Mitochondrial DNA Sequencing Using PacBio SMRT Technology

[1]Shreyasee Chakraborty, [1]William Rowell, [1]Jenny Gu, [2]Victoria Portnoy, [2]Senny Nordmeier, [2]Young Kim, [1]Primo Baybayan, [1]Meredith Ashby
[1]Pacific Biosciences, 1305 O'Brien Drive, Menlo Park, CA 94025
[2]System Bioscience, 265 N Whisman Rd, Mountain View, CA 94043

Introduction

Mitochondrial DNA (mtDNA) is a compact, double-stranded circular genome of 16,569 bp with a cytosine-rich light (L) chain and a guanine-rich heavy (H) chain.

mtDNA mutations have been increasingly recognized as important contributors to an array of human diseases such as Parkinson's disease, Alzheimer's disease, colorectal cancer and Kearns–Sayre syndrome [1]. mtDNA mutations can affect all of the 1000-10,000 copies of the mitochondrial genome present in a cell **(homoplasmic mutation)** or only a subset of copies **(heteroplasmic mutation).** The ratio of normal to mutant mtDNAs within cells is a significant factor in whether mutations will result in disease, as well as the clinical presentation, penetrance, and severity of the phenotype**.** Over time, heteroplasmic mutations can become homoplastic due to differential replication and random assortment. Full characterization of the mitochondrial genome would involve detection of not only homoplastic but heteroplasmic mutations, as well as complete phasing.

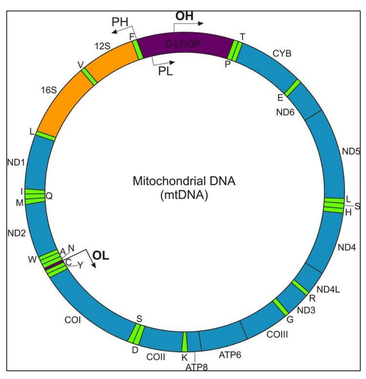


Figure 1. The mitochondrial genome

Previously, we sequenced human mtDNA on the PacBio RS II System with two partially overlapping amplicons [2]. Here, we present amplification-free, full-length sequencing of linearized mtDNA using the Sequel System. Full-length sequencing allows variant phasing along the entire mitochondrial genome, identification of heteroplasmic variants, and detection of epigenetic modifications that are lost in amplicon-based methods [3].

Isolation of Mitochondrial DNA

Isolation of mitochondria and extraction of mtDNA were carried out using System Biosciences mitoISO and mitoDNA kits, respectively. Isolated circular mtDNA was linearized using FastDigest restriction enzyme followed by library construction for SMRT Sequencing [3].

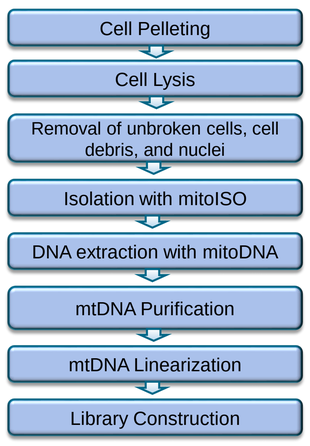


Figure 2. Workflow describing extraction of mitochondria and mtDNA using System Biosciences isolation and DNA extraction kits.

SMRTbell Library

The linearized mtDNA samples were size-selected using BluePippin (Sage Science Inc, Beverly, MA) to remove short fragments ≤13 kb. The recovered DNA (~150 ng) was used to construct a SMRTbell library using the standard library preparation workflow.

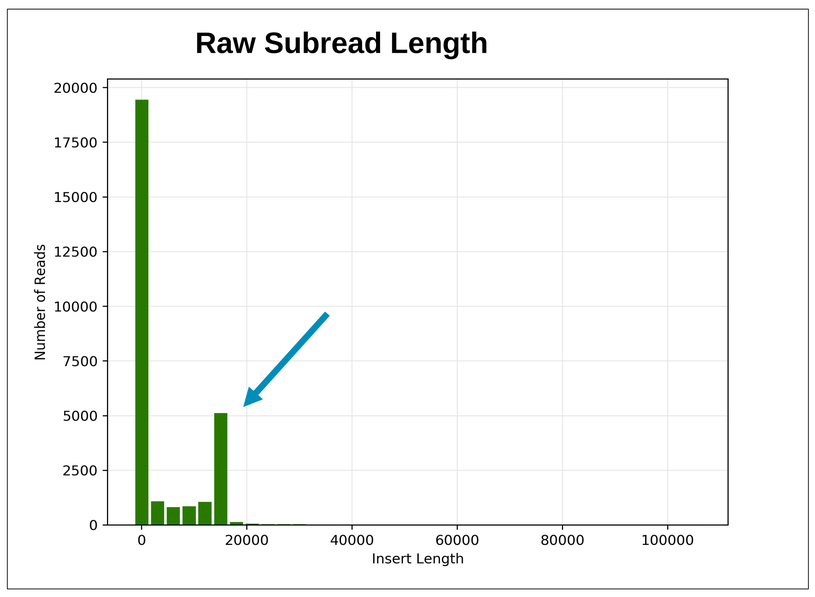
Figure 3. Workflow describing SMRTbell library preparation for linearized mtDNA.

