PacBio

Application note

HiFiViral SARS-CoV-2 kit enables genomic surveillance of COVID-19

Introduction

The COVID-19 pandemic is an ongoing global epidemiological challenge, with the repeated emergence of new strain lineages that are more contagious, virulent, drug resistant and in certain cases, able to evade vaccine-induced immunity. As a result, early and reliable detection of variant lineages is essential for guiding public health interventions.

To this end, researchers perform whole genome sequencing of the SARS-CoV-2 virus to identify and characterize new virus variants. PCR amplification-

based assays are most commonly used to enrich viral genomes for sequencing, however these workflows are complex and hard to scale. Furthermore, they are susceptible to coverage dropout driven by new mutations, which can potentially result in missing novel variants as they emerge.

PacBio

HiFiViral SARS-CoV-2 kit

The HiFiViral SARS-CoV-2 kit is an end-to-end solution for genomic surveillance of COVID-19 variants. The kit utilizes molecular inversion probes (MIPs), a differentiated enrichment technology which combines the simplicity and cost-effectiveness of PCR with the resilience to mutations of target capture. The MIPs approach, coupled with PacBio HiFi reads and SMRT® Link software analysis, delivers several advantages over PCR amplicon sequencing, including:

- A simple workflow with fewer steps and touchpoints, enabling labs to rapidly scale sequencing

- Deep tiling that provides resilience to new mutations and emerging variants

- Robust genome coverage across varying viral input quantities

- The ability to call variants of all types including single nucleotide variants, indels, and structural variants due to high accuracy reads of ~800 bp

- The ability to detect the presence of multiple strains in one sample

Streamlined end-to-end workflow

This application note describes a streamlined workflow and performance characteristics for genomic surveillance of COVID-19 using the HiFiViral SARS-CoV-2 kit, sequencing on the Sequel® IIe system, and analysis using SMRT Link v11.0 (Figure 1).

Figure 1. HiFiViral is a complete, end-to-end workflow

RNA extraction

The HiFiViral workflow was optimized for use with RNA extracted from nasopharyngeal swabs. The kit has been paired successfully with many commercially available viral RNA extraction kits, including MagMax Viral/Pathogen Nucleic Acid Isolation Kit from ThermoFisher, Quick-DNA/RNAViral MagBead from Zymo Research, and QIAamp viral RNA kits.

For best results, extracted RNA should not be subject to repeated freeze-thaw cycles.

Preparation of extracted RNA samples

The kit contains all reagents needed for viral enrichment and barcoding of up to 384 samples that are pooled into one SMRTbell library and run on a single SMRT® Cell 8M (Figure 2). The enrichment assay features a 4-step, addition-only workflow that minimizes plate transfers and includes pre-mixed reagents. Potential pipetting errors are further limited by a color-change feature to give visual confirmation that each reagent has been properly added.
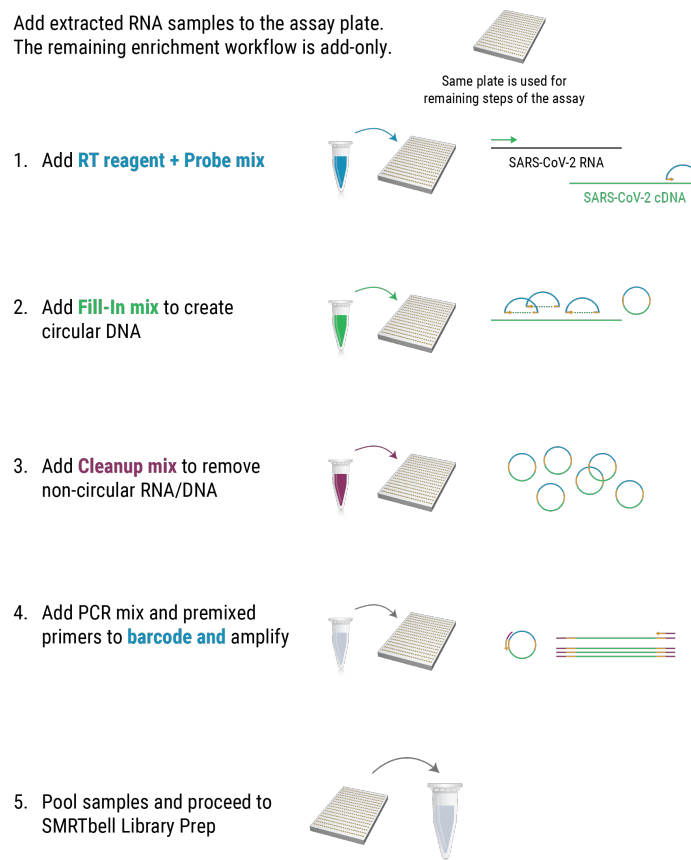

