PACBIO®

Genome Project Starter Pack

Application Brief
Best Practices for Whole Genome Sequencing

Technical Note
Preparing DNA for PacBio Whole Genome Sequencing for *De Novo* Assembly – Quality Control and Shearing

Overview
Assembly Options for your SMRT Sequencing Data

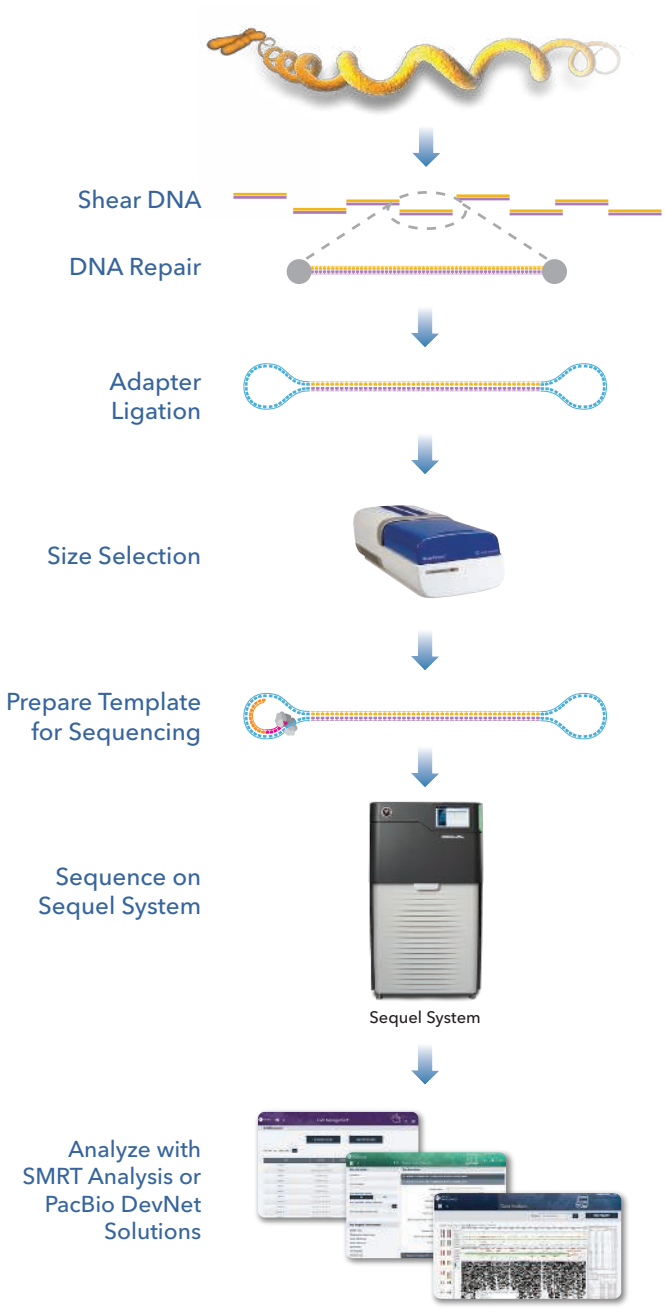
PacBio Certified Service Providers
Global Network of Validated Service Companies that Provide Access to SMRT Sequencing

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%, resulting in the most complete genomes available today.

