Primo Baybayan[1], Mara K. Lawniczak[2], Katharina von Wyschetzki[2], Karen Oliver[2], Craig Corton[2], Shane McCarthy[2], Matt Berriman[2], Mark Blaxter[2], Michelle Smith[2], Mathilde Gendrin[3], Jeremy Herren[4], Sandrine Nsango[5], Sarah B. Kingan[1], Christine C. Lambert[1], Shreyasee Chakraborty[1], Harsharan Dhillon[1], Aaron Wenger[1] and Jonas Korlach[1]

New Advances in SMRT Sequencing Facilitate Multiplexing for *de novo* and Structural Variant Studies

Primo Baybayan[1], Mara K. Lawniczak[2], Katharina von Wyschetzki[2], Karen Oliver[2], Craig Corton[2], Shane McCarthy[2], Matt Berriman[2], Mark Blaxter[2], Michelle Smith[2], Mathilde Gendrin[3], Jeremy Herren[4], Sandrine Nsango[5], Sarah B. Kingan[1], Christine C. Lambert[1], Shreyasee Chakraborty[1], Harsharan Dhillon[1], Aaron Wenger[1] and Jonas Korlach[1]

1. Pacific Biosciences, Menlo Park, CA 2. Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK 3. Institute Pasteur, French Guiana 4. International Centre of Insect Physiology and Ecology, Kenya 5. Faculty of Medicine and Pharmaceutical Sciences, University of Douala, Cameroon

Abstract

The latest advancements in SMRT Sequencing with the Sequel II System Chemistry 2.0 have increased average read lengths up to 50% compared to Sequel II Chemistry 1.0, thus allowing for multiplexing of:

- 2-3 small organisms (<600 Mb) such as insects and worms for producing reference quality assemblies
- Calling structural variants for up to 2 samples with ~3 Gb genomes
- Analysis of 48 microbial genomes
- Characterizing up to 4 microbial communities for metagenomic profiling in a single SMRT Cell 8M

With the improved processivity of the new Sequel II sequencing polymerase, more SMRTbell molecules reach rolling circle replication mode, resulting in longer overall read lengths and thus allowing efficient detection of barcodes (up to 80%) in the SMRTbell templates. Furthermore, multiplexing of genomes larger than microbial organisms is now achievable. In collaboration with the Wellcome Sanger Institute, we have developed a workflow for multiplexing two individual *Anopheles coluzzii* using as low as 150 ng of genomic DNA per individual. The resulting assemblies showed high contiguity (Contig N50s >3 Mb) and completeness (>98% of conserved genes). For microbial multiplexing, we multiplexed 48 microbes with varying complexities and genome sizes ranging from 1.6 to 8.0 Mb in a single SMRT Cell 8M. Using a new end-to-end Microbial Assembly Analysis application in SMRT Link 8.0, the assemblies resulted in complete circularized genomes (>200-fold coverage). Finally, the long-read lengths (>90 kb) allow detection of barcodes in large insert SMRTbell templates (>15 kb), thus facilitating multiplexing of two human samples in 1 SMRT Cell 8M for detecting SVs, large Indels and CNVs.

Here, we present new/updated workflows and describe the results for multiplexing samples for specific applications for SMRT Sequencing using the Sequel II System.

Multiplexing Methods

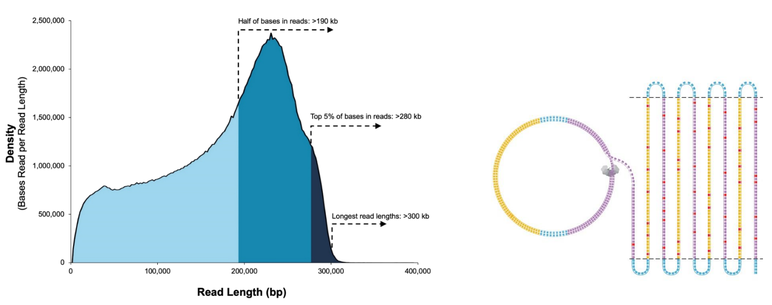


Fig 1. Sequel II Chemistry 2.0 allows more SMRTbell molecules to reach rolling circle replication mode, resulting in longer read lengths, thus improving detection of barcodes in longer insert SMRTbell templates.

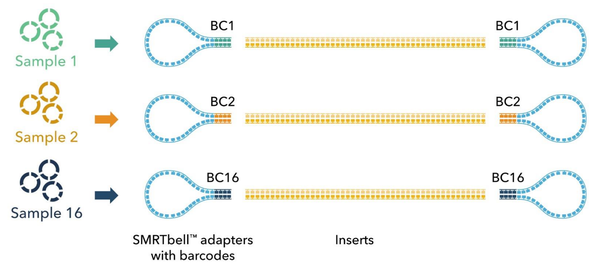
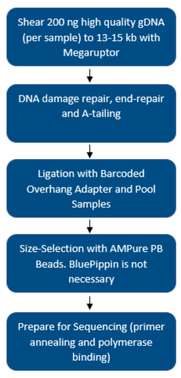


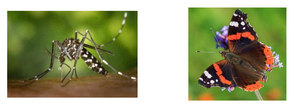
Fig 2. Samples are barcoded during ligation with barcoded overhang adapters containing 16 bp PacBio specific sequences.

