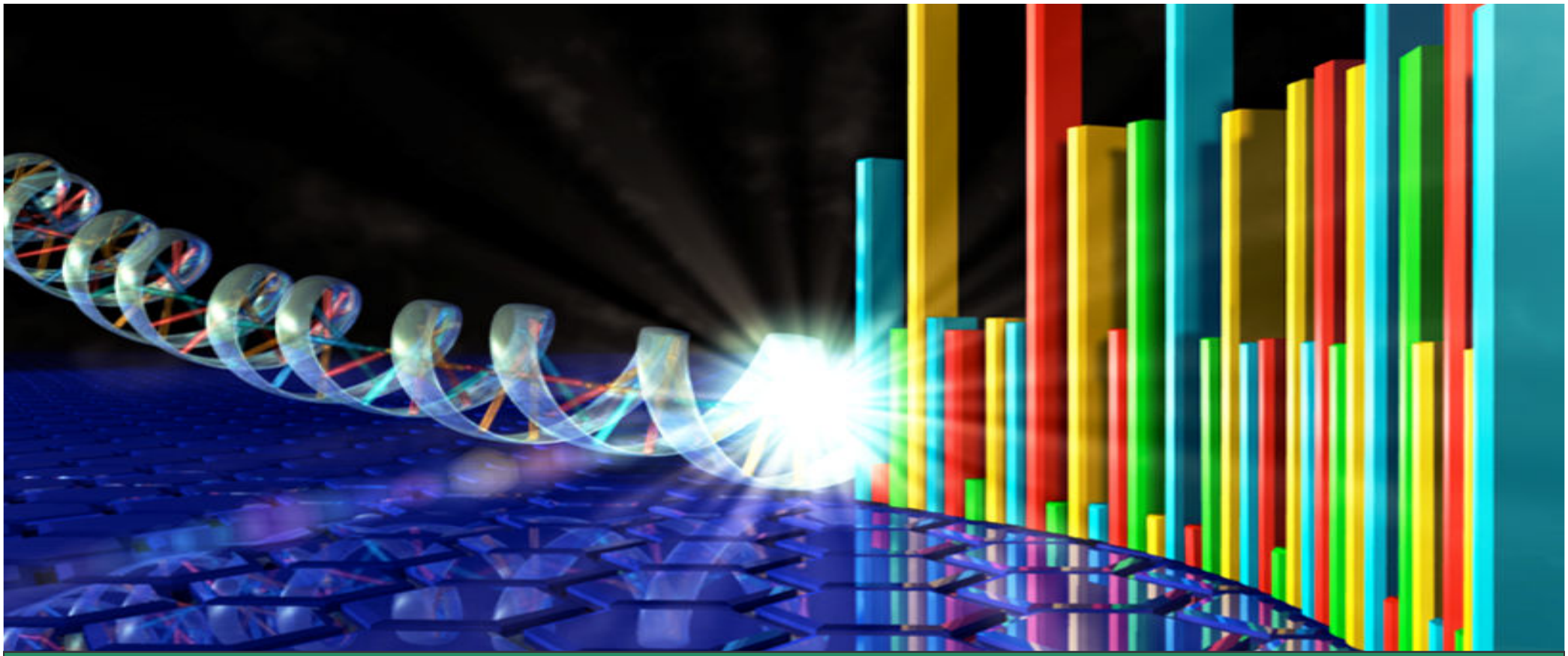
The Genome Sequencer FLX™ Software 2.0.00 Titanium

What is new ?