Figure 2. The HiFiViral enrichment workflow

To begin the assay, RNA samples are added to the reaction plate, and RT + Probe Mix is added to each sample well. The HiFiViral SARS-CoV-2 probe mix uses ~1,000 MIPs to tile the entire genome, such that each base is covered by on average 22 probes.

Next, Fill-in mix is added to capture viral sequences via a cDNA-templated fill-in reaction that creates circular DNA molecules. In step 3, non-circular DNA is digested by the Cleanup mix, which preserves the enriched viral sequences and removes residual probe molecules and contaminating non-SARS-CoV-2 DNA or cDNA.

Finally, pre-mixed dual-indexed primers and PCR Mix are added to the assay plate to create and amplify barcoded, double-stranded DNA from the circular templates. Products from the completed enrichment workflow are pooled for SMRTbell library preparation using SMRTbell prep kit 3.0 (102-182-700) and sequenced on one SMRT Cell 8M (101-389-001) with Binding kit 3.1 (102-333-400) and Sequel II sequencing 2.0 kit (101-820-200) on a Sequel IIe system (101-986-400). The full workflow requires 28-42 hours starting from extracted RNA through the end of analysis.
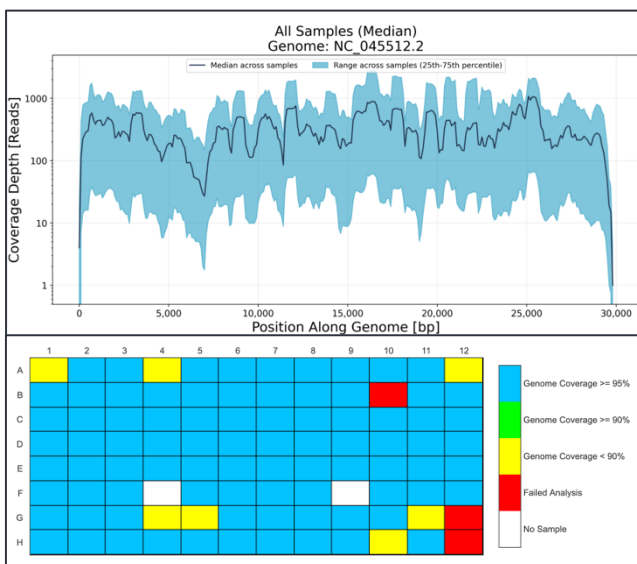

Figure 3. Sample QC is simplified with genome coverage (a) and plate QC plots (b)

Simple analysis with SMRT Link v11.0

The HiFiViral SARS-CoV-2 application in SMRT Link v11.0 enables automated, streamlined detection and analysis of virus variants. The software tools "lima" and "mimux" perform sample demultiplexing, read QC, and trim synthetic bases to produce a set of high-quality sequence reads representing each viral sample.

The processed reads are passed to "pbmm2" for mapping to the Wuhan reference, "bcftools" for variant calling, and "vcfcons" for generating sample consensus sequences. The pipeline requires 4-fold coverage of HiFi reads to produce a consensus base.

