Multiplexing Strategies for Microbial Whole Genome SMRT Sequencing

Christine Lambert[1], John Harting[1], Walter Lee[1], and Primo Baybayan[1],
PacBio, 1380 Willow Road, Menlo Park, CA 94025

Introduction

Increasing throughput of PacBio's single molecule sequencing systems enables multiple microbes to be sequenced on a single SMRT Cell. This multiplexing is readily achieved by incorporating a unique barcode for each microbe into the SMRTbell adapters and using a streamlined library preparation process. Incorporating a barcode without PCR amplification prevents the loss of epigenetic information and the generation of chimeric sequences, while the modified protocol eliminates the need to generate several individual SMRTbell libraries.

We multiplexed the genomes of up to 8 unique strains of *H. pylori*. Each genome was sheared and processed through adapter ligation in a single, addition only reaction. The barcoded samples were then pooled in equimolar quantities and a single library was prepared from the pooled samples.

We demonstrate successful de novo microbial assembly from all multiplexes tested (2- through 8-plex) using data generated from a single SMRTbell library, run on a single SMRT Cell and analyzed with standard tools within SMRT Analysis. This strategy was successful using both small (1.6 Mb, *H. pylori*) and medium (5 Mb, *B. subtilis* and *E. coli*) genomes. This protocol facilitates the sequencing of multiple microbial genomes in a single run, greatly increasing throughput and reducing costs per genome.

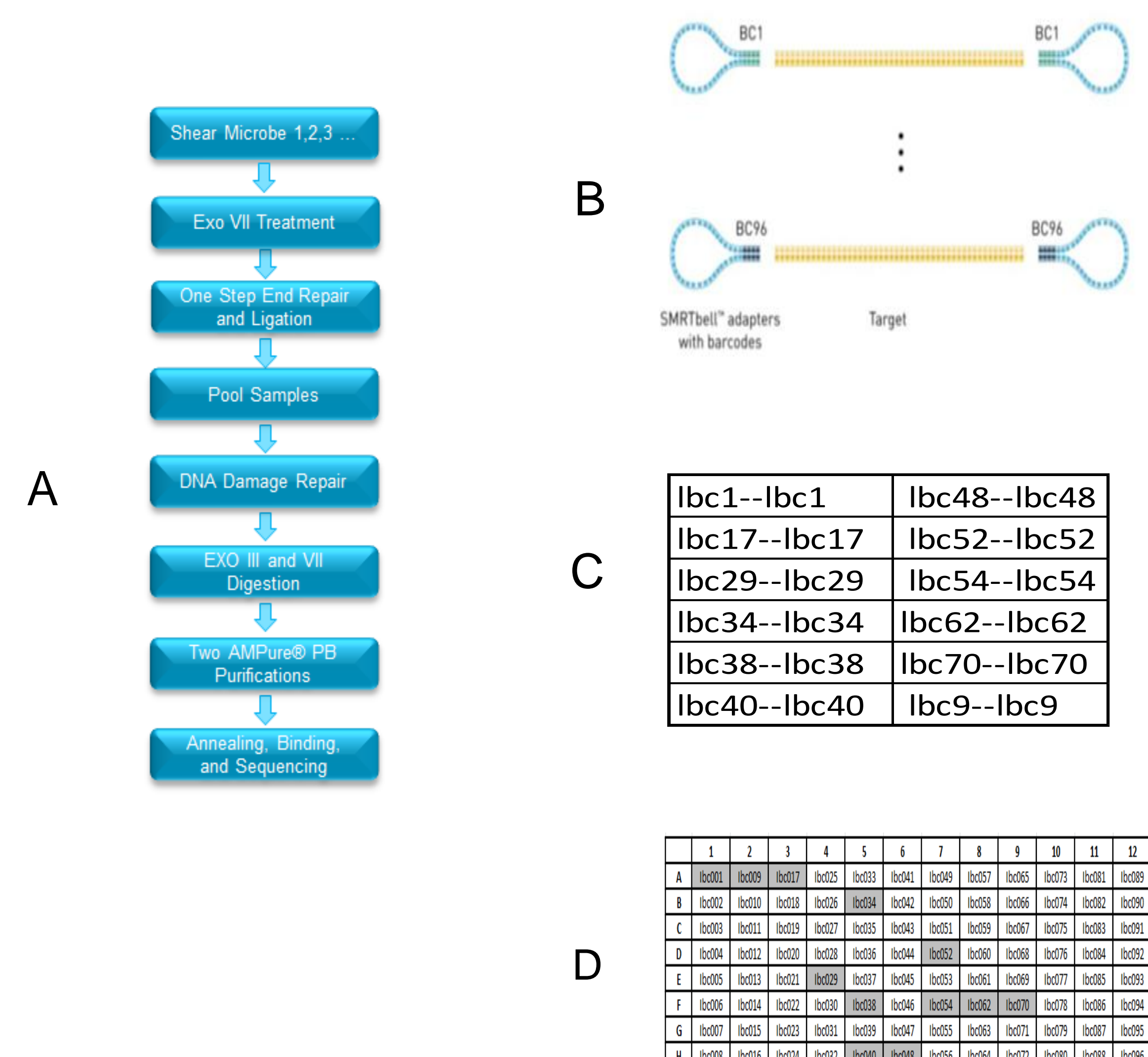## Workflow for Multiplexing Microbial Genomes

