembly error may result in multiple clipping events associated with neighboring nucleotide positions in the reference sequence (Fig. 1a). To minimize false positives in our results, and exclude clipping events due to within-population biological variation, our analyses below primarily focus on locations in long-read assemblies with a 100% clipping rate.

The assembly metrics differed greatly between sample types and assembly algorithms. For instance, all human gut metagenomes were

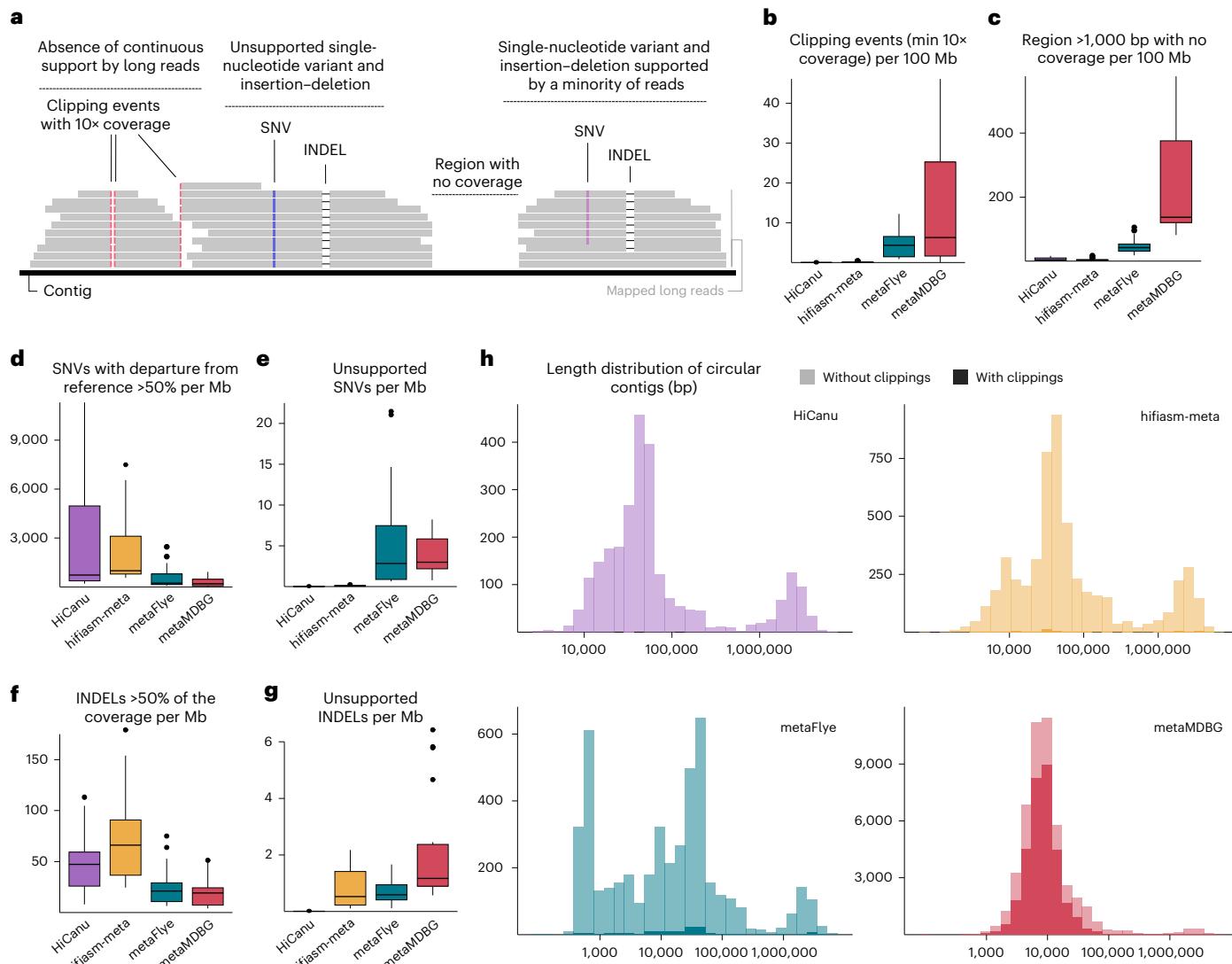
assembled to a similar total size by all assemblers, but not the surface ocean samples (Supplementary Table 2), for which metaMDBG produced 610% more assembled sequences compared with HiCanu on average. We found high-confidence read clipping events (that is, 100% clipping at locations with at least 10 $\times$  coverage) in at least one sample from each assembler (Supplementary Table 2). However, the frequency of these events normalized by the assembly size showed that metaFlye and metaMDBG generated up to 180 times more clipping events compared with HiCanu and hifiasm-meta for the same samples (Fig. 1b). The number of clipping events was particularly high in the surface ocean samples, in which assemblies from metaMDBG had over three orders of magnitude more clipping events compared with those from hifiasm-meta (Supplementary Table 2). Overall, clipping events affected up to 5.6% of contigs longer than 10,000 nucleotides reported by metaMDBG with an error rate up to 46 per 100 million base pairs (Mb) of assembly. We also computed the number of regions longer than 1,000 bp with no apparent coverage by individual long reads and found that the occurrence of regions that are not supported by any read was as pervasive as clipping events. While this issue also affected all assemblers (Fig. 1c and Supplementary Table 2), metaMDBG led the pack with up to 5.3% of all contigs longer than 10,000 nucleotides with zero-coverage regions.

The reporting of contig circularity was a common feature of all assemblers; however, the number of circular contigs in final assemblies also varied between algorithms (Fig. 1h and Supplementary Table 2). MetaMDBG generated substantially more circular contigs than other assemblers, notably from surface ocean metagenomes. Overall, we found at least one clipping event for a large proportion of circular contigs reported by metaMDBG (Fig. 1h), which, in some cases, represented up to 77% of the circular contigs in a sample (for example, the marine sample HADS 013; Supplementary Table 2).

In addition to the clipping events, we reported the frequency of single-nucleotide variants (SNVs) and insertion–deletion events (INDELS) in the assembled contigs that were not fully supported by the long reads. We classified such cases into two categories: (1) 'minority variants', where the final assembly included a nucleotide or an INDEL that did not match the most frequent variant in the supporting long reads (Fig. 1a,d,f), and (2) 'unsupported variants', a more severe case in which the final assembly included a nucleotide or an INDEL that did not occur in any of the long reads at that position (Fig. 1a,e,f). After normalizing based on the assembly size, HiCanu and hifiasm-meta had the most minority variants, while metaFlye and metaMDBG had the most unsupported variants. The latter case affected thousands of genes in all assemblies (Supplementary Table 2), leading to genes with incorrect amino acid sequences owing to the impact of INDELS on open reading frames<sup>37,38</sup>. Our detailed investigation of individual clipping events associated their occurrence with a few recurring classes of erroneous reporting of contigs, including chimeras, premature circularization, haplotyping issues, false duplications and nonexistent sequences, for which we offer an incomplete list of examples below.