In some cases, with highly mutated variants, "bcftools" may filter out reads containing many nucleotide differences from the reference. By using a matched reference rather than the default Wuhan reference, users may avoid coverage gaps or errors in the consensus sequence.

The final output is a report including QC metrics, variant calls, and consensus sequences ready for submission to surveillance databases. In addition, the report flags samples that may contain multiple strains for follow-up experiments and analysis.

HiFiViral SARS-CoV-2 performance on RNA control samples

To demonstrate assay performance, commercial control samples were prepared according to the HiFiViral SARS-CoV-2 kit protocol (102-132-000). Twist Synthetic SARS-CoV-2 RNA controls 14, 15, 16, and 17, representing lineages B.1.1.7, B.1.351, and P.1 were serially diluted to mimic 8 different Ct values ranging from 19 to 35 (Han et al., 2021). Three replicates were run for each condition, for a total of 96 samples. The 96 samples were pooled into one SMRTbell library and sequenced on a single SMRT Cell 8M. This dataset can be downloaded from our website.
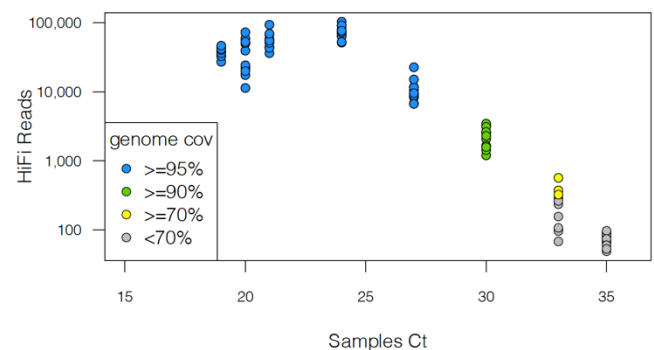

Figure 4. Assay performance on RNA control samples at 96-plex

Figure 4 shows HiFi read count versus sample Ct for the RNA controls. With a minimum of 1,000 mapped HiFi reads, samples achieve complete genome coverage, including all samples with Ct 30 and lower. Some of the Ct 33 samples have 70% genome coverage as indicated with yellow data points, and these may still be used for strain assignment.

HiFiViral SARS-CoV-2 performance on surveillance samples

Performance was also evaluated on surveillance samples. Figure 5 shows the results of processing and sequencing a typical batch of samples with Ct range of 9 to 28, plus 9 samples with unknown Ct values.
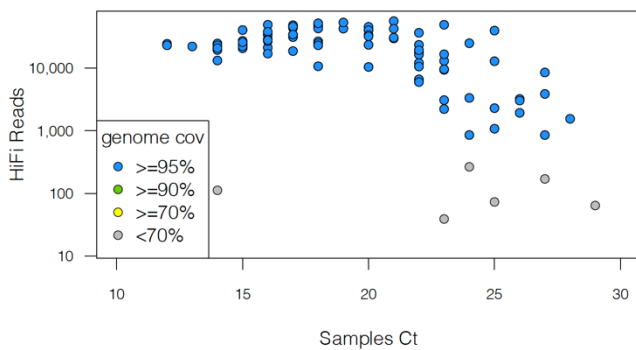

Figure 5. Assay performance on surveillance samples at 96-plex

Again, samples with at least 1,000 mapped HiFi reads yield complete genome coverage. In this experiment, samples with known Ct had an overall success rate of 83%. All controls performed as expected. This dataset can be downloaded from our website.