SMRT Sequencing

Mitochondrial genomes from two different cell lines were sequenced on the PacBio Sequel System (10 hr movie, Sequel Sequencing Kit v2.1, Software v5.0). Initial results yielded reads spanning the complete mitochondrial genome. However, further library preparation development is needed to increase the fraction of intact, linearized templates and boost sequencing yield.

Sample	Mean Polymerase Read Length	Polymerase Read Length N50	Longest Subread Length Mean	Longest Subread Length N50
Sample 1	10,558	16,750	7,411	14,250
Sample 2	42,557	77,250	4,306	15,250

Table 1. Sequencing results showing long read lengths



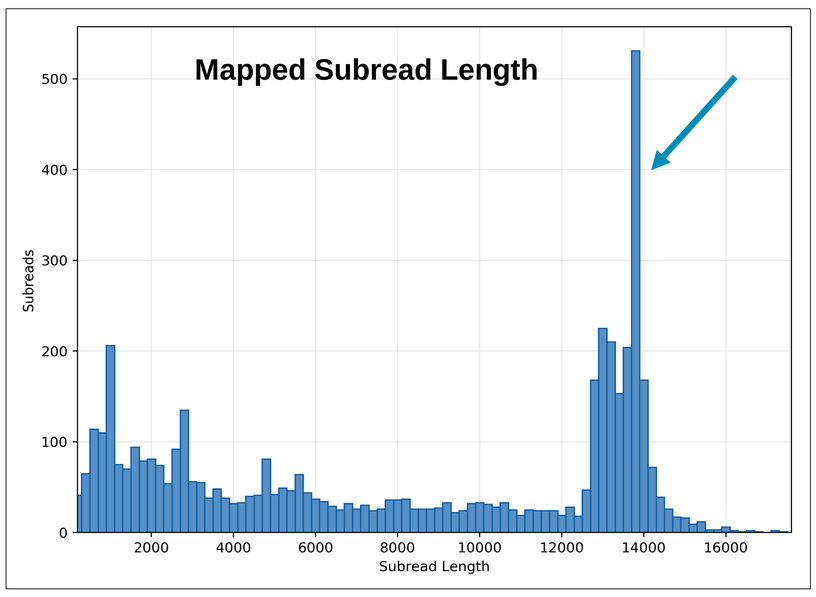


Figure 4. mtDNA Subread length and mapped subread length show distinct peaks at approximately 16 kb (arrows), indicating full length sequences of the mitochondrial genome.

Data Analysis

Subreads from primary analysis between 13 kb and 18 kb in length were analyzed with Long Amplicon Analysis (LAA2) to generate consensus sequences. LAA2 clusters similar sequences and applies the Arrow statistical model for SMRT Sequencing data iteratively to generate highly accurate (>QV50), fully phased, full-length mtDNA consensus sequences over 16 kb in length.

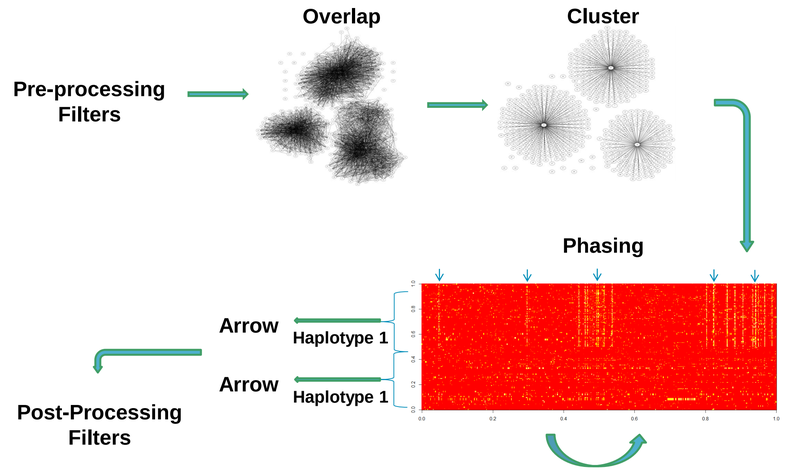


Figure 5. Overall workflow of LAA2 pipeline for PacBio data

Aligned reads in IGV

To simulate the detection of heteroplasmic mtDNA, we generated a 9:1 *in silico* mixture of 5000 reads from two distinct mtDNA variants. LAA2 was used on reads between 13 kb and 18 kb in length to cluster and generate phased consensus sequences. These consensus sequences, as well as the raw subreads used to generate them, were aligned to rCRS mtDNA reference sequence (NC_012920.1) with GraphMap [5] and displayed using IGV 2.4.