Application	Plex	Available Barcoded Overhang Adapters	Additional Info
Small genomes (<600 Mb)	2-3	Barcoded Overhang Adapter Kit 8A (101-628-400) or 8B (101-628-500)	Use set A or B (16 total barcodes)
Structural Variation (Human)	2	Barcoded Overhang Adapter Kit 8A (101-628-400) or 8B (101-628-500)	Use set A or B (16 total barcodes)
Shotgun Metagenomics	4	Barcoded Overhang Adapter Kit 8A (101-628-400) or 8B (101-628-500)	Use set A or B (16 total barcodes)
Microbial Multiplexing	32-48	Oligo Synthesis Provider	See procedure for sequences

Table 1. Samples may be barcoded using PacBio's Barcoded Overhang Adapters. For >16 samples, additional barcoded overhang adapters may be ordered from your Oligo Synthesis Provider.

Low DNA Input Multiplexing of Arthropods for *De novo* Genome Assembly





- Multiplex 2-3 insect genomes (<600 Mb)
- Low DNA Input Multiplexing is supported on the Sequel II System only (30-hr movies, 2-hr pre-extension)

Fig 3. Library prep workflow for multiplexing small genomes.

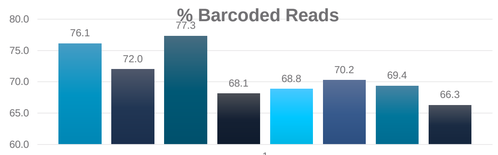


Fig 4. Barcode detection rate is ~70% yield of barcoded reads in a 2-plex pool.

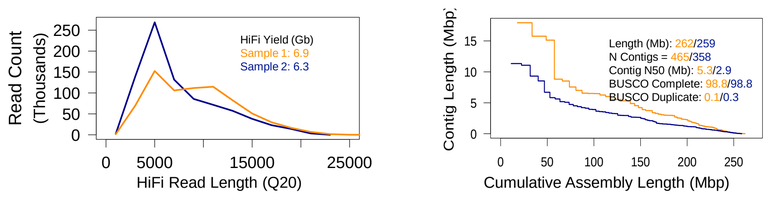
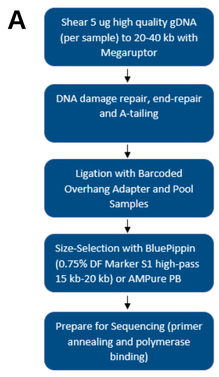


Fig 5. High contiguity (FALCON + Purge-Dups) and genome completeness (BUSCO) for multiplexed insect *de novo* genome assemblies.

Multiplexing Two Human Samples in One SMRT Cell 8M for SV Detection



- Multiplex 2 human samples
- SV Multiplexing is supported on the Sequel II System only (15-hr movies, 4-hr pre-extension)

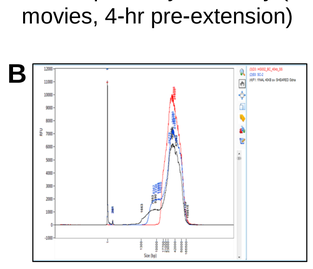


Fig 6. Library prep workflow for multiplexing two human samples **(A)**. Libraries may be size-selected using BluePippin (gold standard) or AMPure PB Beads. **(B)** Femto Pulse sizing QC analysis of samples size selected with BluePippin and AMPure PB Beads.

A

Size-Selection Method	Total Bases (Gb)	Unique Molecular Yield (Gb)	Polymerase RL (bp)	Polymerase N50 (bp)	Longest Subread (bp)	Longest Subread N50 (bp)	P0 %	P1 %	P2 %
BluePippin	220.0	89.1	53630	99322	21912	28946	46.32	51.19	2.49
AMPure	260.3	90.8	52249	95486	18523	26466	32.79	62.16	5.06

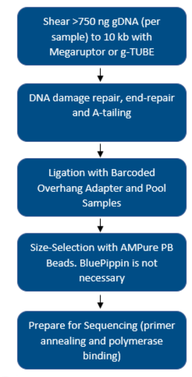
B

Sample	Total Unique Molecular Coverage, 2 samples	Fold UMC/ Barcode	% Barcoded yield
BluePippin	21.53	10.7X	64.0
AMPure	21.76	10.9X	66.0

Table 2. Sequencing **(A)** and coverage **(B)** performance of BluePippin and AMPure PB size-selected 2-plex human libraries. % Barcoded reads for a 30 kb pooled library is ~65%, generating ~10-fold unique molecular coverage (UMC) per sample. Recommended UMC coverage for population genetics is 5- to 10-fold.