FROM GENOMIC DNA TO A COMPLETE GENOME

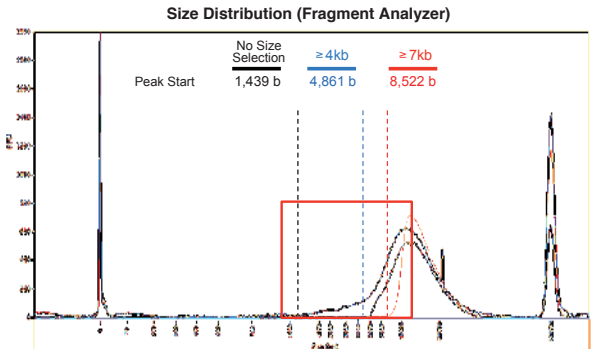


WORKFLOW RECOMMENDATIONS

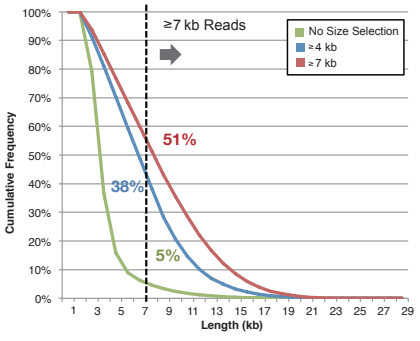
- Start with high-quality, high molecular weight DNA
- Prepare long-insert SMRTbell® libraries[1] for sequencing
- Size select library for optimal yield and read lengths
- Sequence on the Sequel System to depth required for your project (yields up to 10 Gb per SMRT Cell 1M)
  - 40 to 50-fold coverage for inbred organisms
  - 80+ -fold coverage for outbred or heterozygous organisms
- Utilize the PacBio® software suite[2] for genome assembly

SIZE SELECTION ENRICHES FOR LONGEST INSERTS

A



B

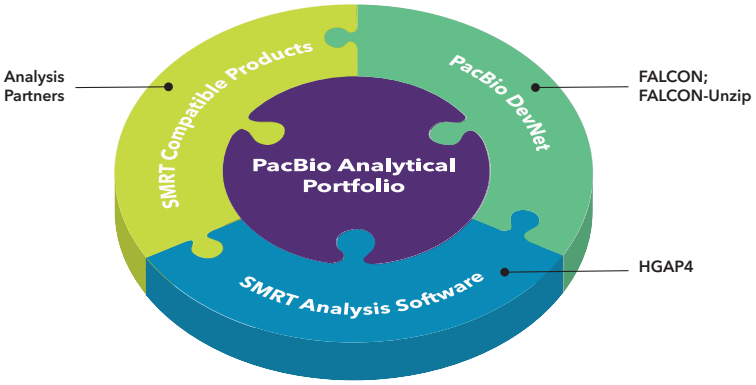


***(A)** Fragment Analyzer™ trace showing SMRTbell library sizes with no size-selection (black), a 4 kb selection (blue) or a 7 kb selection (red) performed with Blue Pippin™ (Sage Sciences). **(B)** Following SMRT Sequencing, the frequencies of reads spanning 7 kb regions increased by 38% for 4 kb size-selected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

Generate highly accurate *de novo* assemblies with megabase-level contiguity

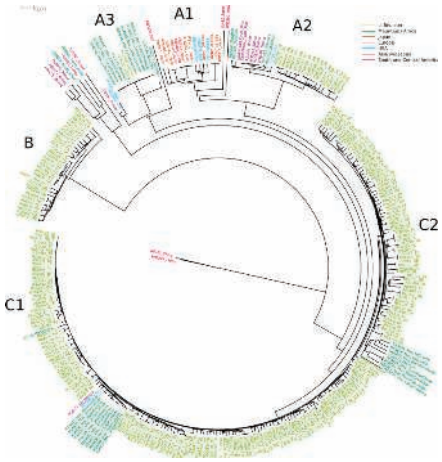
- Push-button assembly for small genomes with HGAP4[3]
- Phased assembly at the command line with FALCON and FALCON-Unzip[4]
- Community-developed tools to customize your *de novo* assembly pipeline[5]
- Network of analysis partners for platform or full-service bioinformatics[6]



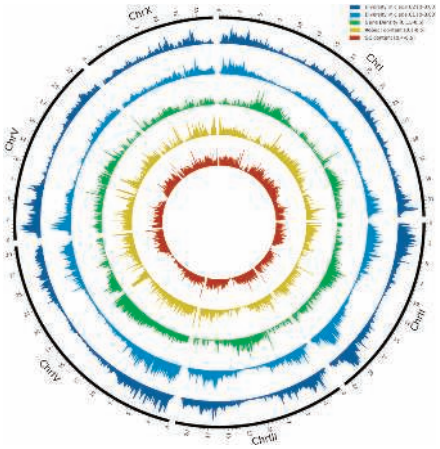
The PacBio analytical portfolio offers solutions for every assembly expertise level, from command line tools to no-hassle full-service genomics.

BUILDING BETTER GENOMES. ENABLING BREAKTHROUGH DISCOVERIES.