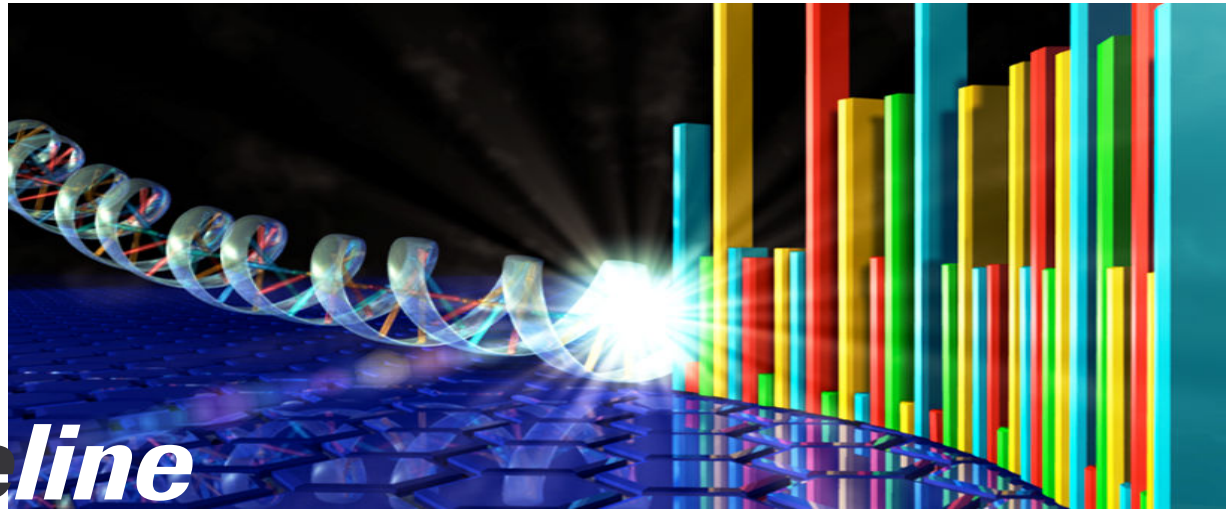
www.genome-sequencing.com

Software Package List

Package	On Instrument	Off Instrument	RPM Available
rigControl / Sequencer	**Only**	Never	No
gsRunProcessor	Required	Yes	OpenMPI-Yes Self Contained -Yes
gsReporter	Required	Yes	Yes
gsRunProcessorManager	Required	Optional*	Yes
gsRunBrowser	Required	Optional	No
gsAssembler	Never	Optional	No
gsMapper	Never	Optional	No
gsAmplicon	Never	Optional	No
gsSupportTool	Optional	Optional	Yes

* Required if users want to launch jobs with gsRunBrowser

Pipeline



Major Changes in the Pipeline

- Elimination of IPC in lieu of industry standard Message Passing Interface (MPI)
  - No more sMinProc.sh
  - No more /usr/local/rig/ required (off instrument)
  - No more pipeline configuration files
  - No more internal shell scripts
  - The pipeline can run on nearly any number of computers simultaneously (“cluster”)
- Retains the familiar runImagePipe, runAnalysisPipe, runAnalysisFilter commands
- Single executable called “gsRunProcessor” contains all the pipeline code
- Introduction of “processing scripts”

Data Set Layout for 2.0.00

- R_2008_06_04_10_11_24_rig2_jkeller_jktestrun
  - rawImages/
  - D_2008_06_04_10_11_32_rig2_imageProcessingOnly
    - regions/
      - 1.cwf, 2.cwf
    - gsRunProcessor.log
  - D_2008_06_04_14_00_10_dataRig10_signalProcessing
    - regions/
      - 1.cwf, 2.cwf
      - Output from gsReporter (.txt, csv, etc.)
    - Sff/
      - FD392FSF01.sff, FD392FSF02.sff
    - gsRunProcessor.log
    - 454DataProcessingDir.xml
    - Output from gsReporter (.txt, .fna., .qual, etc)

