Application-Specific Barcoding Strategies for SMRT Sequencing

Primo Baybayan[1], Christine Lambert[1], Jill Muehling[1], Ting Hon[1], Walter Lee[1], John Harting[1], Swati Ranade[1], Steve Kujawa[1], Anand Sethuraman[1], Paul Kotturi[1]
PacBio, 1380 Willow Road, Menlo Park, CA 94025

Introduction

The number of useable bases per SMRT Cell can exceed 1 Gb with the latest P6-C4 chemistry and 6-hour movies. For applications such as microbial sequencing, targeted sequencing, Iso-Seq full-length isoform sequencing and Roche NimbleGen's target enrichment method, current PacBio RS II SMRT Cell yields could be in excess relative to project requirements.

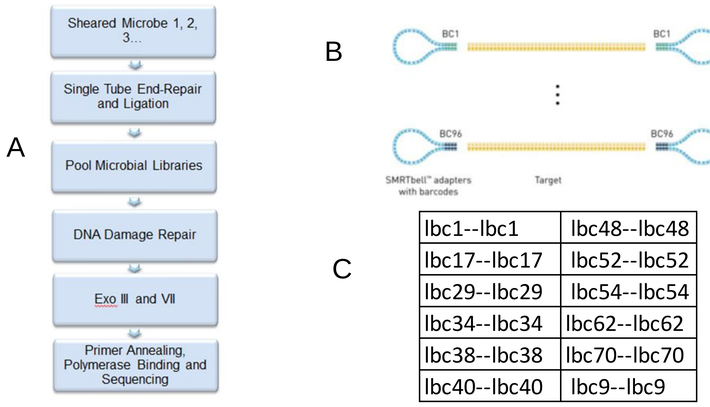
To this end, barcoding is a viable option for multiplexing samples. For microbial sequencing, multiplexing can be accomplished by tagging sheared genomic DNA during library construction with SMRTbell barcoded adapters. Results from 2-, 4-, 6- and 8-plex will be presented.

For HLA typing, full-length HLA genes as large as 5 kb may be barcoded during amplification or during SMRTbell library construction. The preferred barcoding strategy depends on the user's existing workflow and flexibility for changing and/or updating existing workflows. Five Class I/II genes (3.3 – 5.8 kb) x 96 patients can be multiplexed and typed in up to 4 SMRT Cells.

For Iso-Seq analysis, barcodes are incorporated during first-strand synthesis using oligo-dT tailed with 16-bp barcodes. Six maize RNA samples were barcoded, pooled and constructed into a single SMRTbell library.

Finally for target enrichment using NimbleGen's SeqCap EZ Target Enrichment method, linear barcodes are incorporated before the capture step. Results from a 12-plex target enrichment experiment will be presented.

Multiplexing for Microbial Assembly



C

lbc1--lbc1	lbc48--lbc48
lbc17--lbc17	lbc52--lbc52
lbc29--lbc29	lbc54--lbc54
lbc34--lbc34	lbc62--lbc62
lbc38--lbc38	lbc70--lbc70
lbc40--lbc40	lbc9--lbc9

Figure 1:

A. Workflow for multiplexing microbial genomes

B. Barcodes are added to the SMRTbell adapters.

C. Recommended barcoded adapters available from PacBio.

MICROBIAL WHOLE GENOME ASSEMBLY OF MULTIPLEXED *H. PYLORI* STRAINS

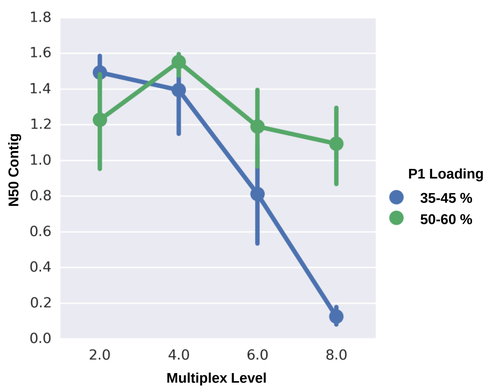


Figure 2: N50 contig vs Multiplex level of sequenced *H. pylori* strains using 10 kb library preparations, P6-C4 chemistry, and sequenced on the PacBio RS II. Increasing P1 loading helped achieve higher multiplex level to 6, possibly 8.