A Nematode phylogenetic relationships revisited



B Genome architecture of *P. pacificus*



C *Pristionchus pacificus* genome assembly

Assembly Statistics	
Platform	Sequel System
SMRT Cells 1M	4
Contig N50	4.6 Mb
Contigs	135

An improved nematode assembly enabled the characterization of chromosomal patterns within isolates and between nematode species. The high-quality genome assembly was achieved using just 4 SMRT Cells on the Sequel System[7].

KEY REFERENCES

1. Procedure & Checklist- Preparing >30 kb Libraries Using SMRTbell® Express Template Preparation Kit
2. BFX Summary for Plant and Animal Sciences- Assembly Options for Your SMRT Sequencing Data
3. SMRT Analysis algorithms for de novo assembly-HGAP
4. FALCON and FALCON-Unzip Information- read the docs
5. Devnet analysis tools for de novo assembly
6. SMRT Compatible Analysis Products and Partners
7. Rodelsperger, C. et al. (2017) Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism *Pristionchus pacificus*. Cell Reports 21(2), 834-844.

PN: BP102-020718

PACBIO®

Technical Note

DNA Prep

PACBIO®

Preparing DNA for PacBio® Whole Genome Sequencing for *de novo* Assembly: Quality Control and Shearing

Introduction

Single Molecule, Real-Time (SMRT®) Sequencing uses the natural process of DNA replication to sequence long fragments of native DNA. As such, starting with high-quality, high molecular weight (HMW) genomic DNA (gDNA) will result in longer libraries and better performance across difficult to sequence regions of the genome. This technical note is intended to give recommendations, tips and tricks for assessing and preserving the quality and size of your gDNA sample, shearing methods, and size selection procedures for samples intended to be used with whole genome sequencing for *de novo* assembly.

Topics Covered

DNA Quality Control (QC)

- Purity, quantification, and concentration
  - Qubit® fluorometer
  - NanoDrop® spectrophotometer
- Size
  - CHEF Mapper® System
  - Pippin Pulse™ System
  - FEMTO Pulse™ System

DNA Shearing

  - Megaruptor System
  - Needle shearing

DNA QC

We recommend using both fluorometric and spectrophotometric methods, for example using both the Qubit and NanoDrop instruments, when assessing your DNA sample. If the sample is pure gDNA, free of any RNA contaminants and other small molecules, the two methods should converge on similar amounts of gDNA.

Purity

Typically, we expect to see the NanoDrop A260/280 ratio between 1.8 and 2.0 and the A260/230 ratio ≥2.0 for ultrapure gDNA.

If the NanoDrop purity readings (A260/280 and A260/230) are out of the range specified above, we recommend performing an AMPure® purification step (see your selected library preparation protocol for AMPure instructions) followed by re-assessment of quantity and purity of the gDNA sample.

Quantification

If the measured NanoDrop concentration is significantly different (>50%) from the Qubit measurement, we recommend doing an AMPure purification step, as specified by your chosen library preparation protocol, followed by a re-measurement with both methods. Typically, a single AMPure purification step resolves the discrepancy.

- If the concentration measurement discrepancy after one AMPure purification step is ***reduced***, but not quite below the 50% difference threshold, we recommend proceeding with library preparation.
- If the concentration measurement discrepancy after one AMPure purification step is ***not reduced***, we recommend trying another cleanup approach such as a salt:chloroform wash protocol before a re-measurement with both methods.

PACBIO®

gDNA Sample	Qubit Conc. (ng/µl)	NanoDrop Conc. (ng/µl)	Concentration Difference	A260/280	A260/230
Example 1	49.0	34.5	34.7%	1.87	2.49
Example 2	72.2	19.4	52.8%	1.84	2.16

Table 1 - Example evaluation of gDNA samples. *Example 1 has a measured concentration difference of ≤50% with NanoDrop purity numbers in the expected range. Therefore, it is greenlighted to proceed to the next step. Example 2 has a measured concentration difference of >50% and therefore should go through an AMPure purification step before being re-evaluated.*

DNA Purity and Quantification – Alternative Methods

If you are unable to measure the quantity and/or purity with both the Qubit and NanoDrop instruments, it is still possible to generate a successful long-insert library following the guidance below.

- If you are using only a NanoDrop instrument and the purity numbers are in the appropriate range (A260/280 of 1.8-2.0 and A260/230 of ≥2.0), we recommend proceeding to library preparation.
- If you are using only a Qubit instrument, we recommend performing one AMPure purification step before a re-measurement of concentration and proceeding to library preparation.