Sample type	No. of samples	≥ 90% Coverage
NP extracts, known Ct	84	83%
NP extracts, unknown Ct	9	44%
Twist controls	2	100%
Negative control	1	0

Table 1. Genome completeness of 96-plex surveillance samples

Performance at high throughput was demonstrated on a combination of control and surveillance samples at 380-plex. For the 380-plex run, 78% of 164 predominantly delta lineage samples had >95% genome completeness.
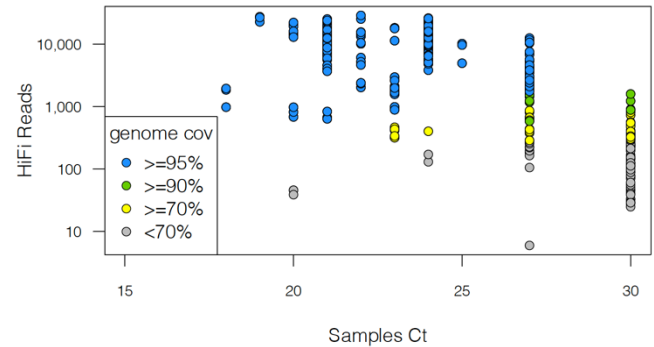

Figure 6. HiFiViral performance at 380-plex.

Comprehensive detection of mutation types and lineages

The HiFiViral assay is able to characterize all mutation types, from SNVs and indels that are key to strain assignment for surveillance to structural variants which contain interesting biology. Table 2 shows precision and recall across 7 different control samples from different variant lineages. With precision greater than 99% and recall 95% or higher, the results yield perfect accuracy for strain assignment with Nextstrain (Hadfield et al., 2018).

Control sample	Nextclade assignment	Precision	Recall	Nextstrain accuracy
Twist 01	19A	1	94.8%	100%
Twist 13	20C	1	99.7%	100%
Twist 14	20I (Alpha, V1)	1	99.9%	100%
Twist 15	20I (Alpha, V1)	1	99.9%	100%
Twist 16	20H (Beta, V2)	1	100%	100%
Twist 17	20J (Gamma, V3)	1	100%	100%
Twist 23	21A (Delta)	99.1%	99.4%	100%

Table 2. SNV calling and strain assignment of controls in 384-plex

PacBio

The HiFiViral assay also has robust performance against Omicron samples. Data from real-world samples analyzed during the fall of 2021 and winter of 2022 at 4 distinct sites show that performance remained consistent for runs comprised of samples from diverse lineages (Site 1, 95-plex), predominantly Delta (Site 2, 380-plex or Site 3, 87-plex), or predominantly Omicron (Site 4, 35-plex), as shown in Figure 7.
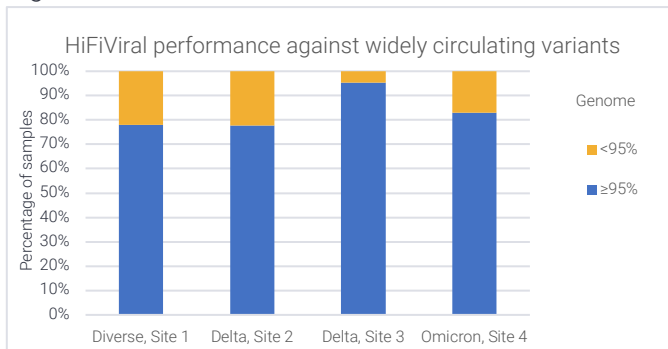

Figure 7. HiFiViral performance against widely circulating variants

The combination of HiFi accuracy and 675 bp captured sequences also enables detection of structural variants. Figure 8 shows VCF tracks of 4 large deletions in the *orf7a* gene, which range from 87 bp to 271 bp. The aligned reads cleanly span the two largest deletions.
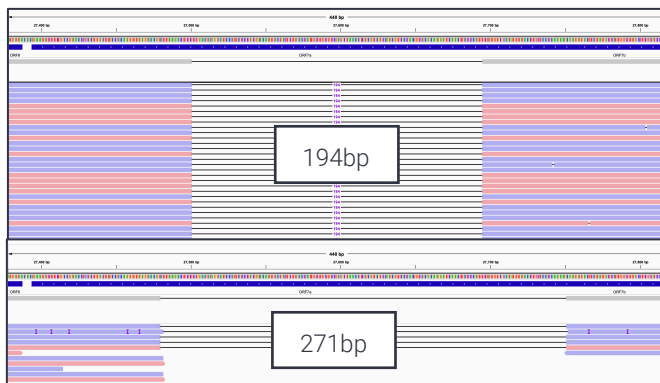