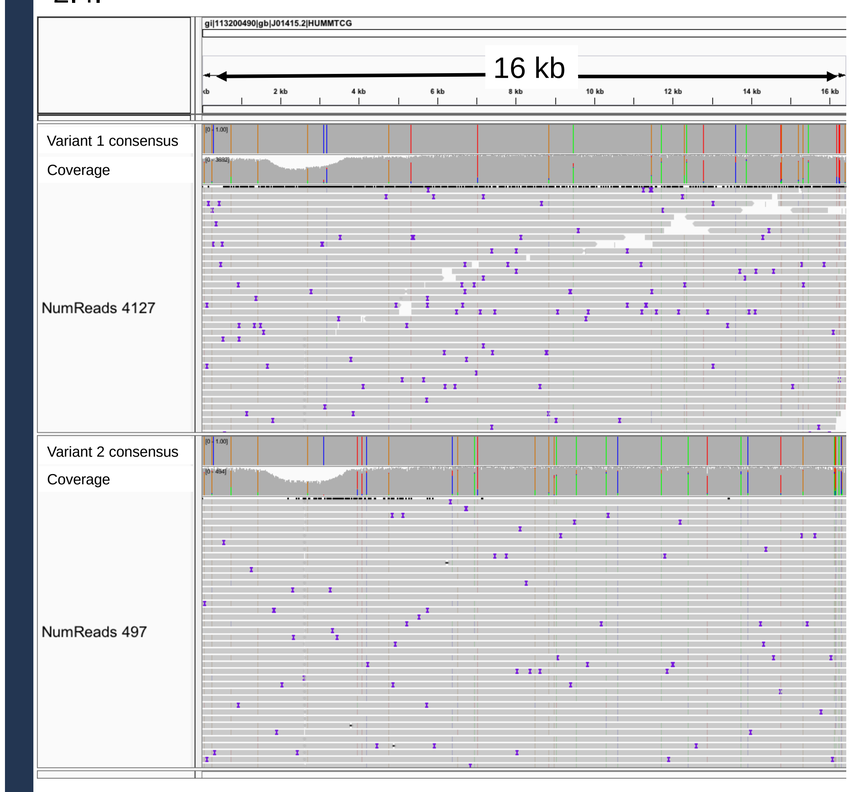


Figure 6. LAA2 clusters and phases an *in silico* mixture of two mitochondrial genome variants.

Conclusions

- Full-length mtDNA can be sequenced with PacBio long read technology in one pass, with no assembly.
- A mtDNA library for SMRT sequencing can be constructed with as little as 150 ng of input DNA.
- Unlike with amplicon-based short read sequencing, PacBio SMRT Sequencing permits phasing of all mutations along the entire mitochondrial genome.

Next Steps:

- Optimize DNA isolation and library preparation protocols to further enrich for full-length mtDNA sequences.
- Explore the limits of heteroplasmy detection with increased read depth and tuned LAA2 parameters.
- Analyze base modifications for epigenetic characterization.

References

[1] Smith, D. R. (2016). The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Briefings in Functional Genomics*, *15*(1), 47–54. [doi: 10.1093/bfgp/elv027]

[2] Vossen R.H.A.M., Buermans H.P.J. (2017) Full-Length Mitochondrial-DNA Sequencing on the PacBio RSII. In: White S., Cantsilieris S. (eds) *Genotyping. Methods in Molecular Biology*, vol 1492. Humana Press, New York, NY [doi: 10.1007/978-1-4939-6442-0_12]

[3] Fendt, L., Zimmermann, B., Daniaux, M., & Parson, W. (2009). Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC Genomics*, *10*, 139. [doi: 10.1186/1471-2164-10-139]

[4] Mühlenhoff, U. (2010) Mitochondria: Methods for Preparation. In: *eLS*. John Wiley & Sons Ltd, Chichester. http://www.els.net [doi: 10.1002/9780470015902.a0002601.pub2]

[5] Sovic, I. (2017). https://github.com/isovic/graphmap