Concentration

Additionally, we recommend having the gDNA at a concentration of ≥100 ng/µl prior to shearing and/or SMRTbell® library construction. This can be achieved by additional AMPure purification steps.

Downstream Step	Optimum Input Concentration
CHEF Mapper System	100 ng/lane
Pippin Pulse System	100 ng/lane
FEMTO Pulse	5-500 pg/µl
Needle Shearing	250 ng/µl
Megaruptor	25-50 ng/µl

Table 2 - Optimal input concentration for various DNA sizing and shearing methods. *DNA sizing and shearing methods are listed in order of preference for HMW gDNA preparation.*

Size

*To determine whether you need to shear the gDNA, we recommend determining the gDNA size distribution by **one** of the following DNA sizing methods:*

1. CHEF Mapper System (BioRad)
2. Pippin Pulse System (Sage Science)
3. FEMTO Pulse System (Advanced Analytical)

- If the gDNA mean fragment size is in the range of 10 kb to 100 kb, we recommend proceeding directly to SMRTbell library preparation without shearing of the DNA. This will ensure obtaining the very longest possible sequencing reads and highest yield from the runs performed on PacBio Systems.
- If the gDNA is well above 100 kb, typically identified by being beyond the resolution of the sizing instrumentation used, we recommend shearing the gDNA.

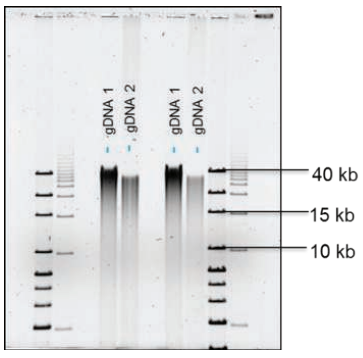


Figure 1. DNA QC example using Pippin Pulse. *The size distribution of gDNA1 is 20 kb to 40 kb. It is possible to construct a library with gDNA 1 without shearing the sample and achieve long sequencing reads. However, gDNA 2 is degraded with the mean fragment size <40 kb. This gDNA is still viable for library preparation but will likely not produce the optimal read lengths or yield on PacBio Systems.*

PACBIO®

DNA Shearing

The two recommended methods for shearing gDNA to the desired size distribution are needle shearing and the Megaruptor System (Diagenode). To ensure sufficient yields of final >30 kb libraries, input gDNA must be sheared carefully so that the average size of fragmented DNA remains well above the desired size selection cut-off. The response of individual gDNA samples to recommended shearing parameters may differ and must be determined empirically and evaluated by pulsed-field gel electrophoresis (PFGE). Test shears are highly recommended.

- For needle shearing, we recommend starting with a gDNA concentration of 250 ng/µl and performing the shearing as recommended in our official guideline protocols (see pages 4-5 in our SMRTbell Express Template Preparation Protocol).
- For shearing with the Megaruptor System, we recommend starting with a gDNA concentration of 25-50 ng/µl and shearing according to the manufacturer's recommendations at either 50-60 kb setting range for >30 kb libraries or 75 kb setting for >40 kb libraries. For further recommendations on using the Megaruptor System for shearing, please see pages 4-5 in our SMRTbell Express Template Preparation Protocol.

AMPure Purification and SMRTbell Library Prep

After shearing the gDNA to the desired mean fragment size, we recommend performing an AMPure purification step and resuspending the gDNA in Elution Buffer (included with PacBio Template Prep Kits) at a concentration of ~140 ng/µl prior to SMRTbell library preparation. The gDNA is then converted to SMRTbell library molecules according to our recommended library preparation protocols, which are posted on the PacBio Documentation webpage. Specific protocols for constructing long-insert libraries for *de novo* assembly applications are listed below.