Image Capture

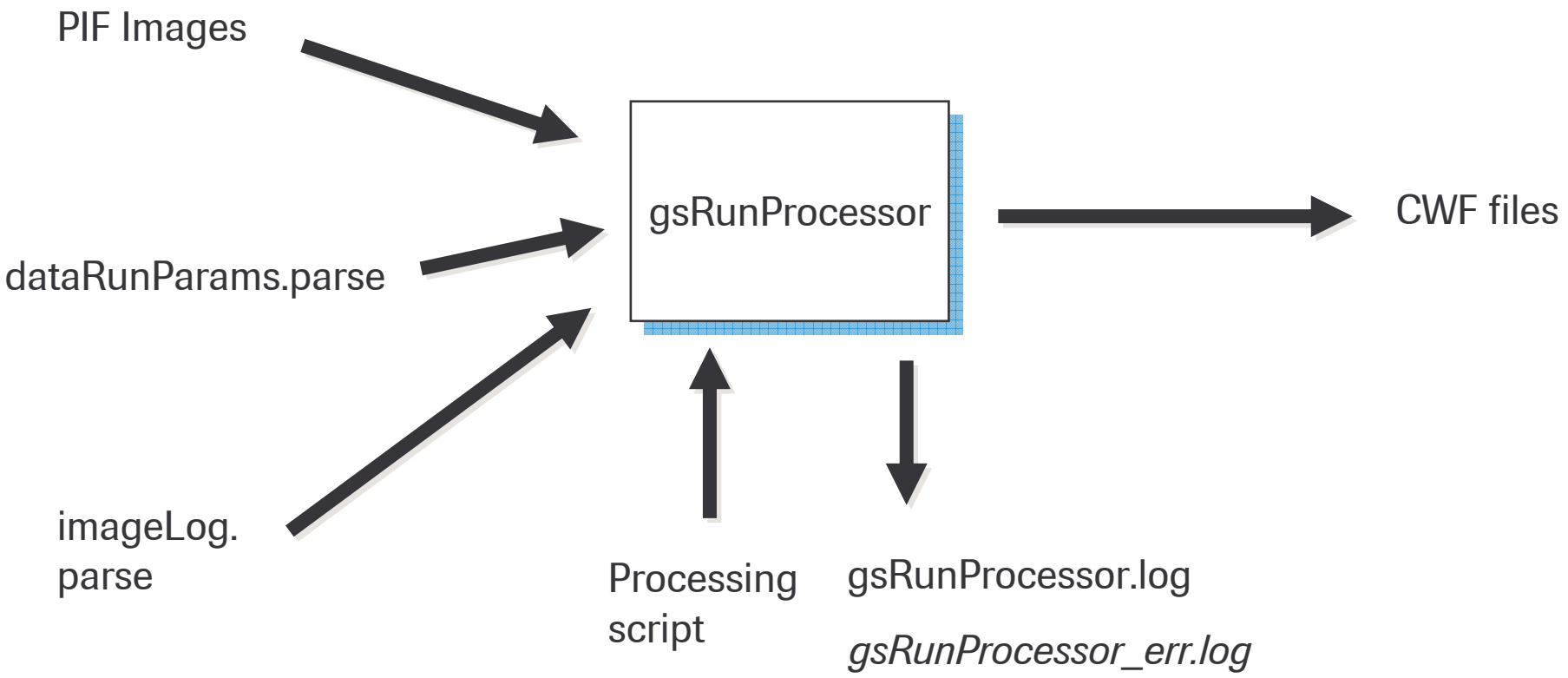
Image Processing Pipeline

Signal Processing Pipeline

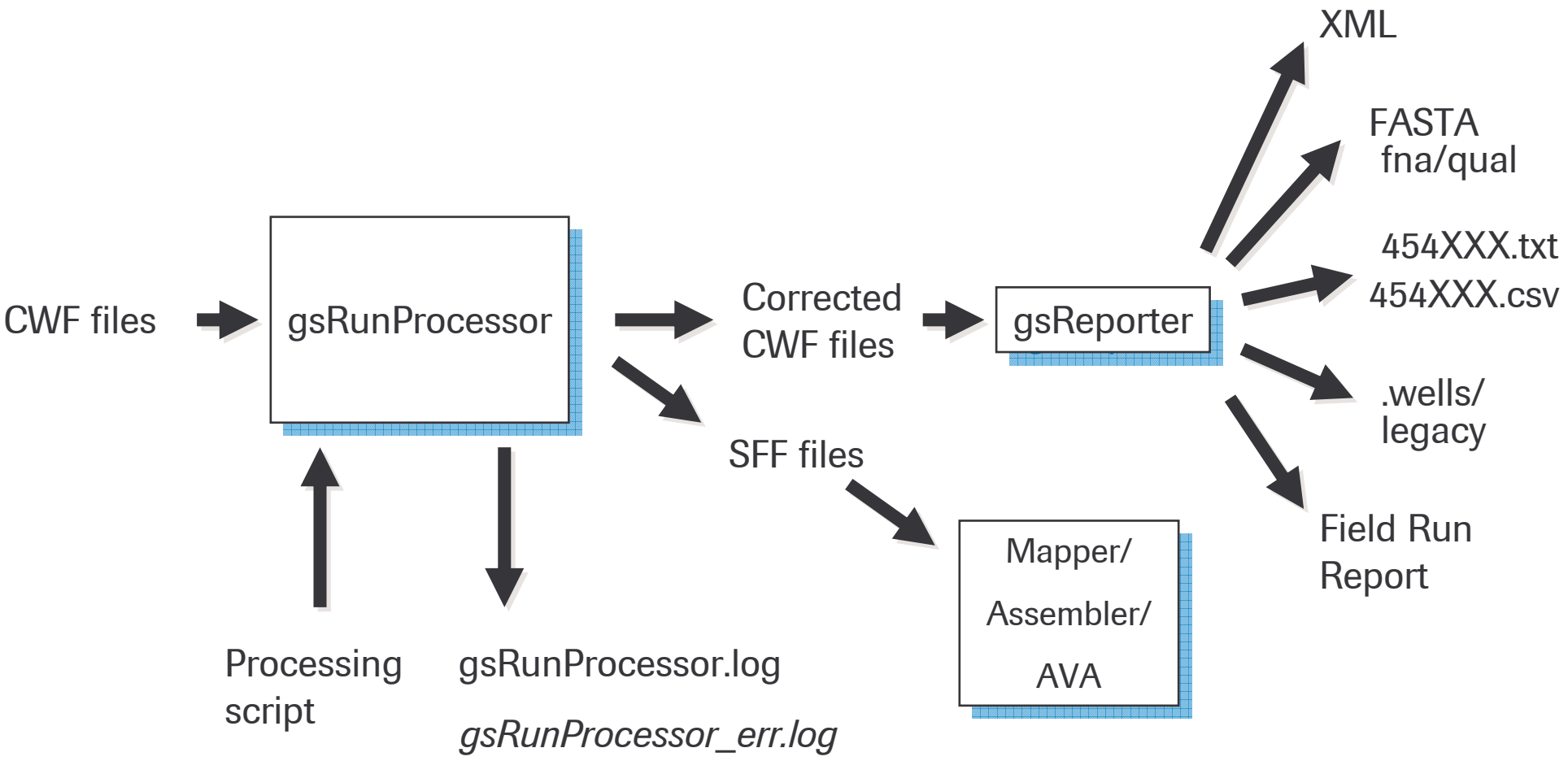
Composite Wells File (.cwf)

- ".wells" file replacement
- Self-contained
- "Package" format based on OpenOffice ".odt"
  - XML based meta-data
  - XML based metrics
  - Full processing history of the data
  - Called bases and quality scores
  - … and the well flowgrams themselves
- Compressed
  - About half the size of a comparable .wells file for the same number of wells and flows

