HiFi Long-Read WGS: Sequencing Performance of Buccal/Saliva Samples and Estimation of Non-human DNA Contamination

Stuti Joshi[1]; Kenny Chen[1]; Sami Belhadj[1]; John Harting[1]; Maria Keith[1]; Tiantian Geier[1]; Erica Smith[1]; Yun-Hua Hsiao[1]; Bojan Losic[1]; David Brohawn[1]; UCI–GREGoR[2]; Emmanuèle Délot[2]; Seth Berger[1]; Eric Vilain[2]; Rachid Karam[1]
[1]Ambry Genetics. 1 Enterprise, Aliso Viejo, CA; [2] Institute for Clinical and Translational Science, University of California, Irvine, CA

BACKGROUND

- Oral specimen collection, e.g., buccal swabs or saliva, offers non-invasive, autonomous, and patient-accessible options for genetic testing.
- Compared to blood, oral specimens are challenging for long-read whole genome sequencing (LR-WGS), yielding a lower mass of genomic DNA (gDNA), lower integrity, and potential microbial contamination.
- Performance of LR-WGS is shown for 129 buccal and 34 saliva samples sequenced at Ambry Genetics for the UCI-GREGoR site.
- We also show a method to estimate non-human DNA contamination prior to sequencing using qPCR, to inform optimal sample multiplexing on the Revio.

METHODS

Isolation: Blood samples were isolated using the PacBio Nanobind HT CBB kit. Buccal swabs and saliva samples were isolated using the Zymo Quick DNA HMW kit.

QC: gDNA integrity was evaluated on Agilent Femto Pulse. The Genomic Quality Number (GQN) represents the proportion of fragments above 10 kb or 30 kb (GQN10 or GQN30, respectively).

Library Preparation: gDNA underwent mechanical fragmentation, HiFi library preparation[1] and gel-based size selection on the YourGene Health LightBench instrument[2].

Sequencing and Analysis: Libraries were sequenced on PacBio Revio using SPRQ chemistry. Raw data was processed via PacBio's HiFi-human-WGS-WDL workflow[3].

Bacterial Quantification: Bacterial content of samples was approximated on the Quant Studio 6 Flex, using the Zymo Femto Bacterial DNA Quantification Kit[4].

TAKE HOME POINTS

- gDNA isolation from buccal and saliva samples yielded suitable HMW DNA for HiFi LR-WGS.
- Variable contamination levels from the oral microbiome reduces mean sequencing depth.
- qPCR offers a means to make reliable pre-sequencing estimates of non-human content in buccal/saliva samples.
- Quantification of contamination prior to sequencing offers opportunities to optimize multiplexing, increasing throughput and reducing the cost of sequencing per sample.

RESULTS I: Performance of Buccal/Saliva Samples from gDNA Isolation to Sequencing

HMW gDNA Isolation from Buccal/Saliva

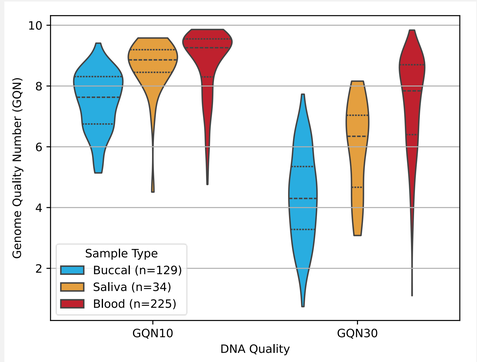

Figure 1. gDNA Quality Across Sample Types. Femto Pulse analysis revealed variable quality of High Molecular Weight (HMW) gDNA from oral specimen types. Acceptable quality for buccal and saliva was set at GQN10 ≥ 6.0 (Buccal: 91%, Saliva: 97%) and GQN30 ≥ 3.0 (Buccal: 78%, Saliva: 100%).