Recommended Protocols

Procedure & Checklist – Preparing Greater Than 30 kb Libraries Using SMRTbell Express Template Preparation Kit

Procedure & Checklist – Preparing Greater Than 30 kb SMRTbell Libraries Using Needle Shearing and BluePippin Size-Selection on PacBio RS II and Sequel Systems

Procedure & Checklist – Preparing Greater Than 30 kb SMRTbell Libraries Using Megaruptor Shearing and BluePippin Size-Selection on PacBio RS II and Sequel Systems

Additional Protocols

Procedure & Checklist – Preparing Greater Than 15 kb Libraries Using SMRTbell Express Template Preparation Kit

Procedure & Checklist – >20 kb Template Preparation Using BluePippin Size-Selection System (15 – 20 kb Cutoff) for Sequel Systems

PN: TN101-041818

ASSEMBLY OPTIONS FOR YOUR **SMRT SEQUENCING DATA**

THE LEADER IN LONG-READ SEQUENCING

PACBIO®

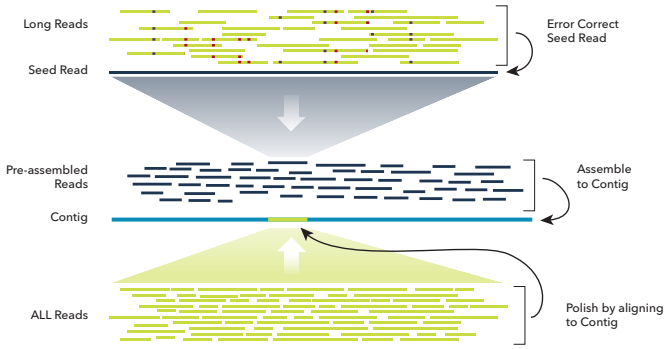
PLANT AND ANIMAL SCIENCES

It is now easier and more affordable than ever to do a genome project with PacBio® Single Molecule, Real-Time (SMRT®) Sequencing—the gold standard for generating contiguous, highly accurate reference genomes. Assembly is no longer a challenge with the wealth of bioinformatics tools developed and optimized for SMRT Sequencing data, enabling you to generate high-quality genome assemblies on **any budget**.

PACBIO GENOME ASSEMBLY

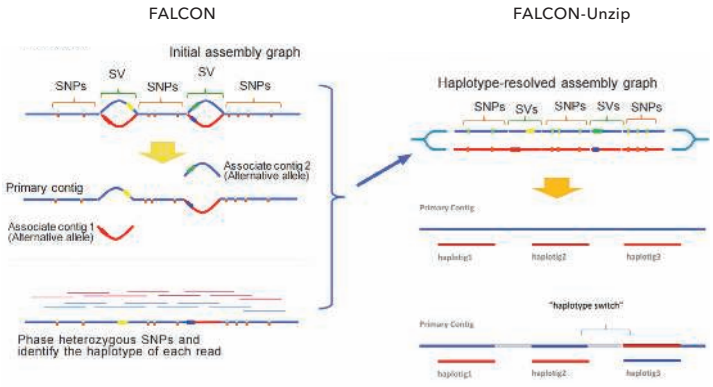
Using the Hierarchical Genome Assembly Process (HGAP) and long-read polishing, SMRT Sequencing data produces genome assemblies with megabase-size contig N50s and >99% accuracy. Explore the options for assembling SMRT Sequencing data below.

SMRT ANALYSIS



- Free for anyone to download and use at pacb.com/software-downloads
- Includes HGAP4, our push-button assembly tool with polishing built into the pipeline
- Use with genomes ≤3 Gb

FALCON AND FALCON-Unzip



- Open source for anyone to use at pacb.com/devnet
- FALCON: specialized assembly parameterization for genomes of any size
- FALCON-Unzip: genome phasing of heterozygous or outbred organisms