Targeted Sequencing Multiplexing

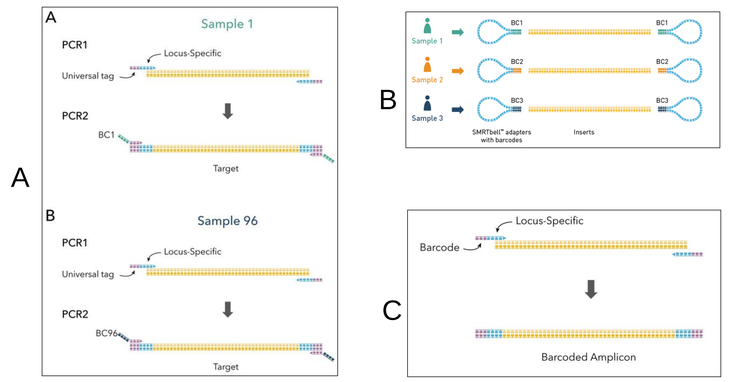


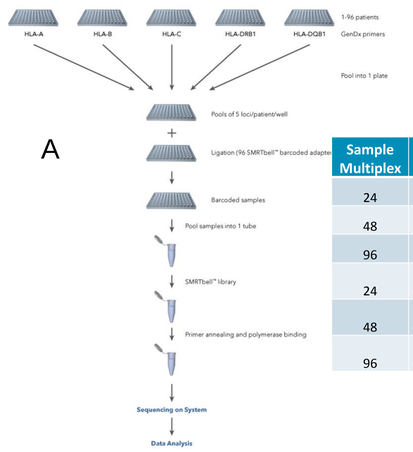
Figure 3: Barcoding options for targeted sequencing. www.pacb.com/wp-content/uploads/2015/09/ProductNote-Barcoded-Adapters-Barcoded-Universal-Primers.pdf

A. Barcoded Universal Primers: Barcode can be incorporated into the amplicon via a two-step tailed primer approach. Barcodes are commercially available from PacBio.

B. Barcoded Adapters: Barcodes are incorporated during ligation with barcoded adapters. Barcodes are commercially available from PacBio.

C. Locus-specific primers tailed with barcodes. Primers may be ordered from any oligo synthesis providers. The first 96 barcodes out of 384 available sequences are available: www.pacb.com/wp-content/uploads/PacBio-PCR-Primer-Barcodes- 0001-to-0096-IDT-Template.xlsx.

MULTIPLEXING FOR HLA SEQUENCING



B

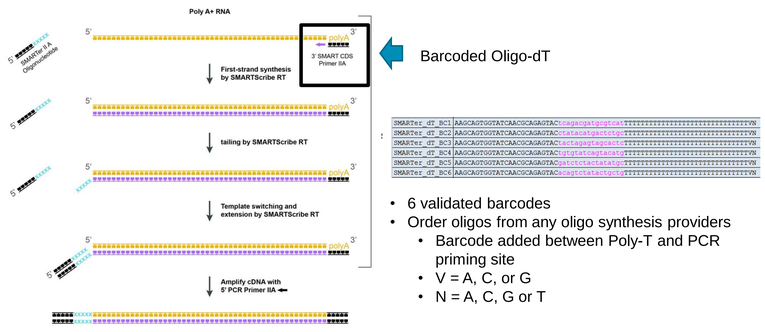
Sample Multiplex	HLA Loci	Amplicon Size	Targeted Region	# SMRT Cells
24	Class I (A/B/C)	~3.3 kb	Full Length	1
48	Class I (A/B/C)	~3.3 kb	Full Length	1
96	Class I (A/B/C)	~3.3 kb	Full Length	2
24	Class I & II (A/B/C/DR/DQ)	~3.3 kb - 5.8 kb	Class I: Full Length Class II: Partial	1
48	Class I & II (A/B/C/DR/DQ)	~3.3 kb - 5.8 kb	Class I: Full Length Class II: Partial	2
96	Class I & II (A/B/C/DR/DQ)	~3.3 kb - 5.8 kb	Class I: Full Length Class II: Partial	4

Figure 4:

A. GenDx HLA-A, HLA-B, HLA-C, HLA-DRB1 and HLA-DQB1 are amplified and pooled for barcoding. All 96 wells are subsequently pooled into a single tube for SMRTbell library construction.