Figure 1: **SMRTbell Library Preparation Workflow**

A. Schematic workflow for multiplexing microbial genomes, achieved through the incorporation of a unique barcode for each microbe into the SMRTbell adapters after shearing genomic DNA.
B. Barcodes are added to the SMRTbell adapters.
C. Recommended best barcoded adapters for microbial multiplexing.
D. Barcode Plate Map of the commercially available barcodes

http://www.pacb.com/wp-content/uploads/2015/09/User-Bulletin-Barcode-Plate-Mapping.pdf

http://www.pacb.com/wp-content/uploads/2015/09/Package-Insert-Barcoded-Adapters-Plate-96.pdf

Materials and Methods

SMRTBELL LIBRARY PREPARATION

For this study *H. pylori* genomes (1.6 Mb) were used:

- The procedure used in this study is derived from the barcoded adapter procedure: *Preparing SMRTbell Libraries using PacBio Barcoded Adapters for Multiplex SMRT Sequencing.*
- A total of 1.5 µg gDNA of the pooled sheared DNA was used as input into library preparation.
- Eight unique strains of *H. pylori* were sheared to 10 kb using Covaris g-TUBES, by centrifugation at 8000 rpm at 50 ng/µl.
- After shearing, the *H.pylrori* strains were quantified and independently underwent a single tube End-Repair and Ligation reaction.
- Following barcoded adapter ligation, the *H. pylori* strains were pooled (2-8 plex) in equimolar quantities.
- The barcoded, pooled SMRTbell library preparation was completed using a modified protocol, including AMPure purification, DNA Damage repair, and digestion with Exo III & Exo VII.

PART NUMBER	PRODUCT NAME
100-538-700-02	Procedure & Checklist – Preparing SMRTbell Libraries using PacBio Barcoded Adapters for Multiplex SMRT Sequencing
100-466-000	SMRTbell Barcoded Adapter Plate - 96
100-259-100	SMRTbell Template Kit 1.0

Table 1: **Procedure and reagents required for SMRTbell library preparation using barcoded adapters for multiplexing**

For information: http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-Preparing-SMRTbell-Libraries-using-PacBio-Barcoded-Adapters-for-Multiplex-SMRT-Sequencing.pdf

