P727

Extracting HMW DNA from saliva for HiFi sequencing applications

Sarah Kingan, Deborah Moine, Nina Gonzaludo, Shreyasee Chakraborty, Heather Ferrao, Kristina Weber, Christina Dillane*, Mike Tayeb*, Duncan Kilburn,
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, *DNA Genotek Inc., Ottawa, ON, Canada

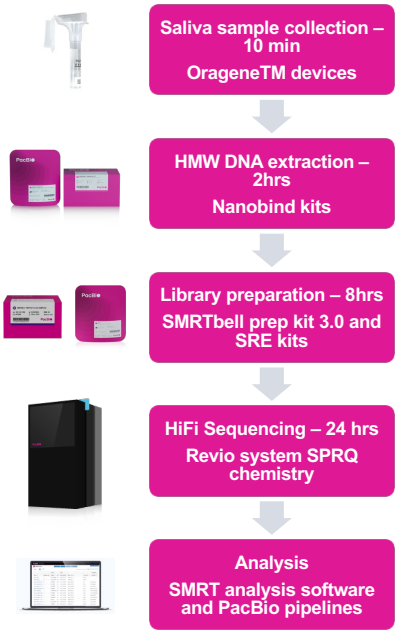
Introduction

Saliva is an attractive DNA source for large-scale genomic and clinical research studies due to its non-invasive collection and ability to be self-collected. However, concerns about DNA quality and yield have limited its use with long-read sequencing, where high-quality, high-molecular-weight (HMW) DNA is key for optimal yield and performance.

In this proof-of-concept study, we demonstrate that high quality PacBio HiFi sequencing results can be obtained from DNA extracted from saliva collected in DNA Genotek Orageneᵀᴹ devices and extracted using the Nanobind PanDNA or CBB kits.

- ✓ **Saliva collected in Orageneᵀᴹ devices is stable at room temperature**
- ✓ **Nanobind kits extract HMW DNA**
- ✓ **DNA from saliva samples is ~75−95% human**

Simple workflow for saliva DNA extraction through analysis



Paired whole blood (200 µL) and saliva were collected from 5 donors and extracted using Nanobind PanDNA or CBB kits followed by library preparation, sequencing and analysis.

High DNA quality from saliva

With Nanobind, 1 to 45 µg of high purity, HMW DNA from 500 µL saliva samples collected in Orageneᵀᴹ devices (N=35). DNA mode size as measured by Femto Pulse is >80 kb for majority of samples .

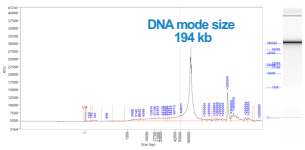


Figure 1. Size distribution of DNA extracted from Saliva donor 1 on Femto pulse system (Agilent technologies)

Similar HiFi sequencing metrics from saliva versus blood

Paired saliva and blood samples were sequenced on a single SMRT® Cell on the Revio® system. We obtained 98 to 135 Gb of HiFi data resulting in 30X - 39X coverage per genome, sufficient for comprehensive WGS variant detection.

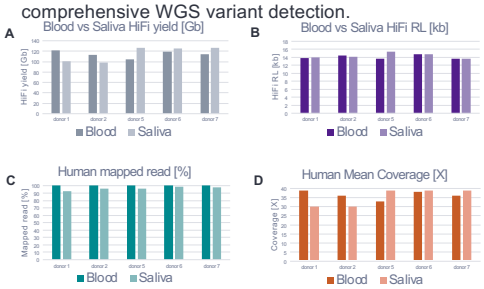


Figure 2. Paired saliva and blood samples from same donor show similar metrics for HiFi yield **(A)**, HiFi read length (RL) **(B)**. For the five saliva samples, percentage of human mapped reads are 92.7% - 98.2% **(C)** giving 30X to 39X human genome coverage **(D)**.

Variant calling with PacBio WGS Variant Pipeline

Each sample was analyzed as a singleton using v3-a1. Code available at:

github.com/PacificBiosciences/HiFi-human-WGS-WDL/

Variant type	Genotyping	Concordance
Single nucleotide variants (SNVs)	Deep Variant v1.8.0	RTG Tools v3.12.1
Structural variants (SVs)	Sawfish v0.12.7	truvari v5.2.0

Table 1. Software tools and version used for genotyping and concordance analysis.

High variant calling concordance

Paired **saliva** and **blood** samples for the same donor were assessed for genotype concordance of high-quality (GQ≥20) variants across GRCh38 without masking.

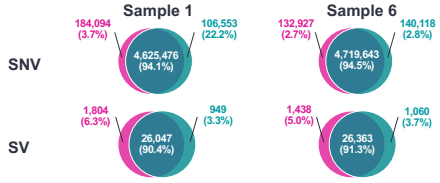


Figure 3. Results for two representative samples #1 and #6. For comparison, a technical replicate of blood from the same donor had concordance of 94.1% for SNVs and 91.5% for SVs and comparison of blood samples from two different donors had concordance of 21.7% for SNVs and 44.6% for SVs (not shown).

Taxonomic profiling of saliva microbiome

Taxonomic profiling analysis on HiFi reads that did not map to human (5 to 25% of reads depending on the sample). Reads were decomposed into k-mers and searched against GenBank db using sourmash.

https://github.com/PacificBiosciences/pb-metagenomics-tools/tree/master/Taxonomic-Profiling-Sourmash

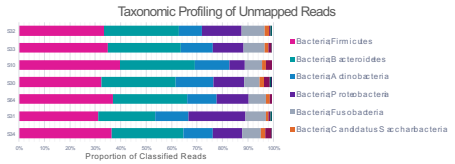


Figure 4. Taxonomic analysis of 7 saliva samples different from paired blood. More than 90% of classified reads are from the five bacterial phyla: Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, and Fusobacteria.

Conclusion

- **Saliva samples collected with Orageneᵀᴹ devices and extracted using Nanobind kits are a good alternative to blood for HiFi sequencing.**
- Similar sequencing performance was obtained from blood and saliva for the same individual.
- Variant calling concordance between blood and saliva from same doner is similar to that between matched blood samples.

Acknowledgements

The authors would like to thank Dan Portik and Billy Rowell for data analysis.

Disclaimer: Oragene is a trademark of DNA Genotek Inc.