PacBio

Application note

CiFi: Long-read multi-contact 3C for chromosome-scale assemblies using PacBio HiFi reads

Chromatin architecture dictates how the genome is organized, accessed, and regulated across cell types, which can influence gene expression and drive phenotypic variation. Comprehensive genome sequencing data integrated with chromatin interaction information is crucial for complete, chromosome-scale genome assembly. CiFi (pronounced “Sci-Fi”) is a new application developed by the sequencing community that integrates chromatin conformation capture (3C) with PacBio® HiFi long-read sequencing to generate highly accurate, multi-contact reads.

Traditional Hi-C methods rely on short reads, limiting mapping resolution in repetitive or structurally complex regions. CiFi overcomes these challenges by producing long, concatemeric HiFi reads while requiring low DNA input. This amplification-enhanced workflow enables chromosome-scale, haplotype-resolved assemblies and comprehensive chromatin interaction analysis using a single-sequencing technology (McGinty, Kaya, et al., 2025).

102-326-662 REV01 JAN2026

PacBio

CiFi method overview

CiFi combines traditional Hi-C approaches with HiFi sequencing. The workflow begins with a 3C library preparation (Figure 1). Formaldehyde crosslinking preserves native chromatin interactions, followed by restriction digestion to fragment the genome. Proximity ligation then joins spatially adjacent fragments into multi-contact concatemers.

After crosslink reversal, a high-fidelity amplification step, using the PacBio Ampli-Fi protocol enriches for intact un-crosslinked ligation products. This step is necessary for improving overall sequencing yield and read length. An optional size selection can be performed prior to amplification to enrich for longer concatemers. After amplification, proximity-ligated fragments enter a standard SMRTbell® library preparation workflow.

Sequencing can be performed on any PacBio HiFi system (e.g., Revio®, Vega™, Sequel® II). Each CiFi read contains multiple chromatin-interacting segments (Figure 1), which are digested *in silico* prior to downstream bioinformatic pipelines used for identifying contact pairs, haplotype phasing, and scaffolding. If desired, a standard HMW DNA sample can be combined with a CiFi sample ahead of library preparation, allowing for a single-library, chromosome-scale (SLC, pronounced "slick") assembly workflow (Figure 2).

See supplemental information of McGinty, Kaya, et. al. (2025) for additional method details.

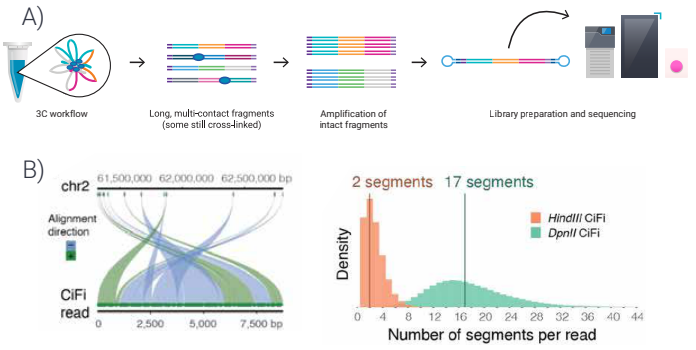

Figure 1. Figure from McGinty, Kaya, et al. (2025). CiFi combines Hi-C with HiFi sequencing to create long multi-contact fragments that can be sequenced on any PacBio long-read system for improved phasing and assembly scaffolding. (A) Overview of the CiFi approach using an amplification step to remove residual cross links from the 3C workflow. (B) Mapping of multi-contact segments (*Dpn*II), a CiFi read against chromosome 2 in the human genome, and the distribution of number of segments for each RE digest (*Dpn*II—4 bp cutter, *Hind*III—6 bp cutter).