SINGLE MOLECULE, REAL-TIME SEQUENCING

- SMRT Sequencing performed on the PacBio RS II using P6-C4 chemistry
- MagBead loading and One-Cell-Per-Well
- Collected 360 minute movies to maximize barcode yield
- Three SMRT Cells run per multiplex level (2-, 4-, 6-, 8-plex)
- Loaded 20-30 pM to achieve P1 = 33-42% and 40-50 pM to achieve P1 = 50-60%.
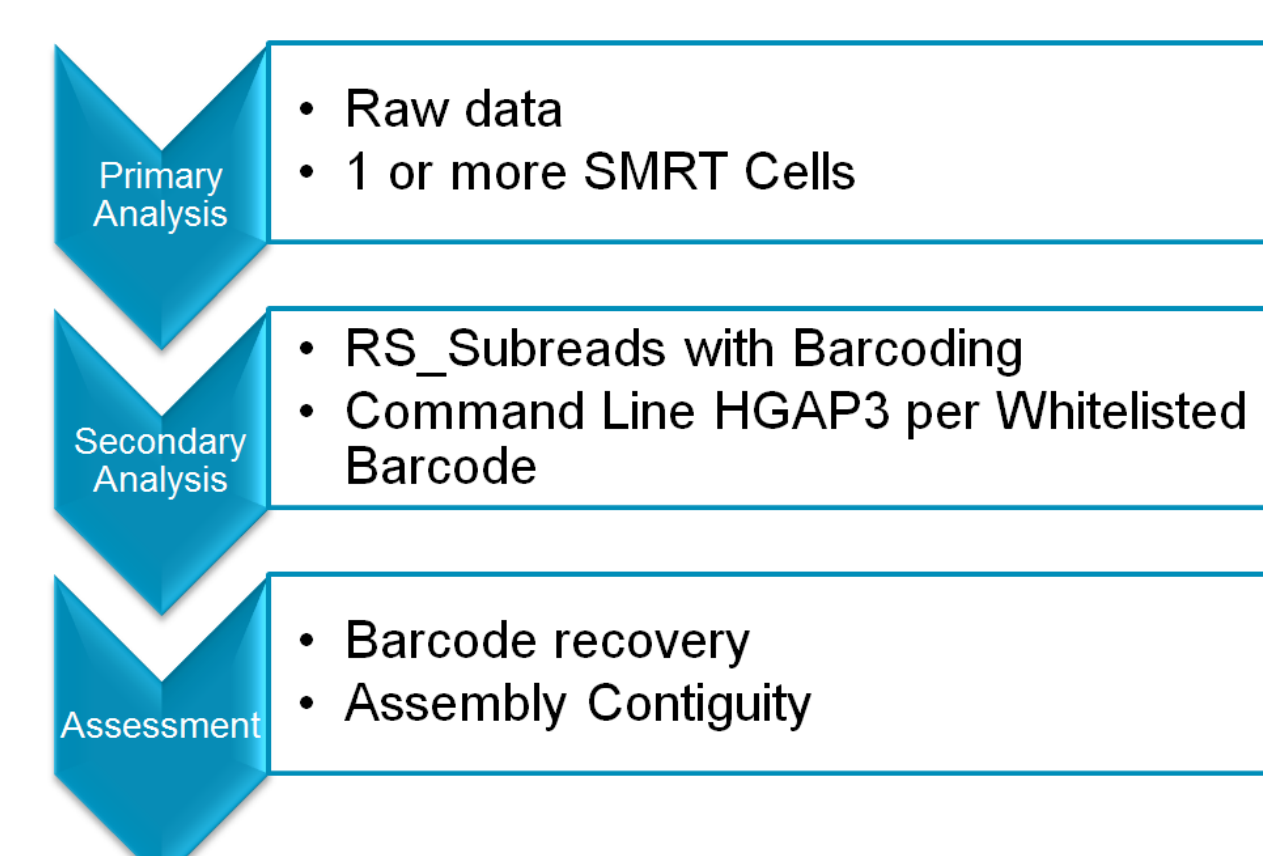
DATA ANALYSIS WORKFLOW

Figure 2: **Analysis workflow for microbial multiplexing**

1. For each SMRT Cell, label each ZMW with a barcode using "pbbarcode labelZmws." Score and align barcodes against all potential barcodes and return the highest scoring for each ZMW.

2. pbbarcode emitFastqs + awk. The "pbbarcode emitFastqs" generates a fastq file of subreads for each barcode. "awk" extracts the subread names from each barcode.fastq file and generate text file as the 'whitelist' per barcode.

3. For each barcode whitelist, run HGAP after including the whitelist as a filter in the P_Filter[Filtering] module.

https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP-Whitelisting-Tutorial

