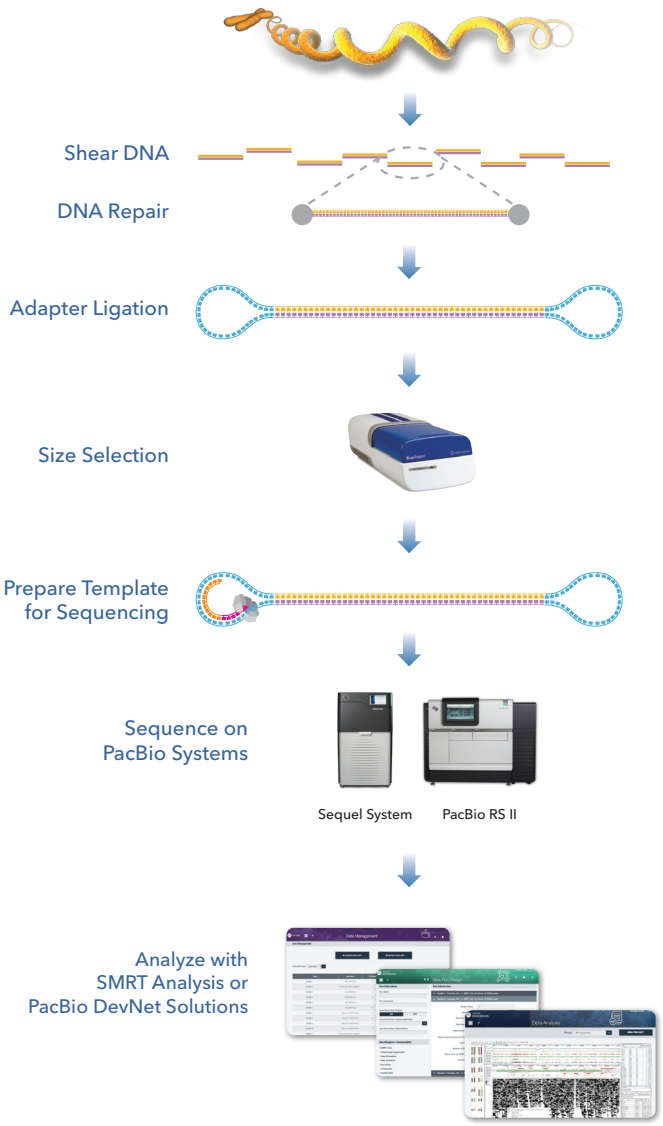
# PLANT AND ANIMAL WHOLE GENOME SEQUENCING
## BEST PRACTICES

PLANT AND ANIMAL SCIENCES

With Single Molecule, Real-Time (SMRT®) Sequencing and the Sequel™ System, you can easily and affordably generate high quality assemblies for even the most complex genomes. Users are regularly achieving plant and animal genome assemblies with megabase-size contig N50s and consensus accuracies >99.999%, resulting in the most complete genomes available today.

### FROM GENOMIC DNA TO A COMPLETE GENOME

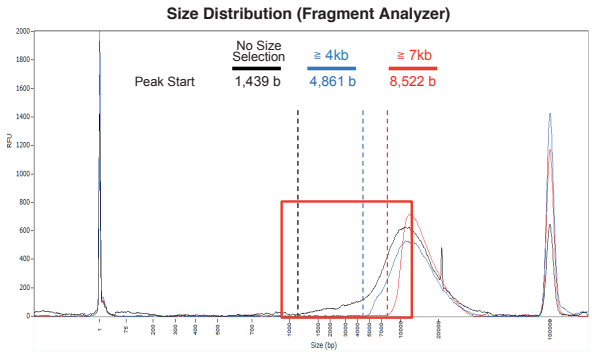



### SAMPLE PREPARATION RECOMMENDATIONS

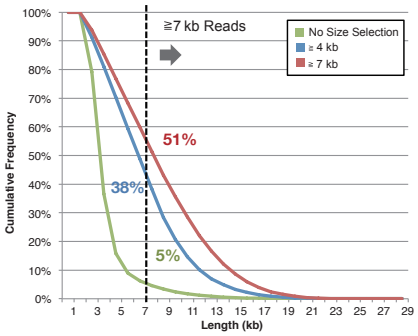
- Use recommend high-quality DNA input (>5 µg)
- Prepare long-insert SMRTbell™ library
  - 20 kb protocol[1] sufficient for most plant and animal genomes
  - >20 kb protocol[2] recommended for high-quality gDNA samples
  - >30 kb protocols[3,4] available for genomes with long, complex repeats
- Size selection recommended
- Sequence to sufficient depth for accurate assembly; recommend:
  - 50-fold coverage for relatively homozygous diploids
  - >80-fold coverage for heterozygous or polyploid organisms
  - ~5 Gb generated per each SMRT Cell 1M on the Sequel System