Figure 8. HiFiViral captures large viral genome deletions

Finally, HiFi reads can detect the presence of minor variants in the sample. The presence of multiple strains may be the result of sample contamination or other lab error, multi-strain coinfection, or the occurrence of a *de novo* viral mutation. A binomial model is used to identify samples with multiple strains. The algorithm can detect multiple strains in samples with Ct <26 when the minor strain is 20% or more of the pool.

Figure 9 shows two examples of multiple strains in surveillance samples. The highlighted boxes in the top sample show minor variants that are distinguishable at 6 sites in a 9 kb region around the *orf1ab* and *spike* genes. The bottom sample shows a 50/50 mix of two strains with distinct SNVs at 7 sites.
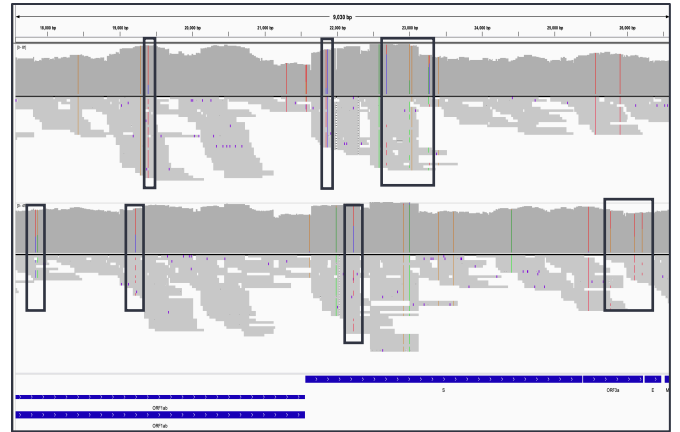

Figure 9. The HiFiViral workflow can flag samples that may have multiple strains present

Reflex sequencing of high Ct samples with low coverage

In cases where very precious samples with high Ct values generate at least 100 reads but not the 1,000 HiFi reads needed to recover a complete genome, several strategies can be employed to improve results. In some cases, rerunning the same enriched sample in a subsequent run with a lower multiplex level will recover the sample. Coverage can be further increased by merging the bam files from the initial and reflex runs and resubmitting to the HiFiViral SMRT Link pipeline. As a final option, the enrichment step can be repeated to generate additional template for sequencing.

Summary

The identification and characterization of emerging variants is central to managing the COVID-19 pandemic and protecting public health.

The HiFiViral SARS-CoV-2 kit offers a highly accurate end-to-end solution that is simple, robust, cost-effective, and well-suited for scalable surveillance of a rapidly evolving virus to inform public health decisions.

Resources and References

Resources

Procedure & checklist – PacBio HiFiViral high throughput multiplexing for full-viral genome sequencing of SARS-CoV-2 using SMRTbell prep kit 3.0

PacBio consumable part numbers:

Part Number	Item
102-132-000	HiFiViral SARS-CoV-2 Kit
102-182-700	SMRTbell® Prep Kit 3.0
101-389-001	SMRT® Cell 8M Tray
102-333-400	Sequel® II Binding Kit 3.1
101-820-200	Sequel® II Sequencing 2.0 Kit
101-986-400	Sequel® IIe System

References

Hadfield, J., et al. (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23), 4121-4123.

Han, M.S., et al. (2021) RT-PCR for SARS-CoV-2: quantitative versus qualitative. *The Lancet Infectious Disease*, 21(2), 165.