### Chimeric contigs

Our inspection of contigs with a high proportion of read clipping events revealed chimeric contigs. In some cases, chimeras brought together sequences from taxa that belonged to distinct phyla (Fig. 2). Most chimeras brought together sequences from two distinct taxa, but cases that merged sequences from more than three organisms were not uncommon (Fig. 2), and they sometimes combined sequences from distinct domains of life. In one extreme case, metaMDBG formed a sequence that included regions from organisms that originated from Euryarchaeota, Pseudomonadota, Bacteroidota and Cyanobacteria (Fig. 2). We acknowledge that the list of contigs we surveyed here for chimerism is far from exhaustive; furthermore, relying on clipping events alone may occasionally miss chimeric contigs. For instance, we stumbled upon a suspiciously long contig. Even though the frequency of read clipping events did not



**Fig. 1 | Assembly errors across the assemblers.** **a**, A schematic representation of long reads mapping to a contig with multiple types of read disagreement with the reference, including INDELS and SNVs representing more than half or all the coverage, and clipping events spanning the entire coverage. All metrics for **b–g** are normalized by assembly size and exclude the two mock community metagenomes ( $n = 19$  assemblies). For all box plots, the box represents the interquartile range (IQR), the central line indicates the median and whiskers

extend to 1.5 $\times$  the IQR. **b**, Number of clipping events supported by at least 10 reads. **c**, Number of regions over 1,000 bp with no apparent coverage.

**d,e**, Number of SNVs representing >50% (**d**) or all (**e**) of the coverage at a given locus. **f,g**, Distribution of INDELS >50% of the coverage (**f**) or all of the coverage (**g**). **h**, Length distribution of circular contigs by each assembler. The darker color represents the distribution of circular contigs with at least one clipping event.

flag this 7.38-Mb contig, anvi'o assigned it a redundancy score of 100% because it contained two copies of every bacterial single-copy core gene. Our manual inspection showed that the contig metaMDBG reported from the sample Human O1 conjoined two sequence-discrete *Lachnospiraceae* populations (Supplementary Fig. 1).

The highly dangerous nature of chimeric contigs for downstream analyses is dampened by the straightforward nature of their identification by anyone who carefully investigates their data: most chimeric contigs that erroneously connect genomic regions from two or more distinct populations can be easily identified using a series of well-understood indicators, such as sudden shifts in GC content, read coverage and gene-level taxonomy, or through unexpected inventories of single-copy core genes (SCGs). Nevertheless, with the increasing tendency of researchers to generate large metagenome-derived genome compendiums<sup>39–43</sup>, such evaluations are rarely, if ever, conducted. Thus, in an ideal world, the burden of resolving chimeras should never fall on the shoulders of the end users of assembly algorithms.

### Premature circularization

To deliver one of the most sought after promises of long-read sequencing, most long-read assemblers contain built-in features to circularize contigs and report potentially complete microbial and mobile genetic element genomes. Premature circularization, the reporting of a contig as circular when it omits parts of the genome it originates from, is a form of error with dangerous downstream implications. Yet, our analyses showed that the algorithmic features that reconstructed circular contigs in long-read assemblers, especially metaMDBG, and to a lesser degree hifiasm-meta, were far from reliable.

For instance, an archaeal genome that belongs to the genus *Methanothrix* recovered from the AD Sludge sample as a circular genome by hifiasm-meta represents a clear example of the nature and implications of premature circularization. Through an analysis with additional genomes from the RefSeq collection of the US National Center for Biotechnology Information (NCBI), we confirmed that the circular genome was missing a large fraction of the core *Methanothrix*



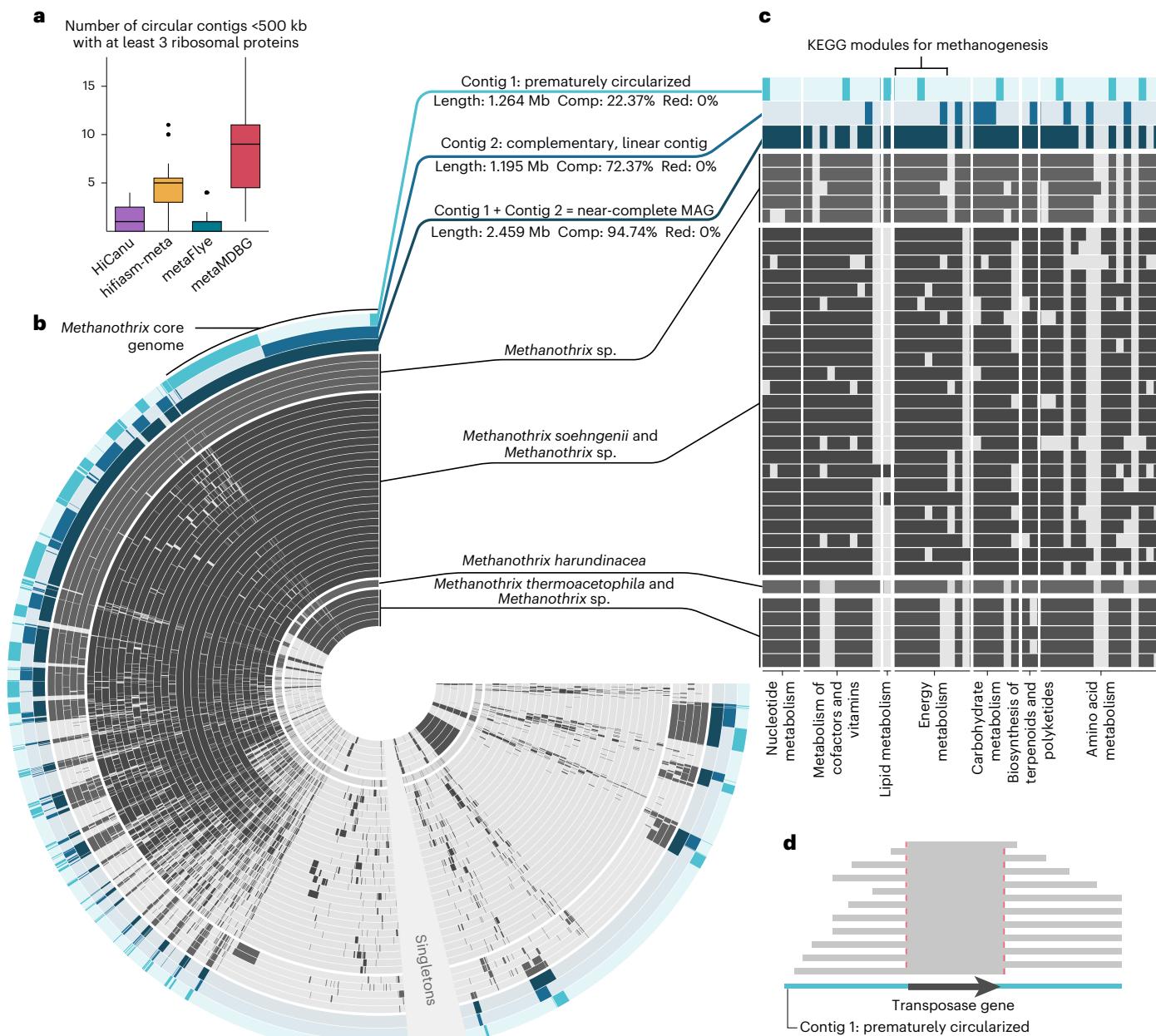
**Fig. 2 | Multi-domain and multi-phylum contigs.** Six contigs from metaMDBG, metaFlye and hifiasm-meta. For each contig, we showed the GC content, coverage in the metagenomics reads used for their assembly and gene-level taxonomy computed with Kaiju and the nr\_euk database. For each assembly breakpoint, we show a zoomed-in detail of the read mapping from IGV. In these subplots, the red