Table 1. Mean GQN By Sample Type (Mean ± SD)

Sample Type	GQN10	GQN30
Buccal	7.5 (± 1.0)	4.3 (± 1.5)
Saliva	8.6 (± 0.9)	6.0 (± 1.5)
Blood	9.4 (± 0.6)	7.4 (± 1.5)

LR-WGS Performance

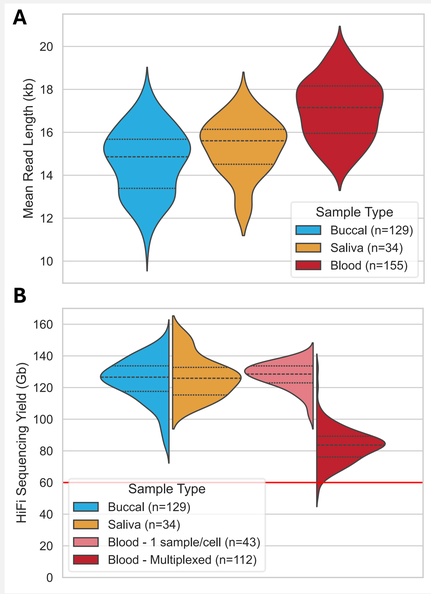

Figure 2. HiFi Sequencing Read Length and Yield. **A)** Mean read lengths across sample types. **B)** HiFi sequencing yield with SPRQ chemistry. Buccal and saliva samples were sequenced as 1 sample/cell, while blood samples were sequenced either as 1 sample/cell or 6 samples/4 cells.

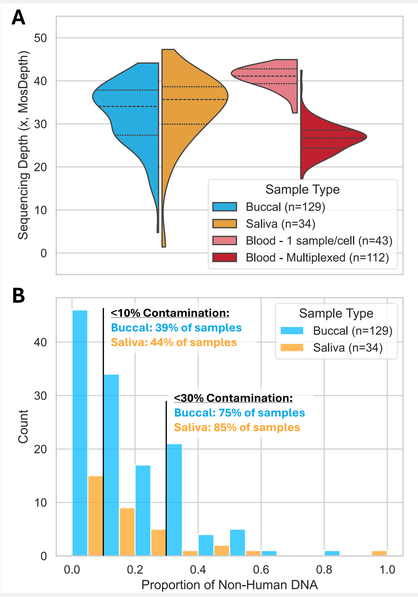

Figure 3. Post-alignment Sequencing Depth and Non-Human Content. A) Observed mean sequencing depth of the human genome aligned to the hg38 reference. **B)** Frequency distribution of non-human content observed in samples from oral collection methods.

Key Observations:

- **Read Length:** Sequencing read lengths of buccal (14.7 ± 1.5 kb), and saliva (15.3 ± 1.3 kb) were shorter than blood (17.1 ± 1.4 kb).
- **Yield:** Revio sequencing with the improved SPRQ chemistry achieved 125 ± 13 Gb HiFi yield per SMRT Cell. 6-plex multiplexing of blood achieved 83 ± 11 Gb and ≥20x sequencing depth.
- **Sequencing Depth:** Oral specimens showed increased variability in mean sequencing depth compared to blood, likely attributed to oral microbiome contamination.

RESULTS II: Estimating Non-Human DNA Contamination in Buccal/Saliva Samples

qPCR Quantification of Bacterial Content in Samples with Known Non-Human Contamination Levels