Data Flow For Image Processing



Data Flow For Signal Processing



Recommended Data Workflow

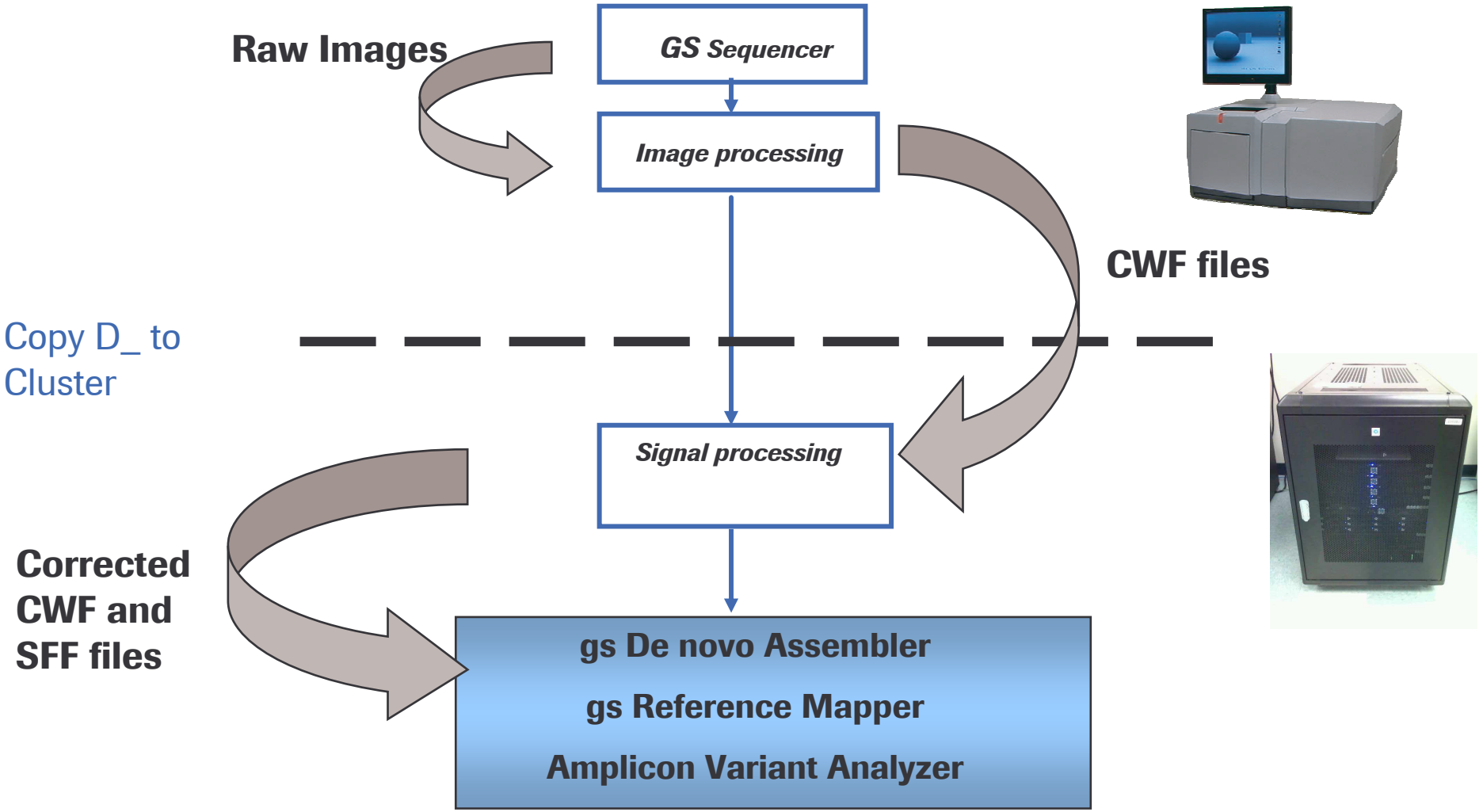
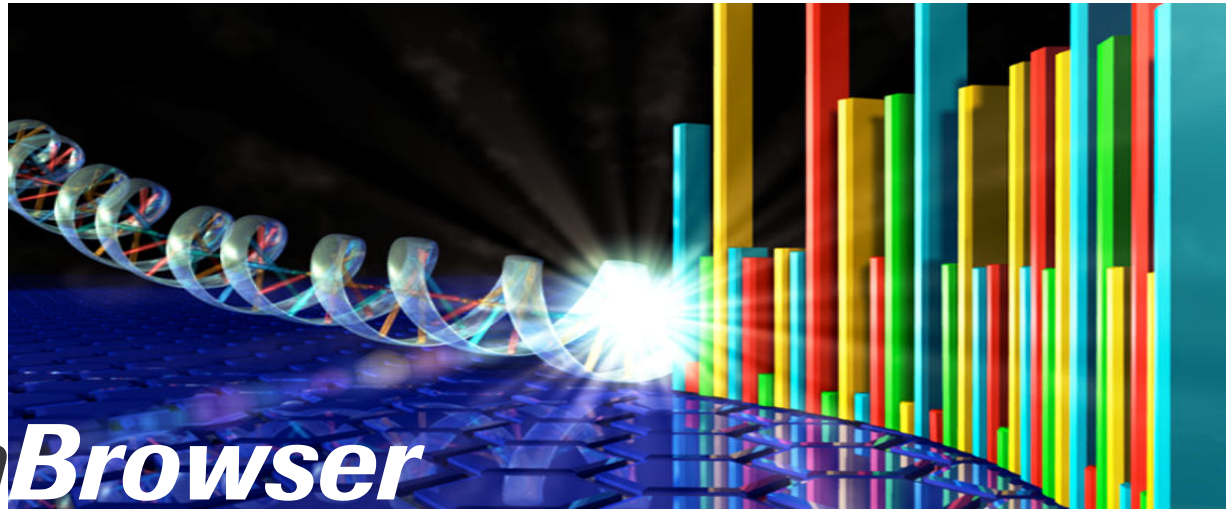