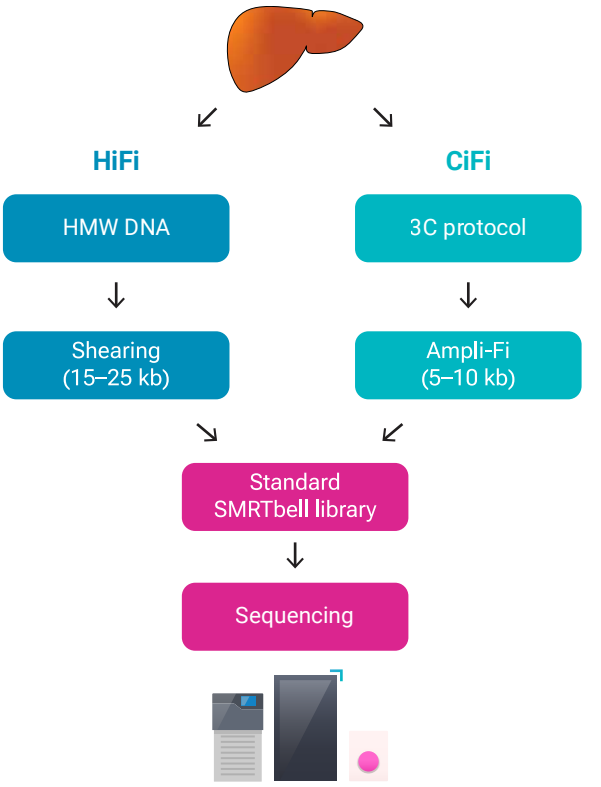

Figure 2. High- level overview of a single-library, chromosome-scale assembly workflow.

Key advantages of CiFi

Improved mapping in repetitive regions

Long multi-contact reads enable alignment across complex genomic landscapes—including centromeres, segmental duplications, and transposable elements—where short-read Hi-C struggles.

Low input requirements

Optimized amplification and library preparation allow robust performance from limited starting material (tens of thousands of cells or single small organisms), expanding access to 3C-based analyses.

Multi-contact reads

Each HiFi read captures multiple proximally interacting segments, providing a more complete view of chromatin structure than conventional paired-end Hi-C.

Single-platform, streamlined solution

CiFi data supports haplotype phasing and chromosome-scale assembly, allowing researchers to obtain both genome sequencing and long-range scaffolding information from a single technology and single sequencing run with the SLC workflow.

102-326-662 REV01 JAN2026

CiFi for haplotype phased chromosome-scale de novo assembly

McGinty, Kaya, et al. (2025) demonstrate that CiFi overcomes two major limitations of previous long-read chromatin-capture technologies: **high DNA input** and **poor mapping resolution in repetitive regions**. Earlier long-read 3C methods such as Pore-C required very large inputs (e.g., ~10 million cells). In contrast, CiFi achieves high performance from as few as 62,000 cultured cells or a single small organism, enabled by the efficient amplification strategy described in the Ampli-Fi protocol.

CiFi also improves mapping resolution because proximity-ligated fragments typically range from ~350 bp to 2 kb depending on the restriction enzyme (e.g. *Dpn*II or *Hind*III). In the McGinty, Kaya, et al. (2025) study, CiFi increased contact resolution across transposable elements, segmental duplications, and centromeres. This improved resolution in complex genomic regions supported performing haplotype-phased, chromosome-scale assembly.

How much CiFi data is needed?

Because each CiFi read contains multiple chromatin contacts, coverage requirements differ from traditional short-read Hi-C, which captures only two contacts per read pair. Depending on the restriction enzyme, the mean contacts per HiFi read can range from 2–17 contacts (McGinty, Kaya, et al., 2025). Because each of those contacts within a single read can be combined into a pair, the number of pairwise interactions per read follows this formula:

$$\frac{n(n-1)}{2}$$

For example, a read with 17 contact segments provides 136 pairwise interactions. Thus, generating ~100 million pairwise contacts from a *Dpn*II experiment (mean of 17 contacts per read) requires fewer than 1 million HiFi reads, about **one-eighth of the output of a typical Revio SMRT® Cell with SPRQ™ chemistry)**. This enables pooling CiFi and WGS libraries on the same SMRT Cell to obtain both assembly and scaffolding data, as demonstrated by McGinty, Kaya, et al. (2025) in the Mediterranean fruit fly and in the vole assemblies described in the right column.