arrows at the end of the mapped read indicate clipping and the coloring at the end of these reads indicates that the following portion of the read mapped to another contig and similar colors indicate that multiple reads continue to map on the same contig. The blue markers indicate large INDELS (>150 bp).

pangenome (Fig. 3b, light blue), including key metabolic modules for methanogenesis that were common to all *Methanotherix* genomes (Fig. 3c). In this case, early circularization seemed to have occurred near a transposase (Fig. 3d and Supplementary Fig. 2). Nearly all the reads clipping on either side of the transposase had supplementary mapping to another contig in the assembly output (Fig. 3b,c, medium blue) encoding the missing metabolic module. The combination of the circular genome and the additional contig matched the other genomes in the genus-level pangenome (Fig. 3b,c, dark blue), suggesting that both contigs belonged to the same population, and neither of them were circular by themselves.

Comparing the overall accuracy of circularization across assemblers is difficult. For instance, low completion estimates based on

single-copy core genes can serve as a quick filter, but not all prematurely circularized genomes will have low completion estimates as missing genomic content will not always contain single-copy core genes. Length comparisons between circular contigs and known genomes in public databases could offer another means of scrutiny, but this strategy will not be effective against poorly studied clades, or those that have no representation in genome repositories. Nevertheless, here we conservatively assumed that circular contigs that were under 500 kb and contained at least 3 ribosomal proteins were most likely circularized erroneously. While this criterion is useful as a proxy for identifying potential assembly artifacts, we note that some naturally occurring small genomes may also meet these criteria<sup>44,45</sup>. Despite this caveat, we reasoned that the frequency with which such contigs



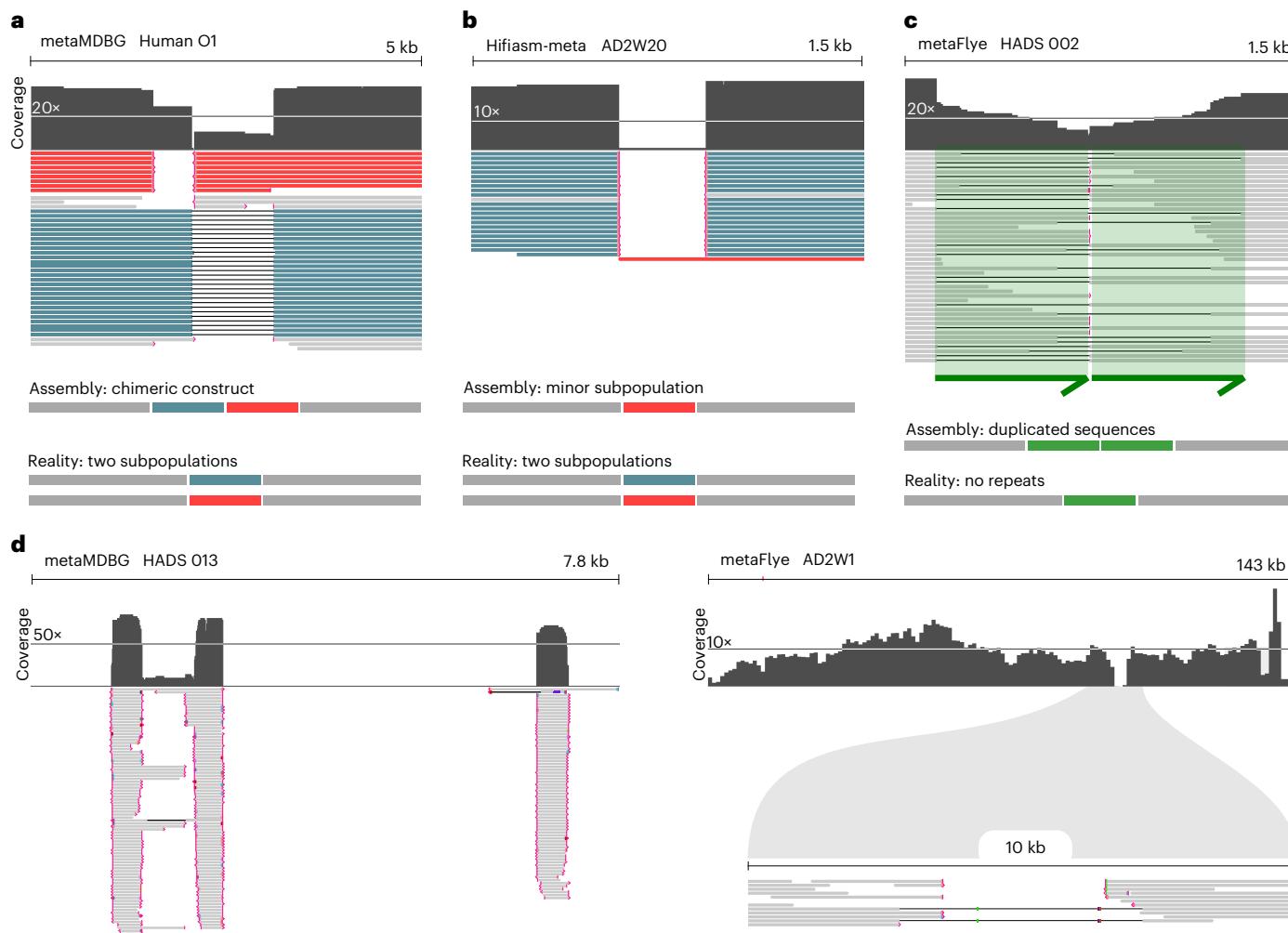
**Fig. 3 | Premature circularization of a *Methanothrix* genome.** **a**, Frequency of circular contigs under 500 kb with a minimum of 3 ribosomal proteins for each assembly, excluding the two mock communities ( $n = 19$ ). The box represents the IQR, the central line indicates the median and the whiskers extend to  $1.5 \times$  the IQR. **b**, A pangenomic analysis of all publicly available *Methanothrix* genomes from the RefSeq database of NCBI completed with the so-called circular genome of *Methanothrix* assembled from the sample AD Sludge by hifiasm-meta (light blue), as well as a contig from the same assembly that corresponds to the rest of the

missing *Methanothrix* genome (medium blue) and the combination of these two contigs (dark blue). **c**, KEGG metabolic module completion of all genomes and contigs in **b**. **d**, A schematic representation of the reads mapping over a transposase gene in the prematurely circularized contigs (light blue in **b** and **c**) showing the lack of read support around the gene; the full figure is available in Supplementary Fig. 2. MAG, metagenome-assembled genome; Comp, completion; Red, redundancy of genomes based on archaeal single-copy core genes.

