SMRT® Link Release Notes (v8.0)

SMRT Link Server Installation

- SMRT Link server software is supported on English-language CentOS 6.x; 7.x and Ubuntu 16.04; 18.04 64-bit Linux® distributions. (This also applies to SMRT Link compute nodes.)
- SMRT Link is **not** guaranteed to work on Linux versions that are no longer supported by the Operating Systems' vendors.
- SMRT Link server software **cannot** be installed on macOS® or Windows® systems.
- Several new SMRT Link v8.0 features are computationally-intensive and require adherence to the computational and storage requirements listed in the document **SMRT Link Software Installation (v8.0)**.
- To install **only** command-line SMRT Tools, use the `--smrttools-only` option with the installation command, whether for a new installation or an upgrade. Examples:
  ```
  smrtlink-*.run --rootdir smrtlink --smrttools-only
  smrtlink-*.run --rootdir smrtlink --smrttools-only --upgrade
  ```

Included with the SMRT Link Installation

Barcode Sequences:

- Sequel_16_barcodes_v1
- Sequel_16_Barcodes_v3
- Sequel_96_barcodes_v1
- Sequel_384_barcodes_v1
- RSII_96_barcodes
- RSII_384_barcodes
- IsoSeqPrimers_v2 (Includes the content of IsoSeqPrimers as well as support for NEB and Clontech primers.)

Reference Sequences:

- Sequel DNA Internal Control
- HIV_HXB2
- HLA_11locus_clustering_guide
- lambdaNEB

New Features

SMRT Link

- SMRT Link is powered by the **Cromwell** scientific workflow engine, which replaces `pbsmrtpipe`.
- The SMRT Link Graphical User Interface (GUI) has a new styling and layout.
- New **Microbial Assembly** application provides *de novo* assembly, circularization and plasmid identification.
- Analysis results visualization is now compatible with **Integrative Genomics Viewer (IGV)**. SMRT View is no longer included with SMRT Link.

- The PacBio RS II instrument is no longer supported by this version of SMRT Link. To analyze PacBio RS II data, use SMRT Link v7.0 or earlier.

SMRT Link - Sample Setup

- Protocols involving low DNA input concentrations now use fixed primer and template concentration targets.
- An alternate number of SMRT Cells to be prepared can now be entered in the final sample dilution step. This allows for preparation of a large amount of library (by selecting a large number of SMRT Cells initially), and a smaller quantity to be used at run time.
- SMRT Link now provides warnings if measured sample amounts, going into sample cleanup or recovery after cleanup, are unrealistically high.

SMRT Link - Run QC

- The accuracy of the UMY estimate was improved.

SMRT Link - SMRT Analysis

- Iso-Seq Application
  - The Iso-Seq Analysis application can now use CCS data as input.
  - The Iso-Seq Classify Only and Iso-Seq with Mapping applications are **deprecated**; their functionality is available in the Iso-Seq application.
- Base Modification Analysis Application
  - The **Base Modification and Motif Analysis** application is **deprecated**; the functionality is available in the Base Modification Analysis application.
  - For data from Sequel II Systems, set the Advanced Analysis parameter **Minimum Qmod score** to `100` instead of the default `30`.

Fixed Issues

- PDFs generated when exporting Data Sets no longer contain the spurious value of `2147483647` for **Raw Data: Polymerase Read Bases**, **Raw Data: Unique Molecular Yield**, and **CCS: Q20 Yield**.

Known Issues

SMRT Link

- When creating a project containing Data Sets that were analyzed with other Data Sets, this error may display: "We're Sorry...Something went wrong." However, the project is still created with the correct Data Sets.
- Date sorting does not work correctly for entries created on the same day.

Run Design

- The Run Design CSV file does **not** allow uploading of two samples with identical names and identical barcodes in the same run. The workaround is to set up the run using the GUI.
- Successful upload of a Bio Sample Name file in the Run Design editor does **not** display a confirmation message.