In terms of coverage across the genome, downsampling results from the McGinty, Kaya, et al. (2025) Mediterranean fruit fly experiment indicate that it may be possible to use as low as ~1.5x coverage of CiFi data before scaffolding performance declines (see Supplementary Fig. S8 in McGinty, Kaya, et al., 2025). However, the fruit fly is a relatively simple genome, and coverage requirements will likely vary depending on the restriction enzyme being used (number and length of contact segments) and the genome complexity of a particular organism. A 5–10x coverage of CiFi data (*Hind*III) may be sufficient to match or outperform traditional short-read Hi-C for phasing and scaffolding human genomes based on early reports from the community (unpublished data).

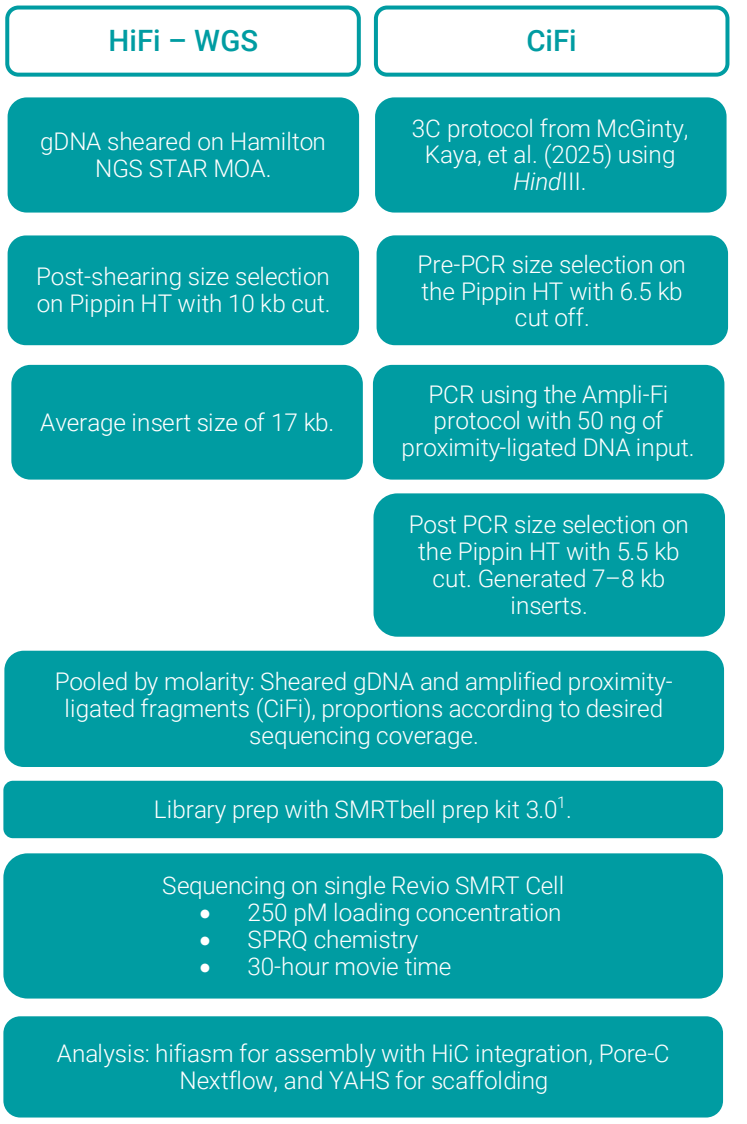

Figure 3. Detailed workflow for Meadow and Prairie vole samples

Chromosome-scale assembly of the prairie and meadow vole using a single Revio SMRT Cell

To establish a practical workflow for generating chromosome-scale assemblies from a single Revio SMRT Cell, PacBio, in collaboration with researchers at UC Davis, sequenced pooled CiFi and WGS libraries from prairie and meadow voles. The detailed workflow is described above in Figure 3.

*Hind*III-digested CiFi libraries were amplified and then spiked into sheared genomic DNA at a target of 10% (by molarity) prior to SMRTbell library preparation. Each pooled HiFi + CiFi library was sequenced on one Revio SMRT Cell using SPRQ chemistry. Pooling accuracy was high, with 11–12% of reads originating from the CiFi fraction, and >99.8% classified as unique reads (Table 1 and Figure 4).