appear across assemblers could provide a useful approximation of each assembler's tendency toward premature circularization. Within this framework, metaMDBG reported about twice as many circular contigs than hifiasm-meta, and about four times more circular contigs than HiCanu and metaFlye (Fig. 3a). Assuming that easily identifiable events of premature circularization errors are reasonable proxies for the rate of all premature circularization errors, this result suggests that metaMDBG and hifiasm-meta are more prone to premature circularization errors compared to other assemblers, such as HiCanu and metaFlye. Our estimates of the rate of false circularization come very close to the benchmarking results shared by the authors in the original metaMDBG

publication<sup>12</sup>, in which they report two times more circular contigs than hifiasm-meta and four times more circular contigs than metaFlye from the human gut. While the high rate of circularization is presented as a strength of metaMDBG in comparison to other assemblers<sup>12</sup>, our observations suggest that the higher rate of circularization may be a result of a higher tendency to report noncircular elements as circular (Figs. 1 and 3, Supplementary Fig. 2 and Supplementary Table 2).

Given the vast difference in quality and perception between a draft genome and a circular genome, it is essential for assembly algorithms that promise circular genomes to be conservative in their decision-making. The identification of prematurely circularized



**Fig. 4 | Prototypical mapping artifacts and their putative origin.** **a**, A chimeric sequence assembled from two subpopulations. At a conserved locus, two subpopulations existed with their own and distinct sequence. The assembled contig contains all or a part of each subpopulation-specific sequence resulting in a chimeric construct. **b**, Another example of a variable genomic site, but in this example, the contig sequence contains the sequence of a very minor subpopulation, supported by only one read. **c**, Duplicated sequence found only

in the assembly, not supported by any long reads. **d**, Two contigs assembled from metaMDBG (left) and metaFlye (right) presenting large regions with no coverage. We BLASTed these regions back to the long reads and found no hits. Coverage visualization was exported from the anvi'o interactive interface (left) or the IGV software (right) and the read mapping visualization was from IGV as well. INDELs smaller than 150 bp as well as mismatches are not shown in the mapping. Red markers at the end of reads indicate read clipping.

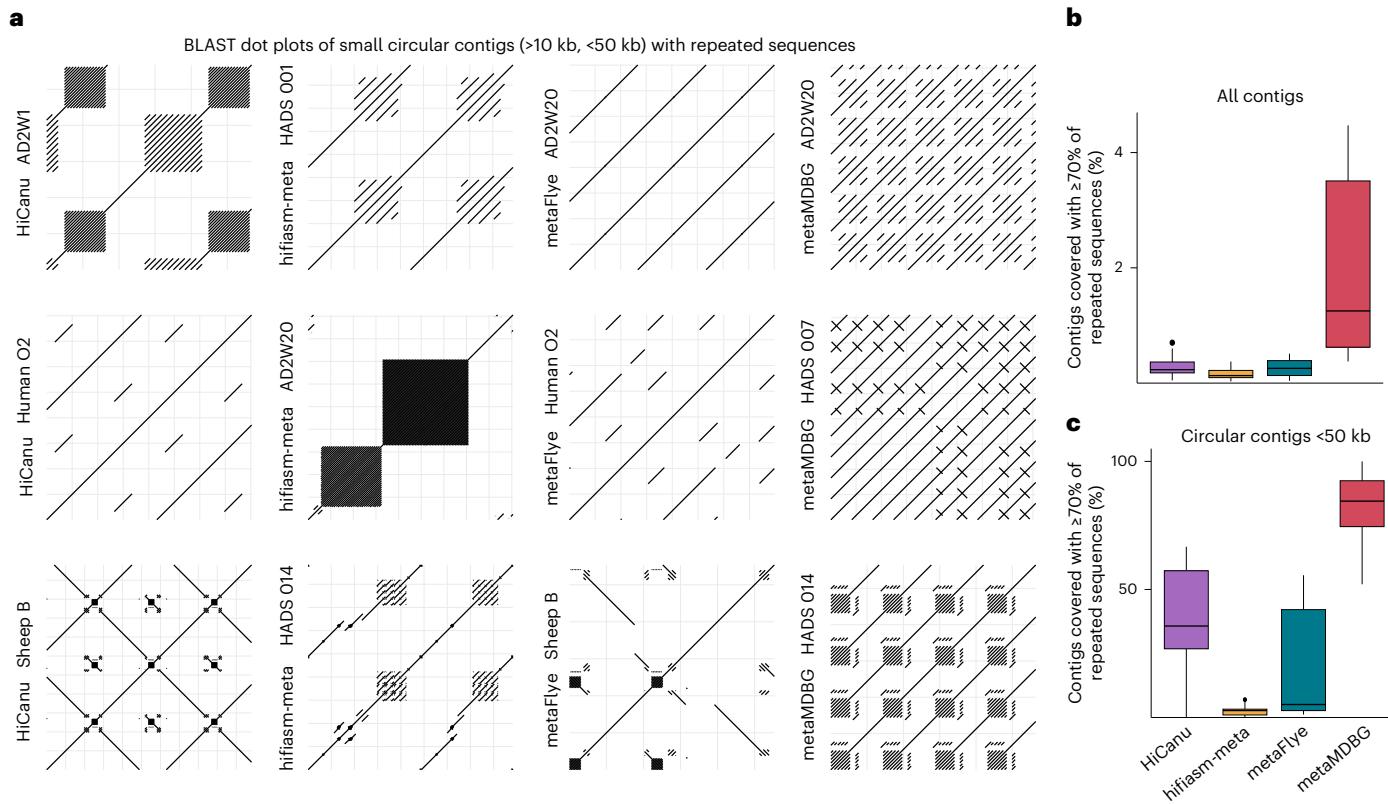
contigs in assembly results will be a notoriously difficult task for end users, especially when the missing genomic context is relatively short, or circular contigs represent plasmids or viruses. Thus, it would be ideal if the circular sequences are validated more rigorously by the assembler before the final reporting.