Results

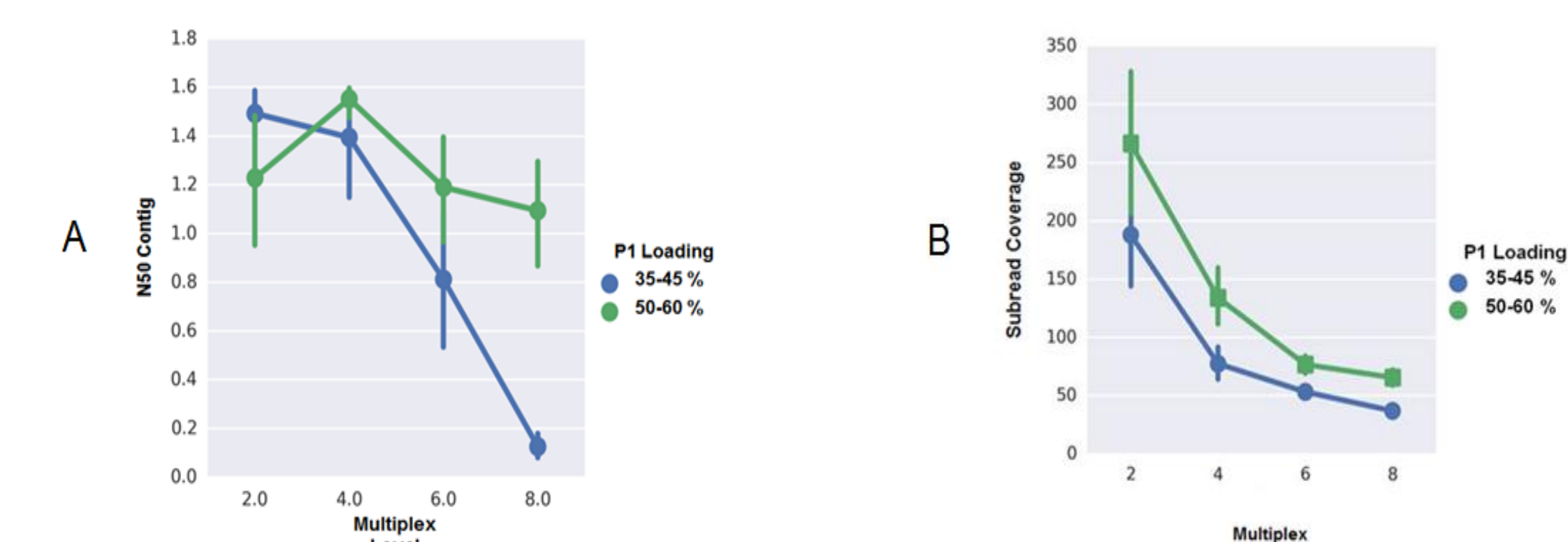### MICROBIAL WHOLE GENOME ASSEMBLY OF MULTIPLEXED *H. PYLORI* STRAINS

Figure 3: **Assembly performance of multiplexed *H. pylori* strains**

A. N50 contig length vs. multiplex level of *H. pylori* strains using 10 kb library libraries. Increased P1 loading helped achieve higher optimal multiplex level to 6, possibly 8.
B. Increased loading generates more subreads resulting in higher coverage for each *H. pylori* strain per multiplex level.

EXO VII TREATMENT IMPROVES LIBRARY QUALITY

Figure 4: **Proposed Multiplex Library Workflow.** Treatment of the sheared genomic DNA with Exo VII before the end-repair and ligation improves library quality.

SAMPLES	TOTAL SUBREADS	PALINDROMIC SUBREADS	% PALINDROMES	# OF CONTIGS
B.sub (-) Exo VII	205684	23689	11.5	204
B.sub (+) Exo VII	300300	4217	1.4	1
E.coli (-) Exo VII	145000	15866	11.0	101
E.Coli (+) Exo VII	347387	7101	2.0	5

Table 2: **Pre-treating sheared gDNA with Exo VII dramatically reduces % palindromes thereby improving assembly.**

Conclusions

- Multiplexing maximizes throughput and efficiency of whole microbial genome sequencing, while reducing cost per sample.
- Achieving higher % P1 loading allows for increased multiplexing levels.
- Treatment with Exo VII before end-repair and ligation significantly reduces the amount of palindromic sequences.
- Smaller genomes like that of *H. pylori*, 6 (possibly 8) strains may be multiplexed.
- A 10 kb, non-size-selected library is sufficient to multiplex a small organism like *H. pylori*.

Future Developments

- Publish procedure and guidelines for multiplexing microbial genomes.
- Compare performance of10 kb and 20 kb SMRTbell libraries.
- Study the effects of size selection of pooled SMRTbell libraries.
- Determine the number of multiplex levels that can be performed on the Sequel System.