Image and Signal Processing Summary

GS Run Processor

Sequencing Kit	Number of Cycles	Raw Images Only	Raw Images Plus Image Processing	Raw Images Plus Full Processing
XLR70	200	28 Gb	30 Gb	37 Gb
XLR70	150	21 Gb	23 Gb	28 Gb
XLR70	100	14 Gb	16 Gb	18 Gb
LR70	100	14 Gb	15 Gb	16 Gb
LR70	42	5 Gb	6 Gb	7 Gb
LR25	100	6 Gb	7 Gb	8 Gb
LR25	42	3 Gb	4 Gb	4 Gb
SR70	42	5 Gb	6 Gb	7 Gb

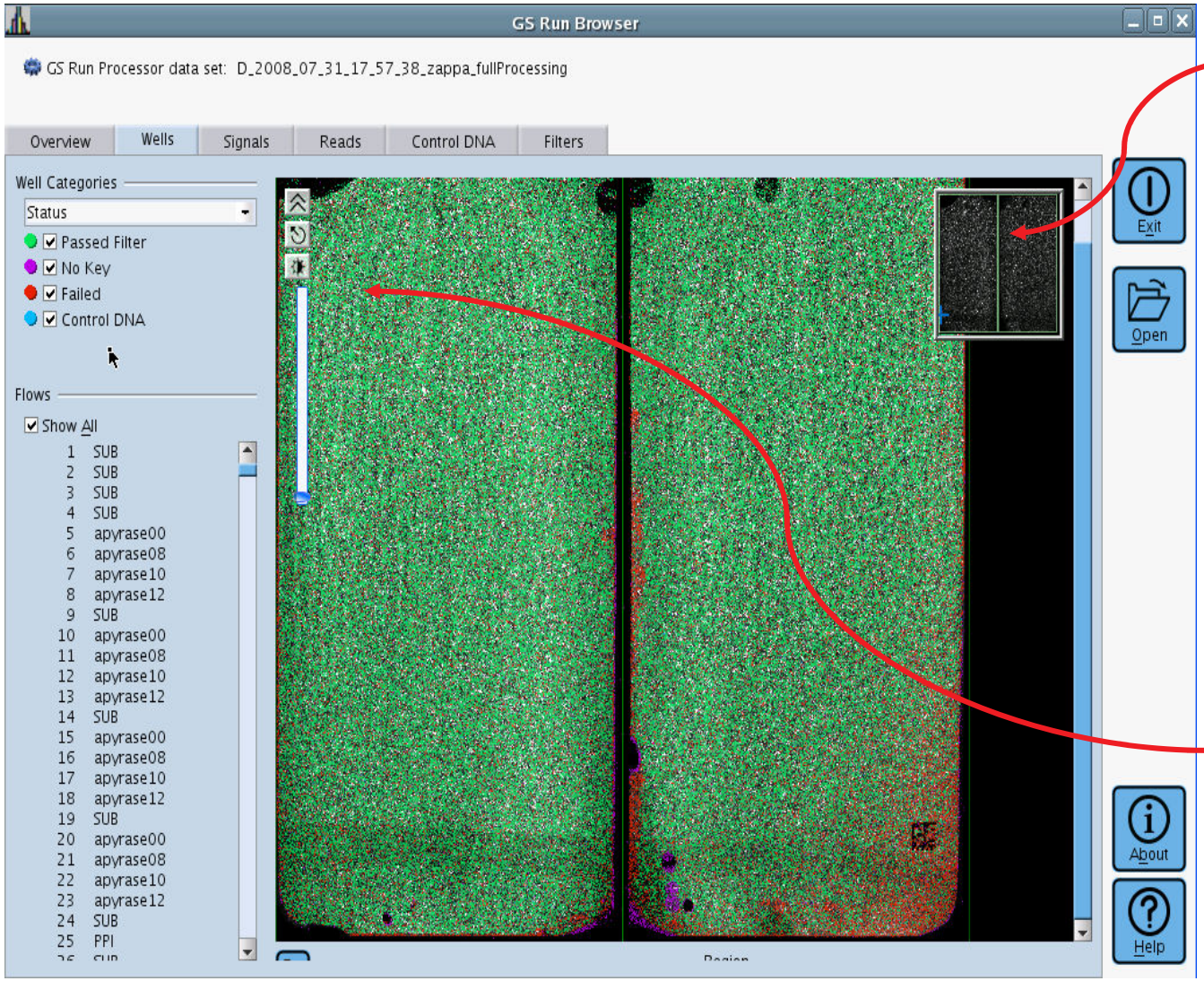
gsRunBrowser



gsRunBrowser – Highlights

- Manage higher well counts
- Ability to browse new cwf data format
- Image and Wells Tab have been combined
- Mapping Tab has been removed
- Improve image and wells rendering and user interaction