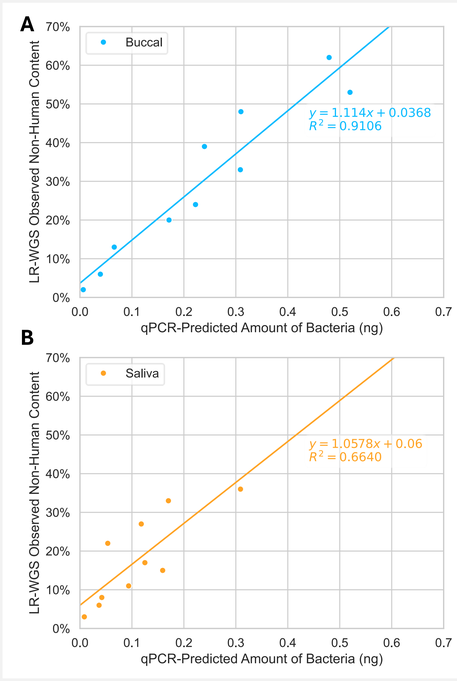

Figure 4. Correlation of qPCR-Predicted Bacterial Contamination with LR-WGS Observed Non-Human Content. Bacterial contamination of previously sequenced buccal (A) and saliva (B) samples was estimated using qPCR and compared to the non-human content observed in PacBio LR-WGS. qPCR estimate and LR-WGS quantification correlated strongly for buccal samples and moderately for saliva samples.

Estimation of Bacterial Content Prior to Sequencing Informs Optimal Sample Multiplexing

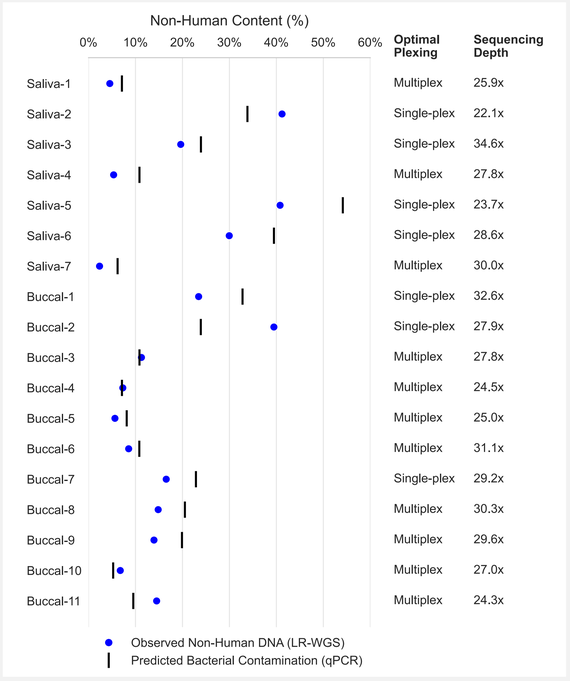

Figure 5. qPCR Quantification of Bacterial Content (Predicted vs. Observed). Bacterial contamination of 18 samples was predicted via qPCR, then was used to determine optimal sample plexing for sequencing. For 16 samples (89%), non-human content was confirmed by LR-WGS to be within ±10% of the qPCR-predicted value. 1 sample reported 14% lower contamination while 1 sample reported 16% higher contamination than their respective predictions. All samples achieved minimum sequencing depth of 20x.

Key Observations

- qPCR quantification of bacterial DNA correlates with the non-human content observed in PacBio LR-WGS.
- This method was able to reliably estimate non-human content within oral samples prior to sequencing.
- This strategy can be leveraged to inform optimal sample plexing for Revio sequencing.

Potential Limitations

- Input DNA mass in each reaction must be constant.
- Potential inter-assay variability may occur.
- Assay is based on *E.* coli standards.
- May not capture microbial diversity leading to underestimations.

REFERENCES

1. HiFi Prep 96 Procedure: https://www.pacb.com/wp-content/uploads/Procedure-checklist-Preparing-whole-genome-libraries-using-the-HiFi-prep-kit-96.pdf
2. Size selection using the LightBench: https://www.pacb.com/wp-content/uploads/Technical-note-Gel-cassette-size-selection-methods-for-WGS-HiFi-libraries.pdf
3. https://github.com/PacificBiosciences/HiFi-human-WGS-WDL
4. Zymo Femto Bacterial DNA Quantification Kit: https://www.zymoresearch.com/products/femto-bacterial-dna-quantification-kit