#### Haplotyping errors, false duplications and nonexistent sequences

Accurate reconstruction of genomic variation is essential to associate within-population structural differences to ecological or evolutionary phenotypes. However, resolving genomic regions that differ between otherwise very closely related subpopulations is a major challenge for de novo assemblers<sup>46</sup>. Assemblers may resolve such structural complexity through three approaches: (1) by reporting separate contigs for variable sequences and conserved regions that flank them, (2) by reporting the most prevalent variable region along with the flanking conserved regions in a single contig and alternative regions in shorter contigs or (3) by duplicating conserved flanking regions in multiple contigs that describe each of the variable regions and their surroundings as separate contigs in a haplotype-aware fashion. Our survey of

long-read assembly results revealed unexpected haplotyping decisions in multiple recurrent forms. In some cases, assemblers concatenated subpopulation-specific variable regions flanked by conserved loci, rather than reporting only one of them accurately (Fig. 4a). As a result, long reads that map to these regions are clipped at the end of their respective subpopulation sequence, a phenomenon that is also known as haplotypic duplication<sup>27,29</sup>. In other cases, the final contig represented a variable region found in a minor subpopulation supported by a small number of long reads or a single one (Fig. 4b), violating the logical expectation to recover a consensus sequence that represents the most abundant subpopulation.

We also identified cases in which assemblers reported false genomic duplications in assembled contigs that were not supported by any long read. Such false repeats were often manifested by high-frequency read clipping events and appeared as long direct repeats that had low likelihood to be present in the target genomic context owing to the sudden decrease in read coverage and/or massive inserts in mapped reads (Fig. 4c). Yet another anomaly was the reporting of sequences that did not exist. Searching for zero-coverage regions from contigs that are longer than 500 bp against the database



**Fig. 5 | Repeats across short contigs.** **a**, Dot plots of reciprocal BLAST results from 12 short circular contigs showing high amounts of repeated sequences. **b,c**, Proportion of contigs with at least 70% of their length duplicated at least once. Result computed for all contigs (**b**) and only for circular contigs under

50 kb (**c**) across all assemblies, excluding the two mock communities ( $n = 19$ ). For all box plots, the box represents the IQR, the central line indicates the median and whiskers extend to  $1.5 \times$  the IQR.

of raw long reads using the NCBI's Basic Local Alignment Search Tool (BLAST) with the flag '-dust no' to include sequences of low complexity (Supplementary Table 3), we confirmed that the assembly outputs by metaMDBG and metaFlye occasionally included up to over-5,000-bp-long regions that have no homology to any of the input long reads (Fig. 4d). We further confirmed this observation by comparing the  $k$ -mer ( $k = 21$ ) content between these regions and long reads, and found that over 90% of the  $k$ -mers in zero-coverage regions were absent in the long reads (Supplementary Table 3). False duplications and reporting of nonexistent sequences in assembly results are unexpected behaviors from any assembler and can lead to spurious open reading frames or the omission of genuine ones.

### Excessive repeats

A recurrent and puzzling observation throughout our manual inspections of assembly results was the astonishing number of repeats that occasionally made up the entirety of some contigs. Yet, these repeats were not caught by our survey of read clipping event frequencies to mark regions of concern as repeats rarely resulted in 100% read clipping events to pass our filter. Thus, we characterized repeats without any read mapping data but by aligning each contig to itself. We marked any region of a contig as a 'repeat' if it was longer than 200 bp and occurred multiple times in the same contig with at least 80% identity, based on observed similarity in naturally occurring repeats<sup>47</sup>. Our survey of the frequency and distribution of such repeats across all contigs showed that each assembler reported contigs with repeats. As repeats are common in nature, and the improved ability to resolve repeats is one of the strengths of long-read sequencing, this finding alone is not concerning. However, the nature of repeats in assemblies revealed by dot plots often showed unexpectedly intricate patterns, suggesting a high likelihood that they were assembly artefacts, rather than natural

genomic organizations (Fig. 5a). While naturally occurring tandem repeats, inversions or palindromes could lead to similar dot plots, our manual inspection of individual cases revealed a variety of errors that spanned from duplicated reporting of known circular plasmids to contigs with multiple repeats that are not supported by any long read.

Repeats differed in their length, identity and frequency across algorithms. The per-sample average length and sequence identity of repeats varied between 600 bp and over 1,450 bp, and 88% and 92%, respectively. MetaMDBG generated more repeats than any other assembler (Fig. 5b), up to over 300,000 repeats in a single assembly (Supplementary Table 2), exceeding the number of repeats reported by HiCanu for the same sample 235 times. The average length of repeats was relatively short, yet we found repeats that were up to 225,520 nucleotides long, and repeats occurred as much as over 990 times in a single contig (Supplementary Table 2). To summarize the number and proportion of contigs reported by each assembler with an excessive number of repeats, we conservatively searched for assembled sequences in which at least 70% of the sequence was composed of repeats. This search revealed that metaMDBG generated on average 2% of contigs with a high number of repeats, a rate that was over 14 times higher than its runner up, hifiasm-meta (Fig. 5b). When we limited our survey to circular contigs under 50 kb, which represented over 90% of all circular contigs, the proportion of repeat-rich contigs skyrocketed across all sample types for all assemblers, with the exception of hifiasm-meta (Fig. 5c and Supplementary Table 2). For instance, 87% of all circular contigs under 50 kb reported by metaMDBG were largely composed of repeats, and in some samples, such as chicken, this number reached 100%, suggesting that artifactual repeats represent at least one of the factors that lead to false circularization (Fig. 5c). Marine metagenomes were particularly difficult for all assemblers. In this sample type, metaMDBG reported the highest number of circular contigs



**Fig. 6 | Assembled *E. coli* from a mock community does not match reference genomes.** **a**, A pangenome analysis of the *E. coli* reference genome used in the Zymo mock community and the *E. coli* circular genome assembled by metaMDBG. Each layer corresponds to a genome, and the coloring represents the presence or absence of a gene cluster from each genome. The gene clusters

are organized by the synteny of genes in the metaMDBG contig. **b,c**, Genome pairwise average nucleotide identity (**b**) and alignment coverage (**c**). **d**, Details of the long-read mapping over two genomic regions highlighted in **a**. These regions included genes not found in *E. coli* B1109 (closest reference genome by ANI metrics), but rather found in other *E. coli* genomes.

(Supplementary Table 2); however, 92% of them under 50,000 bp were composed of repeats (Supplementary Table 2). While metaFlye performed as well as hifiasm-meta in most other sample types, up to 55% of circular contigs metaFlye generated from marine metagenomes were largely composed of repeats (Supplementary Table 2).

These results show that the exciting prospects of recovering complete and circular plasmid and virus genomes from long-read sequencing of complex metagenomes are still far from being realized, and that the state-of-the-art long-read assemblers often fall short when handling short circular elements.

