A High-Quality PacBio Insect Genome from 5 ng of Input DNA

Brendan Galvin[1], Douglas Shoue[2], Lei Zhu[1], Christine Lambert[1], Primo Baybayan[1], Sarah Kingan[1], Michelle Vierra[1], Jonas Korlach[1], Mary Ann McDowell[2] & Stephen Richards[3]

[1]Pacific Biosciences, Menlo Park, CA; [2]University of Notre Dame, Notre Dame, IN; [3]University of California Davis, Davis, CA

Abstract

High-quality insect genomes are essential resources to understand insect biology and to combat them as disease vectors and agricultural pests. It is desirable to sequence a single individual for a reference genome to avoid complications from multiple alleles during *de novo* assembly. However, the small body size of many insects poses a challenge for the use of long-read sequencing technologies which often have high DNA-input requirements. The previously described PacBio Low DNA Input Protocol starts with ~100 ng of DNA and allows for high-quality assemblies of single mosquitoes among others and represents a significant step in reducing such requirements. Here, we describe a new library protocol with a further 20-fold reduction in the DNA input quantity. Starting with just 5 ng of high molecular weight DNA, we describe the successful sequencing and *de novo* genome assembly of a single male sandfly (*Phlebotomus papatasi*, the main vector of the Old World cutaneous leishmaniasis), using HiFi data generated on the PacBio Sequel II System and assembled with FALCON. The assembly shows a high degree of completeness (>97% of BUSCO genes are complete), contiguity (contig N50 of 1 Mb), and sequence accuracy (>98% of BUSCO genes without frameshift errors). This workflow has general utility for small-bodied insects and other plant and animal species for both focused research studies or in conjunction with large-scale genome projects.

Motivation & Workflow

Library Preparation for DNA-Limited Samples:

- DNA from single individuals much preferred for *de novo* assembly to avoid complications from >2 alleles
- Standard SMRT Sequencing protocols require > 5 µg HMW gDNA
- Low DNA Input Protocol[1] reduces requirement to ~100 ng
- For some samples (small organisms), even ~100 ng HMW DNA is not possible

Workflow:

DNA isolation & shearing
Starting with 5 ng gDNA
>10 kb average size
Megaruptor 3 shearing

↓

Amplification
Proprietary (PCR-based) SMRTbell gDNA Sample Amplification Kit

↓

Library preparation & size selection
HiFi SMRTbell Libraries using SMRTbell Express Template Prep Kit 2.0[2]

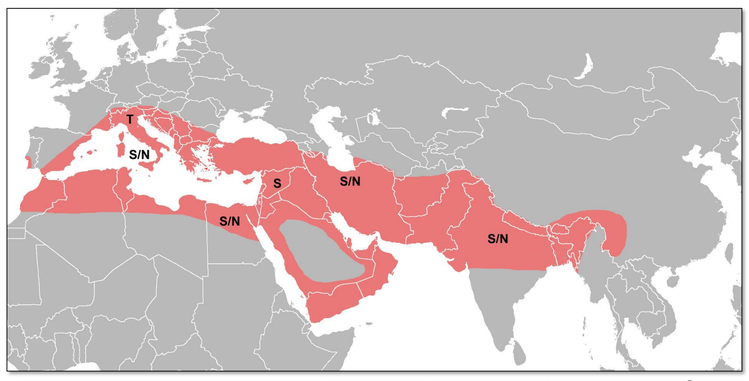
↓

PacBio HiFi sequencing
Sequel II System

Example - Sandfly

Phlebotomus papatasi:

- Main vector of Old World cutaneous leishmaniasis and Phlebotomus fever

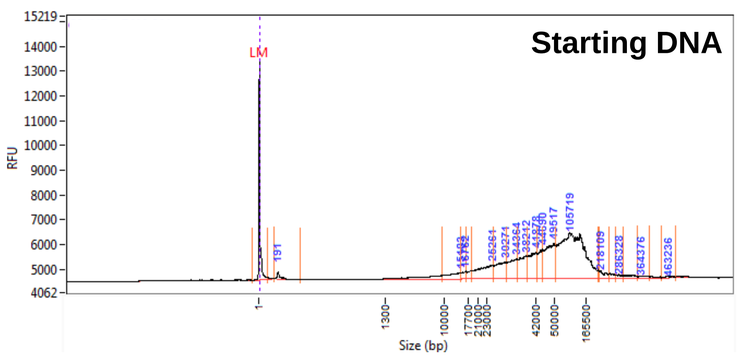


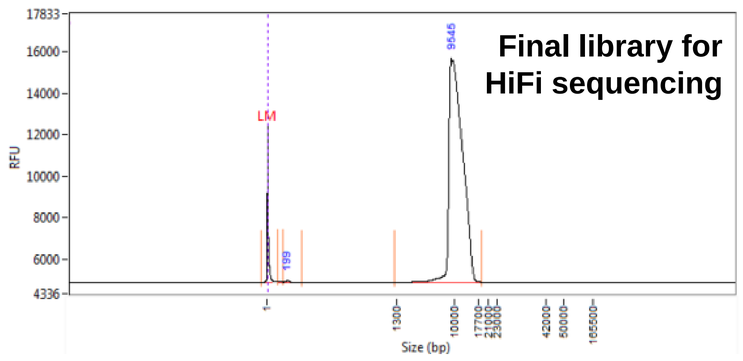
Phlebotomus fever distribution (Toscana, Sicilian, Naples serotypes)[3]