Table 1. Sequencing results for pooled HiFi WGS and CiFi libraries for the meadow and prairie voles. (A) Primary sequencing metrics for the HiFi and CiFi reads for the Meadow and Prairie vole.

Sample	Meadow vole		Prairie vole	
Total yield	98.2 Gb		102.7 Gb	
Total reads	7,334,014		7,216,449	
Reads with SMRTbell index	96.9%		97.4%	
Demultiplexed yield	84.2 Gb HiFi (86%)	11.8 Gb CiFi (12%)	89.5 Gb HiFi (87%)	11.2 Gb CiFi (11%)
Demultiplexed reads	5,563,158 HiFi (76%)	1,544,185 CiFi (21%)	5,662,482 HiFi (79%)	1,366,318 CiFi (19%)
Unique reads		1,542,123 (99.9%)		1,363,853 (99.8%)

CiFi data were digested *in silico* and supplied to hifiasm (Cheng et al. 2021) together with HiFi WGS reads for a Hi-C–integrated assembly. CiFi pairs are then quality controlled and aligned to the assembly using the Pore-C nextflow workflow and scaffolding performed with YaHS. The resulting un-curated assemblies (Table 2) show good continuity and genome size comparable to the Vertebrate Genome Project's mMicPen (2024) reference for the meadow vole.

Telomeric sequence was detected on both ends of many scaffolds indicating that the combined HiFi + CiFi strategy can routinely generate chromosome-scale, reference-quality assemblies. CiFi strategy can routinely generate chromosome-scale, reference-quality assemblies.

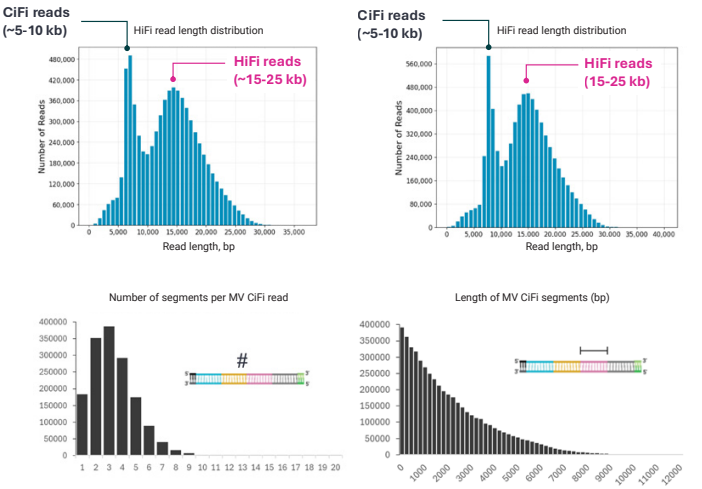

Figure 4. (A) Read length distribution on Revio for the SLC libraries (HiFi + CiFi) of the meadow (left) and prairie (right) voles. (B) Histograms for the number of *Hind*III segments per CiFi read and the length segments for the meadow vole sample.

Table 2. Un-curated assembly stats for the meadow and prairie voles.

Sample	Meadow vole		Prairie vole	
Statistic	Haplotype 1	Haplotype 2	Haplotype 1	Haplotype 2
Size (bp)	2.185 B	2.324 B	2.334 B	2.495 B
Scaffolds	106	56	95	63
N50 (bp)	113 M	129 M	114 M	115 M
L50	8	7	9	8
auN	112 M	128 M	104 M	137 M

References

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170-175. https://doi.org/10.1038/s41592-020-01056-5

McGinty, S. P., Kaya, G., et al. (2025). CiFi: accurate long-read chromosome conformation capture with low-input requirements. *Nature Communications*. https://doi.org/10.1038/s41467-025-66918-y

Microtus pennsylvanicus genome assembly mMicPen1.hap1 (GCF_037038515.1). *NCBI Datasets*. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_037038515.1/

Resources

PacBio Ampli-Fi protocol

102-326-662 REV01 JAN2026