### Mock datasets are useful but can yield misleading insights into the accuracy of algorithms in real-world applications

Popular mock datasets such as Zymo-HiFi D6331 and ATCC MSA-1003 are commonly used for benchmarking long-read assemblers. While their known composition constitutes a reasonable starting point, the mock datasets do not represent the complexity of the natural samples, as also noted by the authors of metaFlye<sup>13</sup>. Their utility to test assemblers is further reduced when benchmarks that use mock communities simply align contigs that emerge from the assemblies to reference genomes without comprehensive reporting of other assembly metrics<sup>11–13</sup>. For instance, while hifiasm-meta reports favorable outcomes given the reference genomes in mock datasets<sup>11</sup>, in our tests, the algorithm generated massive assemblies for each mock dataset, in which the final assembly was 270 Mb instead of the expected size of 93 Mb for the Zymo-HiFi D6331, and it was 948 Mb instead of the expected size of 66.44 Mb for the ATCC MSA-1003, resulting in the lowest N50 values and the highest number of clipping events across all assemblers. Given its performance with the mock datasets, one may expect hifiasm-meta to perform poorly in its applications to complex metagenomes. Yet, in marine samples, hifiasm-meta was first in N50, and second to metaMDBG in assembly size with two orders of magnitude fewer clipping errors, suggesting that the performance of an assembler with mock communities may not predict its performance with real-world datasets.

While mock communities do not represent the diversity and complexity found in real-world samples, they shine in one fundamental way: the known genomic makeup of input organisms to identify glaring issues post-assembly. The Zymo-HiFi D6331 includes five different strains of *Escherichia coli*, which yields sequencing data with complex cases for assemblers owing to the presence of highly conserved and divergent genomic regions. HiCanu and hifiasm-meta were both successful at reconstructing at least one of the five *E. coli* genomes (Supplementary Fig. 3), while metaMDBG reported a circular contig that corresponded to a chimeric genome (Fig. 6). On the basis of pairwise average nucleotide identity (ANI) comparisons, this genome appeared to be most similar to B1109, one of the *E. coli* strains in the Zymo-HiFi D6331 mock dataset (Fig. 6b and Supplementary Table 4). However, the pangenome of the five *E. coli* genomes and the metaMDBG circular contig showed that it was not only missing a portion of the B1109 genome, but also including genomic regions exclusive to other *E. coli* strains and absent in B1109 (Fig. 6a). Clipping events also captured chimeric regions and revealed -10-kb locus with no coverage (Fig. 6c,d). That region without coverage was duplicated in two other contigs, suggesting that long reads were preferentially recruited there (Supplementary Table 5). Using ANI values calculated from local alignments of assembled contigs to reference genomes can mask critical assembly errors, such as phantom sequences, chimeras and unexpectedly large assembly outputs, and invalidate perhaps the only useful aspect of mock datasets while inflating the reported accuracy of assembly algorithms.

### Discussion

Biotechnology, biomedical and basic research communities rely on high-quality assemblies, and trustworthy results are critical for the quality of public genome databases. Our findings highlight the need for

rigorous evaluation of long-read assembly algorithms beyond benchmarks that typically prioritize runtime, contig length or the number of circular contigs. We show that the quantification of read clipping events offers effective means to identify the most severe assembly errors. Assembly algorithms should use input reads for more aggressive post-assembly error correction (rather than offloading this burden onto end users who may lack the time, expertise or computational resources to perform such refinements themselves), and consider offering additional options to adjust their heuristics for researchers willing to sacrifice faster runtimes for fewer errors in their final assembly output.

While our analyses focus on errors that occur during assembly, we note that different assembly errors will differ in their likelihood to influence final genome reconstructions. Some error types, such as premature circularization of contigs, have a higher probability of propagating into high-quality genomes. By contrast, others, such as chimeric contigs, will probably be more effectively filtered out during the binning process, especially when binning tools leverage both sequence composition and differential coverage information. Understanding how specific classes of assembly errors affect genome recovery represents a critical consideration for those who rely on assemblies of long-read metagenomes. Since the initial dissemination of our study as a preprint, an updated version of metaMDBG (v1.2) and a new long-read assembler, myloasm<sup>48</sup>, became available. Even though these assemblers are not error free, their consideration of diverse errors their predecessors have suffered, and their adoption of higher scrutiny that includes the characterization of read clipping events to detect inconsistencies have led to substantially lower numbers of errors in their output (Supplementary Table 2). This paints a positive picture of the future of assembly algorithms and genome-resolved genomics in the era of long-read sequencing technologies, and shows both the practical value of error-aware strategies and the continued need for systematic error-detection frameworks to guide algorithm development, use and downstream interpretations when long-read assemblers meet metagenomes.

### Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-025-02971-8>.

### References

1. Tyson, G. W. et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
2. Eren, A. M. & Banfield, J. F. Modern microbiology: embracing complexity through integration across scales. *Cell* **187**, 5151–5170 (2024).
3. Tørresen, O. K. et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res.* **47**, 10994–11006 (2019).
4. Olson, N. D. et al. Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Brief. Bioinform.* **20**, 1140–1150 (2019).
5. Shaiber, A. & Eren, A. M. Composite metagenome-assembled genomes reduce the quality of public genome repositories. *mBio* **10**, e00725-19 (2019).
6. Chang, T., Gavelis, G. S., Brown, J. M. & Stepanauskas, R. Genomic representativeness and chimerism in large collections of SAGs and MAGs of marine prokaryoplankton. *Microbiome* **12**, 126 (2024).
7. Chen, L. X., Anantharaman, K., Shaiber, A., Murat Eren, A. & Banfield, J. F. Accurate and complete genomes from metagenomes. *Genome Res.* **30**, 315–333 (2020).