ANALYSIS PARTNERS

BLUEBEE

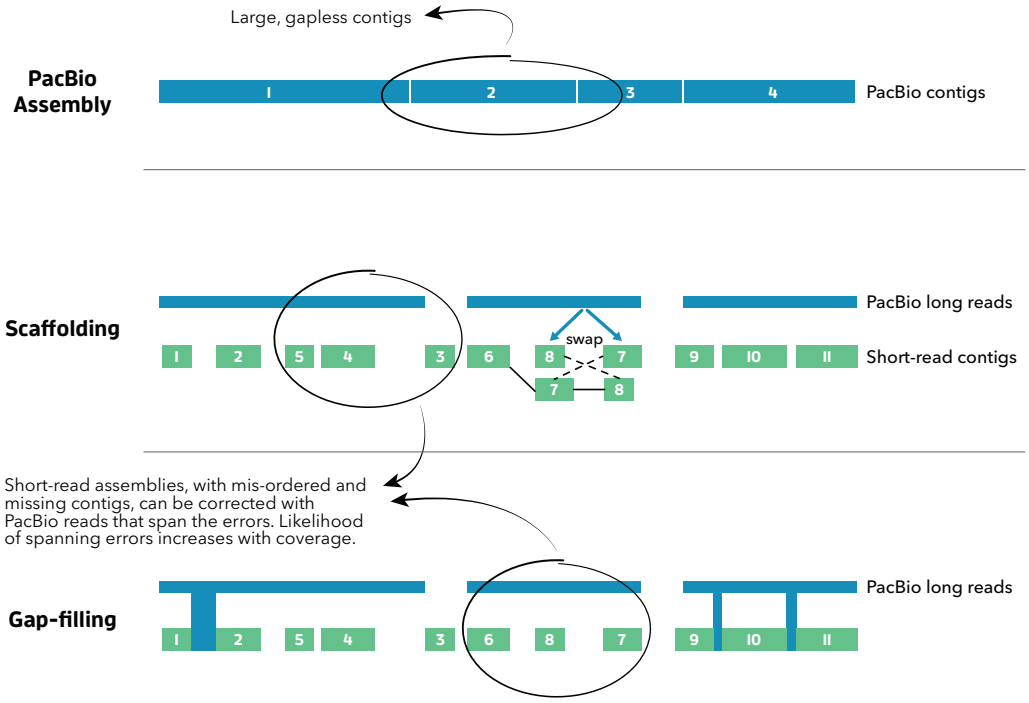
computomics®

DNAnexus

With these PacBio analysis partners, you have the option to completely outsource your bioinformatics or compute resource needs. Options range from cloud-based, self-serve analysis to no-hassle, full-service assembly.

HYBRID GENOME ASSEMBLY

The most contiguous and accurate contig assemblies are generated with SMRT Sequencing data alone. Alternatively, for significantly lower cost than a full PacBio assembly, you can also use SMRT Sequencing data to improve highly fragmented draft genomes through scaffolding of contigs or filling gaps left in short-read assemblies.



PUBLICATION SPOTLIGHT

Browse this GigaScience paper[1] where both PacBio and hybrid assemblies were generated for the complex, hexaploid wheat genome.

COVERAGE REQUIREMENTS

Although every genome project is unique in complexity and downstream utility, we have compiled a quick guide below to help you assess the coverage requirements and expected contiguity for genome assembly using SMRT Sequencing data or hybrid assembly approaches.

METHOD	PACBIO COVERAGE	EXPECTED CONTIGUITY	GAPS IN ASSEMBLY
PacBio Assembly	40 to 50-fold per haplotype	Megabases	NO
Scaffolding[2]	10 to 30-fold	Tens to hundres of kb	YES
Gap-filling[3]	5 to 10-fold	Tens to twenties of kb	YES

KEY REFERENCES

1. Zimin, A. et al. (2017) The first near-complete assembly of the hexaploidy bread wheat genome, *Triticum aestivum*. GigaScience 6 (11), 1-7.
2. Uliano-Silva, M. et al. (2018) A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*. GigaScience 7 (2), 128.
3. Xu, S. et al. (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. PNAS 114 (23), 6133-6138.

PN: SW101-061318

PACBIO CERTIFIED SERVICE PROVIDER

CERTIFIED SERVICE PROVIDERS

North American Certified Service Providers

Arizona Genomics Institute
Tucson, AZ
www.genome.arizona.edu
email: dkudrna@email.arizona.edu
ph: +1.520.626.9596

City of Hope Integrated Genomics Core
Monrovia, CA
www.cityofhope.org/research/shared-resources/integrative-genomics-core
email: xwu@coh.org
ph: +1.626.256.4673 ext. 85071

GENEWIZ
South Plainfield, NJ
www.genewiz.com
email: ngs@genewiz.com
phone: +1.908.222.0711

Genomics Resource Center, Institute for Genome Sciences, University of Maryland
Baltimore, MD
www.igs.umaryland.edu/resources/grc/index.php
email: grc-info@som.umaryland.edu
ph: +1.410.706.5668

Icahn Institute for Genomics and Multiscale Biology at Mount Sinai
New York, NY
www.icahn.mssm.edu/genomics
email: robert.sebra@mssm.edu