- Sufficient DNA for low DNA input protocol[1] could not be obtained from a single individual

Library preparation:

- DNA extracted from a single sandfly
- Made library **starting with 5 ng** using the ultra-low DNA input protocol, followed by BluePippin size selection (8-15 kb)





Sequencing Results:

- Run stats from one SMRT Cell 8M:

Total yield:	626 Gb
Number of reads:	6.4 million
Raw average RL:	97.6 kb
Mean subRL:	13.2 kb

>Q20 HiFi reads:	3,327,164
HiFi yield:	**35.9 Gb**
Mean HiFi RL:	10.8 kb
Mean HiFi read quality:	**Q36 (99.975%)**

Example - Sandfly

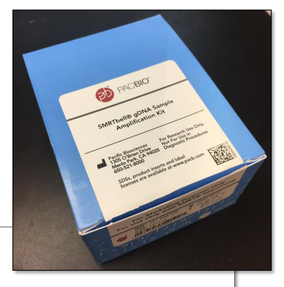
Assembly Results:

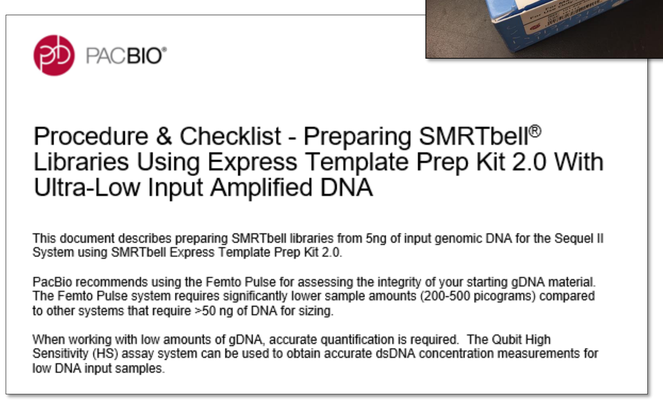
- Falcon assembler + Purge_dups

	This assembly	Previous assembly[4]	Change
Assembly size	**363 Mb**	364 Mb	
# of contigs	**977**	139,199	**142x less fragmented**
Contig N50	**1,012 kb**	5.8 kb	**174x more contiguous**
BUSCO (arthropoda)[5] analysis (n=1,066)	**C:97.2% [S: 95.9%, D: 1.3%] F: 0.7%, M: 2.1%**	C: 87.3% [S: 84.3%, D: 3.0%] F: 10.4%, M: 2.3%	**9.9% more complete genes**
Conserved gene frameshift error analysis (BUSCO gene set)	**14 frame shift inducing errors (808 kb alignment space)**	25 frame shift inducing errors (708 kb alignment space)	**11 less errors (in ~100 kb larger alignments space)**

Protocols & Reagents

- SMRTbell gDNA Sample Amplification Kit & detailed workflow protocol are available





PACBIO®

Procedure & Checklist - Preparing SMRTbell® Libraries Using Express Template Prep Kit 2.0 With Ultra-Low Input Amplified DNA

This document describes preparing SMRTbell libraries from 5ng of input genomic DNA for the Sequel II System using SMRTbell Express Template Prep Kit 2.0.

PacBio recommends using the Femto Pulse for assessing the integrity of your starting gDNA material. The Femto Pulse system requires significantly lower sample amounts (200-500 picograms) compared to other systems that require >50 ng of DNA for sizing.

When working with low amounts of gDNA, accurate quantification is required. The Qubit High Sensitivity (HS) assay system can be used to obtain accurate dsDNA concentration measurements for low DNA input samples.

Conclusions

- Demonstrating efficient HiFi read-based PacBio sequencing from as low as 5 ng of input DNA
- High-quality *de novo* assembly of single sandfly individual
- Kits and protocols available

References & Acknowledgments

[1]Kingan et al. (2019) *Genes* 10: 62
[2]https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-HiFi-SMRTbell-Libraries-using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf
[3]https://commons.wikimedia.org/w/index.php?curid=5576847
[4]https://www.ebi.ac.uk/ena/data/view/GCA_000262795.1
[5]https://busco.ezlab.org/

The authors would like to thank everyone who helped generate data for the poster.