Fig 7. Longer inserts maintain barcoded read count **(A)** and produce more unique molecular yield **(B)**. For SV detection using CLR (continuous long read) sequencing mode, we recommend shearing genomic DNA to approximately 30 kb and size-selected with lower cutoff of 15 kb.

Shotgun Metagenomics for Species Detection, Gene Profiling and Assembly



- Multiplex up to 4 samples depending on project goals
- Metagenomics Shotgun Multiplexing is supported on the Sequel II System only (30-hr movies, 2-hr pre-extension)

Fig 9. Shotgun metagenomics library prep workflow

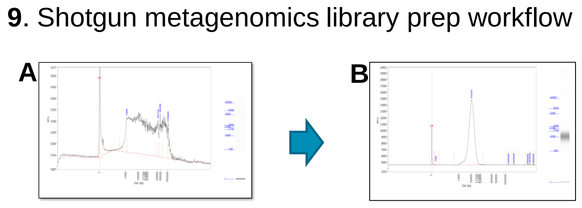


Fig 10. Metagenomics samples **(A)** are often degraded but can still be sheared to 10 kb **(B)** for shotgun sequencing.

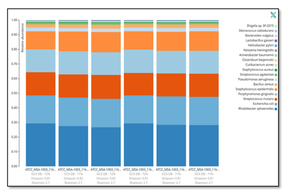
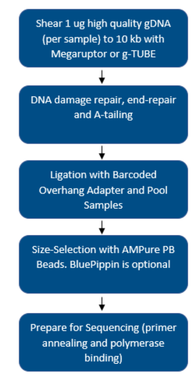


Fig 11. Relative abundances is the same across MSA-1003 samples tagged with 6 different barcodes. Species down to 0.018% abundance were detected successfully in a 6-plex pool.

Multiplex Up to 48 Microbes on One SMRT Cell 8M



- Multiplex up to 48 microbial genomes
- Multiplexing (32-48) in Sequel II System only (15-hr movies, 2-hr pre-extension)

Fig 12. Library prep workflow for microbial multiplexing

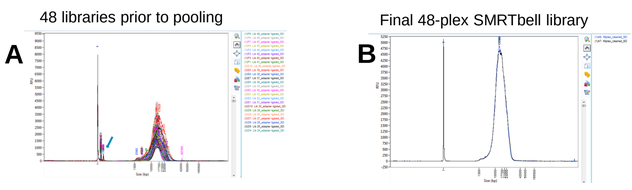


Fig 13. 48 microbial samples sheared to 10 kb **(A)** with the Megaruptor and pooled **(B)** for sequencing on Sequel II System

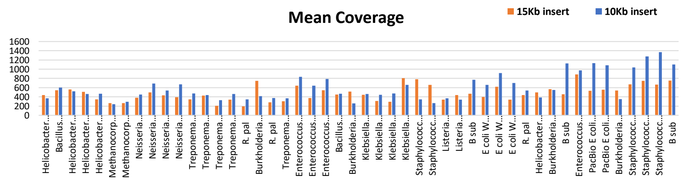


Fig 14. A 10 kb insert library performs similarly well to a 15 kb insert library with mean (total base) coverage of >200-fold.

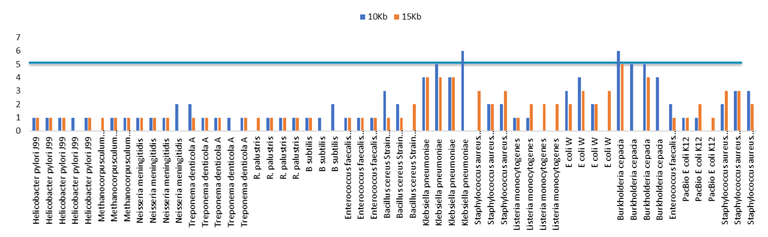


Fig 15. Genomes assembled to less than 5 contigs using PacBio's Microbial Assembly pipeline in SMRT Link. This tool is a push-button *de novo* assembly analysis application for small genomes between 2.0-10 Mb, which includes circularization and assembly of plasmids between 2-220 kb.

Conclusions

- Sequel II Chemistry 2.0 generates longer overall read lengths enabling multiplexing of larger genomes
- 2- to 3-Plex for low DNA input *de novo* assembly of arthropods
- 2-Plex for human structural variant detection
- 4-plex for metagenomics shotgun sequencing
- 32- to 48-Plex for microbial multiplexing

Acknowledgements

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