8. Jain, M. et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338–345 (2018).
9. Koren, S. et al. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol.* **14**, R101 (2013).
10. Wenger, A. M. et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162 (2019).
11. Feng, X., Cheng, H., Portik, D. & Li, H. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nat. Methods* **19**, 671–674 (2022).
12. Benoit, G. et al. High-quality metagenome assembly from long accurate reads with metaMDBG. *Nat. Biotechnol.* **42**, 1378–1383 (2024).
13. Kolmogorov, M. et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* **17**, 1103–1110 (2020).
14. Nurk, S. et al. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, 1291–1305 (2020).
15. Meyer, F. et al. Critical Assessment of Metagenome Interpretation: the second round of challenges. *Nat. Methods* **19**, 429–440 (2022).
16. Delgado, L. F. & Andersson, A. F. Evaluating metagenomic assembly approaches for biome-specific gene catalogues. *Microbiome* **10**, 72 (2022).
17. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
18. Phillippy, A. M., Schatz, M. C. & Pop, M. Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol.* **9**, R55 (2008).
19. Vezzi, F., Narzisi, G. & Mishra, B. Reevaluating assembly evaluations with feature response curves: GAGE and assemblathons. *PLoS ONE* **7**, e52210 (2012).
20. Lai, S. et al. metaMIC: reference-free misassembly identification and correction of de novo metagenomic assemblies. *Genome Biol.* **23**, 242 (2022).
21. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* **32**, 1088–1090 (2016).
22. Ding, Y., Xiao, J., Zou, B., Yang, C. & Zhang, L. DeepMM: identify and correct metagenome misassemblies with deep learning. Preprint at bioRxiv <https://doi.org/10.1101/2025.02.07.637187> (2025).
23. Chen, Y., Zhang, Y., Wang, A. Y., Gao, M. & Chong, Z. Accurate long-read de novo assembly evaluation with Inspector. *Genome Biol.* **22**, 312 (2021).
24. Li, K., Xu, P., Wang, J., Yi, X. & Jiao, Y. Identification of errors in draft genome assemblies at single-nucleotide resolution for quality assessment and improvement. *Nat. Commun.* **14**, 6556 (2023).
25. Madrigal, G., Minhas, B. F. & Catchen, J. Klumpy: a tool to evaluate the integrity of long-read genome assemblies and illusive sequence motifs. *Mol. Ecol. Resour.* **25**, e13982 (2025).
26. Chen, Q., Yang, C., Zhang, G. & Wu, D. GCI: a continuity inspector for complete genome assembly. *Bioinformatics* **40**, btae633 (2024).
27. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
28. Schelkunov, M. I. Mabs, a suite of tools for gene-informed genome assembly. *BMC Bioinform.* **24**, 377 (2023).
29. Ko, B. J. et al. Widespread false gene gains caused by duplication errors in genome assemblies. *Genome Biol.* **23**, 205 (2022).
30. Liao, W.-W. et al. A draft human pangenome reference. *Nature* **617**, 312–324 (2023).
31. Schoelmerich, M. C. et al. Borg extrachromosomal elements of methane-oxidizing archaea have conserved and expressed genetic repertoires. *Nat. Commun.* **15**, 5414 (2024).
32. Tucker, S. J. et al. A high-resolution diel survey of surface ocean metagenomes, metatranscriptomes, and transfer RNA transcripts. *Sci. Data* **12**, 1913 <https://doi.org/10.1038/s41597-025-06166-3> (2025).
33. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
35. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
36. Trigodet, F., Sachdeva, R., Banfield, J. F. & Eren, A. M. anvi-script-find-misassemblies: a program in anvi'o software ecosystem to quantify read clipping events. *GitHub* <https://github.com/merenlab/anvio> (2025).
37. Watson, M. & Warr, A. Errors in long-read assemblies can critically affect protein prediction. *Nat. Biotechnol.* **37**, 124–126 (2019).
38. Hackl, T. et al. proovframe: frameshift-correction for long-read (meta)genomics. Preprint at bioRxiv <https://doi.org/10.1101/2021.08.23.457338> (2021).
39. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019).
40. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
41. Ma, B. et al. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources. *Nat. Commun.* **14**, 7318 (2023).
42. Nayfach, S. et al. A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* **39**, 499–509 (2021).
43. Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
44. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2011).
45. Harada, R. et al. A cellular entity retaining only its replicative core: hidden archaeal lineage with an ultra-reduced genome. Preprint at bioRxiv <https://doi.org/10.1101/2025.05.02.651781> (2025).
46. Liao, X. et al. Current challenges and solutions of de novo assembly. *Quant. Biol.* **7**, 90–109 (2019).
47. Kresser, T. & Remm, M. Characterization of species-specific repeats in 613 prokaryotic species. *DNA Res.* **19**, 219–230 (2012).
48. Shaw, J., Marin, M. G. & Li, H. High-resolution metagenome assembly for modern long reads with myloasm. Preprint at bioRxiv <https://doi.org/10.1101/2025.09.05.674543> (2025).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Methods

The URL <https://merenlab.org/data/benchmarking-long-read-assemblers/> presents our bioinformatics workflow for reproducing our findings or applying the same approaches to evaluate additional assemblers or datasets.

## Datasets

We downloaded a set of publicly available HiFi PacBio metagenomes matching the ones used in long-read assembler publications. To complete and expand the set of biomes, we included eight surface ocean metagenomes. Supplementary Table 1 includes a comprehensive description of these data and their accession numbers.

## Assembly algorithms

For our primary benchmarks, we used four different assemblers: (1) HiCanu v2.2 (ref. 14), (2) hifiasm-meta v0.3 (ref. 11), (3) metaFlye v2.9.5 (ref. 13) and (4) metaMDBG v1 (ref. 12). In the case of HiCanu, we used the same parameters<sup>11</sup> previously used to assemble metagenomes with HiCanu: maxInputCoverage=1000 genomeSize=100 m bat-Memory=200 and -pacbio-hifi. We used hifiasm-meta with the default parameters. For metaFlye, we used the parameters –meta–pacbio-hifi. We included a total of four versions of metaMDBG in our analyses: v0.3 (the original, published version of the software), v1.0 (released on GitHub in August 2024), v1.1 (released on GitHub in December 2024) and v1.2 (released on GitHub in August 2025). The results for these additional versions can be found in Supplementary Information while the results from v1 are shown in the Article. For metaMDBG v1 and later versions, we used the parameters –in-hifi. In a late revision to this paper, we included the results of myloasm v0.2 in Supplementary Table 2. We used the metagenomic reads without previous filtration or processing, for consistency with previously published assembly benchmarks<sup>11–13</sup>.

## Read mapping

We mapped the metagenomics long reads back to their respective assembly by the four assemblers using minimap2 v2.28 (ref. 33) with the following parameters: -ax map-hifi -p1-secondary-seq. This set of parameters allows secondary mapping when the alignment score is as good as the primary mapping score, that is, multi-mapping, as well as to keep the sequence for secondary mapping in the output files so that secondary mapping is properly considered in downstream analyses. We processed the resulting alignment files using samtools v1.17 (ref. 35).

## Processing of assembled contigs and read mapping results

We used anvi'o development branch of v8.1 (ref. 49) to generate contig databases for each assembly with the command ‘anvi-gen-contigs-database’, which performed gene calls using prodigal v2.6.3 (ref. 50). The anvi'o programs ‘anvi-run.ncbi-cogs’ and ‘anvi-run-kegg-kofams’ annotated genes with functions using the Clusters of Orthologous Genes<sup>51</sup> database of NCBI and KOfams<sup>52</sup> by the Kyoto Encyclopedia of Genes and Genomes (KEGG), respectively, while ‘anvi-run-hmms’ identified single-copy core genes in these sequences, which we associated with taxonomy data from the Genome Taxonomy Database (GTDB)<sup>53</sup> using ‘anvi-run-scg-taxonomy’. We also used Kaiju v1.10.1 (ref. 54) to get gene-level taxonomy with the nr\_euk database. We estimated the completeness of KEGG KOfam modules using ‘anvi-estimate-metabolism’. We finally used ‘anvi-profile’ to process the read mapping data to recover coverage values, SNVs and INDELs.