B. Recommendations for multiplexing NGS-go® amplicons. www.pacb.com/applications/targeted-sequencing/hla/

Iso-Seq Application Multiplexing



- 6 validated barcodes
- Order oligos from any oligo synthesis providers
  - Barcode added between Poly-T and PCR priming site
  - V = A, C, or G
  - N = A, C, G or T

Figure 5: Barcodes are incorporated into the cDNA during first-strand synthesis. Double-stranded cDNAs are pooled during SMRTbell library construction.

Isoform Sequencing Multiplexing – cont.

GENOME ANNOTATION OF MULTIPLEXED MAIZE TISSUES

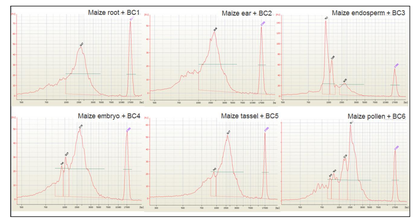
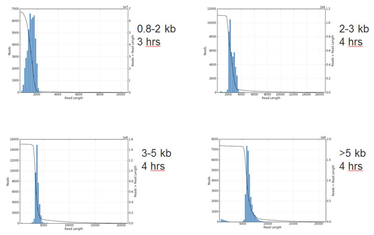


Figure 6: Bioanalyzer traces of barcoded cDNA from 6 Maize tissues prior to pooling.

A



B

Size Bin	Primer	# FL reads	Primer %
1to2k	1	21585	17
	2	16032	13
	3	20104	16
	4	22823	18
	5	18545	15
	6	28287	22
2to3k	1	20823	15
	2	21250	15
	3	14048	10
	4	29355	21
	5	22092	16
	6	11371	23
3to5k	1	20004	14
	2	24862	17
	3	12717	9
	4	35822	25
	5	24384	17
	6	27940	19
5to10k	1	12249	11
	2	17630	16
	3	11372	10
	4	37528	34
	5	16242	15
	6	16454	15

Figure 7:

A. Full-length read distribution from pooled maize library.

B. Full-length yield summary per fraction per sample.

Target Enrichment Multiplexing

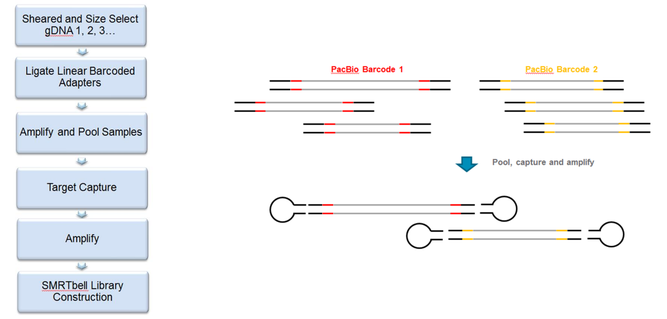


Figure 8: Target enrichment multiplex workflow. Sheared DNA samples are tagged with linear barcoded adapters. Workflow in development.
For information: www.pacb.com/products-and-services/consumables/target-enrichment/

MULTIPLEXING WITH NIMBLEGEN NEUROLOGY PANEL

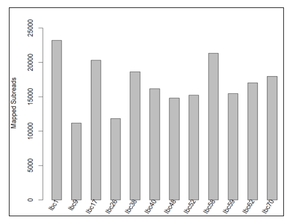


Figure 9: Multiplex of 12 samples. Number of mapped subreads of 5 kb target enrichment using Nimblegen's neurology panel.

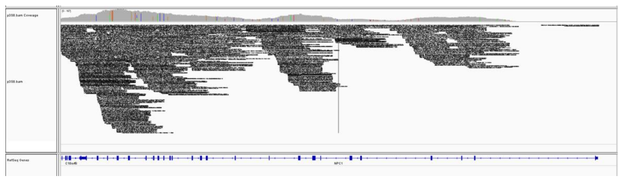


Figure 10: A 60 kb fragment showing intron-exon regions of the NPC1 gene.

Conclusions

- Variety of barcoding methods available for multiplexing samples for microbial whole genome sequencing, HLA typing, isoform sequencing, and targeted sequencing applications with PacBio Systems.
- Barcoding increases efficiency and sample throughput thus reducing the cost of library preparation and sequencing.

The authors would like to acknowledge and thank everyone Involved with the data generation for this poster.