- For insert sizes, entering a decimal value may cause an error. Integer values must be used**.**
- CLR Run Designs created prior to release of SMRT Link v7.0 may appear incorrectly as CCS runs when viewed or edited in SMRT Link v8.0.
- Creating or importing Run Designs that use more than two chemistries causes an error, as only two different chemistries may be loaded on to the instrument.
- Inserting any characters in the Bio Sample Name file other than the allowed alphanumeric, space, dot, underscore, and hyphen characters causes the Demultiplex Barcodes application to fail. Editing the Bio Sample Name file in a plain text editor such as Notepad is recommended, as Microsoft Excel and Word may insert illegal characters invisibly.
- For runs created for Sequel instruments, the summary list of reagents and consumables fails to distinguish between SMRT Cells and SMRT Cells LR and instead reports the total number of cells needed for a run. This will **not** impact instrument behavior; runs containing a mix of SMRT Cells and SMRT Cells LR will load correctly on the instrument, and the instrument will report the correct number of each cell type required.

Data Management

- SMRT Link v8.0 no longer includes weekly automatic database backups. A database backup is still automatically performed once, during installation or upgrade. Failure to back up on a regular schedule risks losing **all** records in SMRT Link (including users, Data Sets, analyses, barcodes, and references) if a file system or reconfiguration error occurs. It does **not** affect underlying sequencing or analysis files, such as BAM files. We **strongly** recommend asking your local Linux System Administrator to schedule regular weekly backups of the SMRT Link database using standard Linux utilities. For additional details, please contact PacBio Technical Support.
- Selection of Data Sets using the **Select All** checkbox incorrectly overrides any existing filters.

SMRT Analysis

- If a Bio Sample Name file is **not** provided during run design and only a subset of barcodes from the Barcode Set is used, the Demultiplex Barcodes analysis will output **all** barcodes in the Barcode Set. When designing a run, provide a Bio Sample Name file with rows containing Bio Sample Names **only** for the barcodes used. Alternatively, rerun the Demultiplex Barcodes application with the advanced parameter **Ignore BioSamples records from Run Design** set to **On**.
- If a wrong or incomplete list of barcode Bio Sample names is included in Run Design, or if a wrong Barcode Set is used during Run Design, the demultiplexed output will contain the wrong number of barcodes. To obtain correct demultiplexing results, rerun the Demultiplex Barcodes application with the advanced parameter **Ignore BioSamples records from Run Design** set to **On**.

- The Resequencing, CCS with Mapping, HGAP4, Structural Variant Calling, and Base Modification Analysis applications fail when launched on extremely large genomes containing chromosomes > 500 Mb.
- The Structural Variant Calling application fails when launched with demultiplexed Data Sets. The workaround is to run `pbmm2` and `pbsv` at the command line, using `--sample` with `pbmm2` to indicate the Bio Sample Name for a demultiplexed Data Set.
- In the instrument touch screen software, inserting non-alphanumeric special characters, such as `&`, in free text fields available to the Administrator (such as **Transfer Scheme description**) causes secondary analysis failures on Data Sets generated by that instrument.
- Some sections of an analysis results page are sometimes reordered upon relaunch. All content remains correct.
- In the Assembly (HGAP4) application, the help text for the advanced parameter **Filters to add to the DataSet** incorrectly says to separate multiple entries with semicolons or commas. Instead, multiple entries should be separated by `AND`.
- Command-line usage of the Assembly (HGAP4) application requires **all** default options to be captured as strings when running Cromwell.
- Using the Minor Variants application through the SMRT Link GUI may result in failed analyses for larger Data Sets. Data can be successfully analyzed using command-line tools. For additional workarounds, please contact PacBio Technical Support.
- In cases of extremely high genome coverage (>1000 fold), the Microbial Assembly application may produce a very high number of contigs, or fail with an out-of-memory error. For additional workarounds, please contact PacBio Technical Support.
- SMRT Tools: Configuration files require **full** file paths to work correctly with Cromwell.
- The `pbcromwell --help` command includes examples with some errors. For correct information, please contact PacBio Technical Support.
- SMRT Analysis fails on compute environments using LSF as a job scheduler. For details on fixing the issue, please contact PacBio Technical Support.