### SIZE SELECTION ENRICHES FOR LONGEST INSERTS

A




B

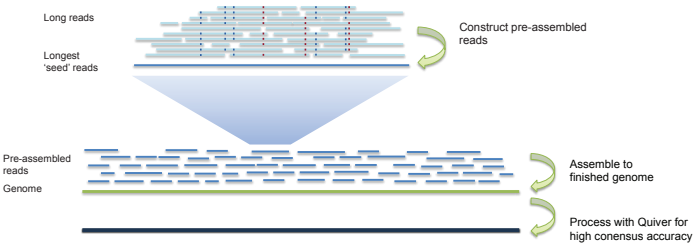



**(A)** *Fragment Analyzer™ trace showing SMRTbell library sizes with no size-selection (black), a 4 kb selection (blue) or a 7 kb selection (red) performed with Blue Pippin™ (Sage Sciences).* **(B)** *Following SMRT Sequencing, the frequencies of reads spanning 7 kb regions increased by 38% for 4 kb size-selected libraries or 51% for 7 kb size-selected libraries[5]. Data provided by courtesy of Okinawa Institute of Advanced Sciences, Japan.*

## DATA ANALYSIS SOLUTIONS WITH SMRT® ANALYSIS AND PACBIO® DEVNET

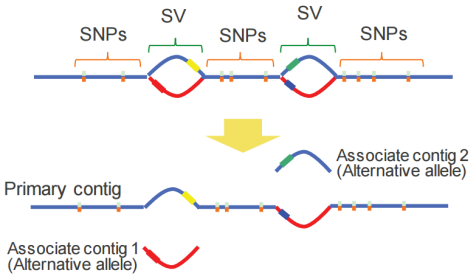
- Exceptional *de novo* genome assemblies with megabase-size contig N50s
- High-quality consensus accuracies >99.999%
- Diploid- and polyploid-aware genome assembly; long reads provide phasing information for heterozygous genomes
- Open-source software
- Advanced data visualization and mining
- Assembly file in BAM and FASTA/Q formats allow for seamless input into downstream tools for further analysis

## HIERARCHICAL GENOME ASSEMBLY PROCESS (HGAP)




*HGAP utilizes all PacBio data using the longest reads for contiguity and all reads to generate high-quality* de novo *assemblies with high consensus accuracy (>QV50)[6].*

## DIPLOID-AWARE GENOME ASSEMBLY PROCESS

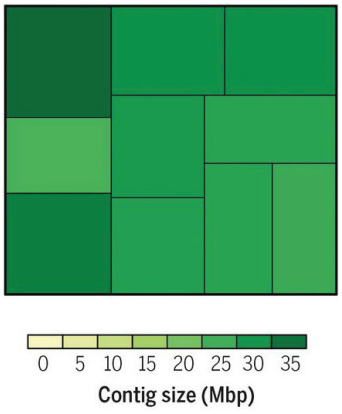



*In diploid-aware assembly, error-corrected reads are assembled using a string graph of the read overlaps, generating primary and alternative contigs that represent the alternative alleles, or structural variants (SVs), between the haplotypes[7]. Recommended for heterozygous, non-inbred genome assembly.*

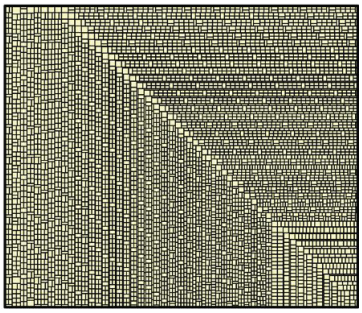
## ACHIEVE HIGHLY CONTIGUOUS GENOME ASSEMBLIES

**A Susie**
Reference sample

**B Long-read assembly**
(Susie 3)




**C Short-read assembly**
(gorGor3)



*A new gorilla genome assembly was generated with SMRT Sequencing, resulting in a >150-fold improvement over previous assemblies. The rectangles in the treemaps in panels B and C are the largest contigs that cumulatively make up 300 Mb (~10%) of the assembly, showing the improved contiguity of the long-read assembly[8].*

## KEY REFERENCES


1. Procedure and Checklist – 20 kb Template Preparation Using BluePippin™ Size-Selection System.
2. Procedure & Checklist >20 kb Template Preparation Using BluePippin™ Size-Selection System (15 - 20 kb Cutoff) for Sequel™ Systems.
3. Preparing >30 kb SMRTbell™ Libraries Using Needle Shearing and BluePippin™ Size-Selection System.
4. Procedure & Checklist - Preparing > 30 kb SMRTbell™ Libraries Using the Megaruptor® Shearing and BluePippin™ Size-Selection System.
5. Okinawa Institute of Advanced Sciences, Japan and Nippon Genetics (2014) Application Note: Long-fragment library preparation for PacBio RS P5-C3 Chemistry. Available at http://www.sagescience.com/wp-content/uploads/2014/10/application_note_2014_07_Eng.pdf
6. Chin, C.S. et al. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT Sequencing data. *Nature Methods*. 10(6), 563-569.
7. Chin, C.S. et. al. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*. 13(12), 1050-1054.
8. Gordon D. et al. (2016) Long-read sequence assembly of the gorilla genome. *Science*. 352 (6281), aae0344.





PN: BP102-011017