McGill University and Génome Québec Innovation Centre
Montreal, Canada
www.gqinnovationcenter.com/index.aspx
email: infoservices@genomequebec.com
ph: +1.514.398.7211

MOgene LC
St. Louis, MO
http://www.mogene.com/
email: srangwala@mogene.com
ph: + 1.314.669.2591

National Center for Genome Resources (NCGR) - Sequencing Center
Sante Fe, NM
www.ncgr.org
email: seq@ncgr.org
ph: +1.505.995.4449

RTLGenomics, LLC
Lubbock, TX
www.rtlgenomics.com
email: info@researchandtesting.com
ph: +1.806.771.1134

University of Delaware - DNA Sequencing & Genotyping Center (Delaware Biotechnology Institute)
Newark, DE
http://www.udel.edu/dnasequence
email: DNAsequence@udel.edu
ph: +1.302.831.0823

University of Minnesota Genomics Center
Minneapolis, MN
genomics.umn.edu
email: next-gen@umn.edu
phone: +1.612.625.7736

Washington State University (WSU) - Laboratory for Biotechnology and Bioanalysis - DNA Sequencing Core
Pullman, WA
lbb.wsu.edu
email: dnaguy@mail.wsu.edu
ph: +1.509.335.1174

APAC Certified Service Providers

DNA Link, Inc.
Seoul, South Korea
www.dnalink.com
email: office@dnalink.com
ph: +82.2.3153.1500

Genome Institute of Singapore
Singapore
www.a-star.edu.sg/gis
email: soonwjw@gis.a-star.edu.sg
ph: +65.6808.8108

Frasergen Bioinformatics
Wuhan, China
www.frasergen.com
email: support@frasergen.com
ph: +86.27.8722.4696

Kazusa DNA Research Institute,
Chiba, Japan
http://www.kazusa.or.jp
email: ngs-jutaku@kazusa.or.jp
ph: +81.438.52.3900

Macrogen Inc.
Seoul, South Korea
www.macrogen.com
email: ngs@macrogen.com
ph: +82.2.2113.7000

National Instrumentation Center for Environmental Management (NICEM)
Seoul, South Korea
www.nature.snu.ac.kr
email: choii@snu.ac.kr
ph: +82.2.880.4871

The Ramaciotti Centre
Kensington, Australia
www.ramaciotti.unsw.edu.au
email: nextgenseq@unsw.edu.au
ph: +61.2.9385.1241

Takara Bio Inc.
Shiga, Japan
www.takara-bio.co.jp
email: info-sales@takara-bio.co.jp
ph: +81.77.565.6999

European Certified Service Providers

Centre for Genomic Research, University of Liverpool
Liverpool, United Kingdom
www.liv.ac.uk/genomic-research
email: CGR.Enquiries@liverpool.ac.uk
ph: +44.151.795.4509

Earlham Institute
Norwich, United Kingdom
www.earlham.ac.uk/genomic-services
email: projects@earlham.ac.uk
ph: +44.1603.450.001

GATC Biotech
A Eurofins Genomics Company
Konstanz, Germany
www.gatc-biotech.com
email: customerservice@gatc-biotech.com
ph: +49.7531.8160.0

PacBio RS II Sequel® System

For additional facility information, please visit www.pacb.com/SMRTproviders. All information listed was provided voluntarily by facilities willing to provide external services.

400+ PacBio Systems Worldwide

SMRT Sequencing is now more accessible than ever.
New service providers are added regularly.

Contact us to find a PacBio System near you

Headquarters
1305 O'Brien Drive
Menlo Park, CA 94025
United States
Phone: +1.650.521.8000

Singapore Office
20 Science Park Road
#01-22 TeleTech Park
Singapore 117674
Phone: +65.6778.5627

Customer Service
Phone: +1.877.920.PACB (7222) option 1
Fax: +1.650.618.2699
Email: orders@pacb.com

Technical Support
Phone: +1.877.920.PACB (7222) option 2
Email: support@pacb.com

PacBio® Sequencing Providers
www.pacb.com/SMRTproviders

Inquiries
North America:
nasales@pacb.com

South America:
sasales@pacb.com

Europe/Middle East/Africa:
emea@pacb.com

Asia Pacific:
apsales@pacb.com

THE LEADER IN LONG READ SEQUENCING

www.pacb.com

PN: IN102-053018