## Pangenomic analyses

To compute a pangenome for *Methanothrix*, we first acquired publicly available genomes from the RefSeq database of NCBI using the program ‘ncbi-genome-download’ (available at <https://github.com/kblin/ncbi-genome-download>) with the parameters ‘–assembly-level all-genera Methanothrix’. For the *E. coli* pangenome, we downloaded the original genomes used to create the mock dataset (available at

<https://s3.amazonaws.com/zymo-files/BioPool/D6331.refseq.zip>). For both pangenomics analysis, we used the program ‘anvi-run-workflow’ to run the anvi'o pangenomics workflow implemented in Snakemake<sup>55</sup>, which used DIAMOND v2.1.8 (ref. 56) to identify gene clusters as described previously<sup>57</sup>. We used the program ‘anvi-display-pan’ to visualize and summarize the pangenomes.

## Identification of assembly errors

To identify potential assembly errors based on clipping events, we developed a program within the anvi'o platform, ‘anvi-script-find-misassembly’<sup>36</sup> (help page: <https://anvio.org/m/anvi-script-find-misassemblies>), which takes a single BAM file of long-read mapping results. The script searches for premature end of alignments, that is, clipping events, and reports positions in which a proportion of reads that are clipped exceeds a user-defined threshold. The script also reports regions in contigs with no coverage. We investigated the region with no apparent coverage by using BLAST v2.16.0 (ref. 58) to search for assembled sequences over 500 bp with no apparent coverage against the long reads directly. We used the flag ‘-dust no’ to include regions with low sequence complexity. We compared the *k*-mer content (*k* = 21) of these regions with the original long read using Meryl v1.3 (ref. 17). In addition, we used BLAST v2.16.0 (ref. 58) to identify contigs covered by at least 70% of repeated sequences. We BLASTed each contig against itself using BLASTN with default parameters, excluded the perfect reciprocal hit, transformed the remaining hits into a Browser Extensible Data (BED)-formatted file, and used bedtools v2.31.1 (ref. 59) to compute the breadth of coverage. We computed additional statistics using a Python script available in our reproducible workflow.

## Manual inspection of mapping results

We used IGV v2.17.4 (ref. 60) and the anvi'o interactive interface to manually inspect genomic regions of interest and generate figures.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All metagenomes used in our study are publicly available through the NCBI, and Supplementary Table 1 lists their accession numbers. DOI URLs for intermediate data products are available at <https://merenlab.org/data/benchmarking-long-read-assemblers/>. They include the assemblies (<https://doi.org/10.6084/m9.figshare.29107748.v3> (ref. 61)), the anvi'o contigs and profile databases (<https://doi.org/10.6084/m9.figshare.29246210> (ref. 62)), the outputs of the script anvi-script-find-misassemblies (<https://doi.org/10.6084/m9.figshare.29279228> (ref. 63)) and the two pangenomes of *Methanothrix* and *E. coli* (<https://doi.org/10.6084/m9.figshare.29864903> (ref. 64)).

## Code availability

The script anvi-script-find-misassemblies used in this study is available in anvi'o<sup>36</sup> (<https://github.com/merenlab/anvio>). A fully reproducible workflow is available at <https://merenlab.org/data/benchmarking-long-read-assemblers/>.

## References

49. Eren, A. M. et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* **6**, 3–6 (2021).
50. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
51. Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, D261–D269 (2015).

52. Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
53. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
54. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
55. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
56. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
57. Delmont, T. O. & Eren, A. M. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* **6**, e4320 (2018).
58. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
59. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
60. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant review with the Integrative Genomics Viewer. *Cancer Res.* **77**, e31–e34 (2017).
61. Trigodet, F. Long-read assemblies. *figshare* <https://doi.org/10.6084/m9.figshare.29107748.v3> (2025).
62. Trigodet, F. Anvi'o databases. *figshare* <https://doi.org/10.6084/m9.figshare.29246210.v16> (2025).
63. Trigodet, F. Outputs of anvi-script-find-misassemblies. *figshare* <https://doi.org/10.6084/m9.figshare.29279228.v1> (2025).
64. Trigodet, F. Pangenomes of *Methanothrix* and *E. coli*. *figshare* <https://doi.org/10.6084/m9.figshare.29864903.v1> (2025).

## Acknowledgements

We recognize that developing assembly algorithms, especially for metagenomes, is a notoriously complex and difficult task, and we express our deep appreciation of those who invest their time and skills in creating and maintaining them.

## Author contributions

F.T., R.S., J.F.B. and A.M.E. conceptualized the study. F.T. curated data, performed formal analyses and developed software tools. F.T., R.S., J.F.B. and A.M.E. analyzed the data and interpreted the findings. F.T. and A.M.E. wrote the original draft of the study, and all authors commented on the paper, made suggestions and approved the final paper.

## Funding

Open access funding provided by Max Planck Society.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41587-025-02971-8>.

**Correspondence and requests for materials** should be addressed to Jillian F. Banfield or A. Murat Eren.

**Peer review information** *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted
  - Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection No new data were collected. Our study made use of publicly available metagenomes, accession IDs of which are reported in Table 1.

Data analysis We used HiCanu v2.2, hifiasm-meta v0.3, metaFlye v2.8.5 and metaMDBG v0.3, v1, v1.1 to assemble publicly available PacBio metagenomes. We used minimap2 v2.28 and samtools v1.17 for long-read mapping. For all downstream analyses, we used anvi'o (development branch of v8.1 <https://anvio.org>), prodigal v2.6.3, DIAMOND v2.1.8, BLAST v2.16.0, bedtools v2.31, and IGV v2.17.4. Ad hoc or in-house scripts we have implemented are available at <https://merenlab.org/data/benchmarking-long-read-assemblers/> as a part of our reproducible bioinformatics workflow.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All metagenomes used in our study are publicly available through the NCBI, and Table 1 lists their accession IDs. DOIs for intermediate data products are available at <https://merenlab.org/data/benchmarking-long-read-assemblers/>

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

### Reporting on sex and gender

*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*

### Reporting on race, ethnicity, or other socially relevant groupings

*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.). Please provide details about how you controlled for confounding variables in your analyses.*

### Population characteristics

*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*

### Recruitment

*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*

### Ethics oversight

*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Study description

Our study evaluates assembly software designed to process long-read metagenomes.

### Research sample

Publicly available PacBio metagenomes that originate from animal guts, anaerobic digesters, and marine systems.

### Sampling strategy

No sampling strategy was considered as the study primarily focused on datasets that were either used by the original publication of the algorithms evaluated or what was available from systems that were not considered by individual publications, such as the marine system.

### Data collection

Data collection was limited to downloading publicly available metagenomes.

### Timing and spatial scale

N/A

Data exclusions	N/A
Reproducibility	Our study does not include experiments.
Randomization	N/A
Blinding	N/A

Did the study involve field work?  Yes  No

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies	<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Eukaryotic cell lines	<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	Palaeontology and archaeology	<input checked="" type="checkbox"/>	MRI-based neuroimaging
<input checked="" type="checkbox"/>	Animals and other organisms		
<input checked="" type="checkbox"/>	Clinical data		
<input checked="" type="checkbox"/>	Dual use research of concern		
<input checked="" type="checkbox"/>	Plants		

## Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.