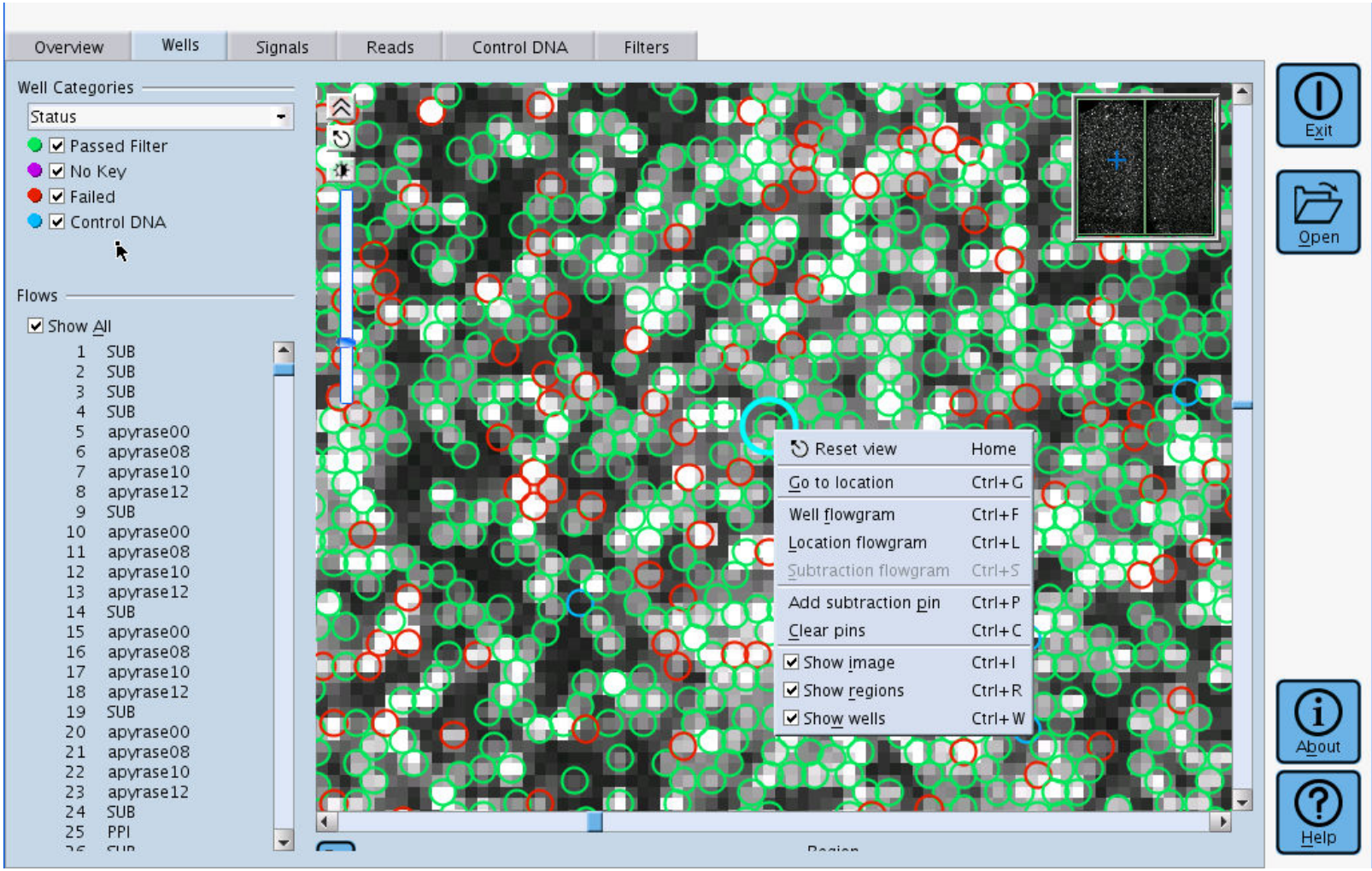
gsRunBrowser – Wells Tab



- Scaled out overview

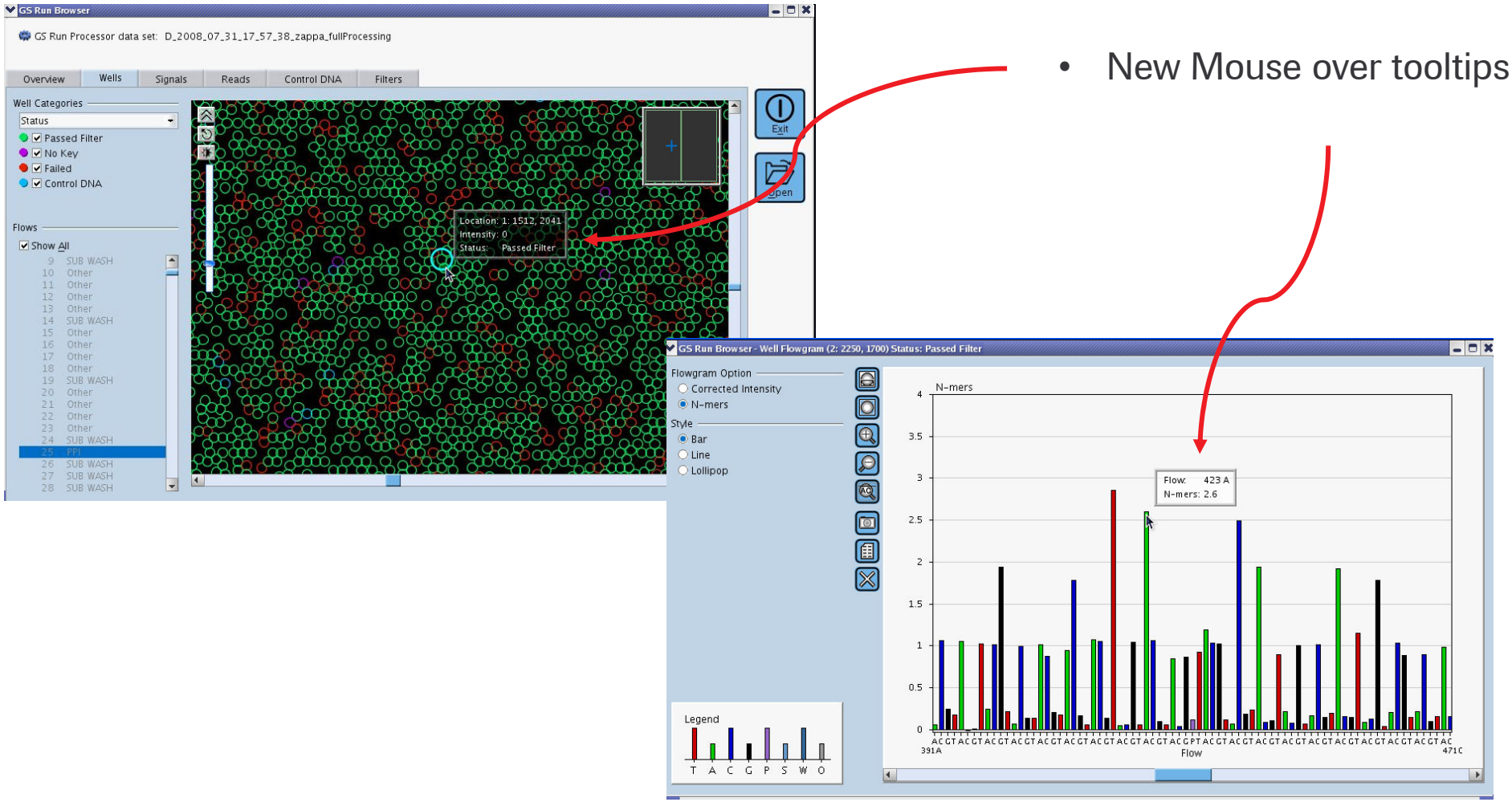
- Collapse button
- Reset button
- Set contrast button
- Scaling slider

gsRunBrowser – Wells Tab



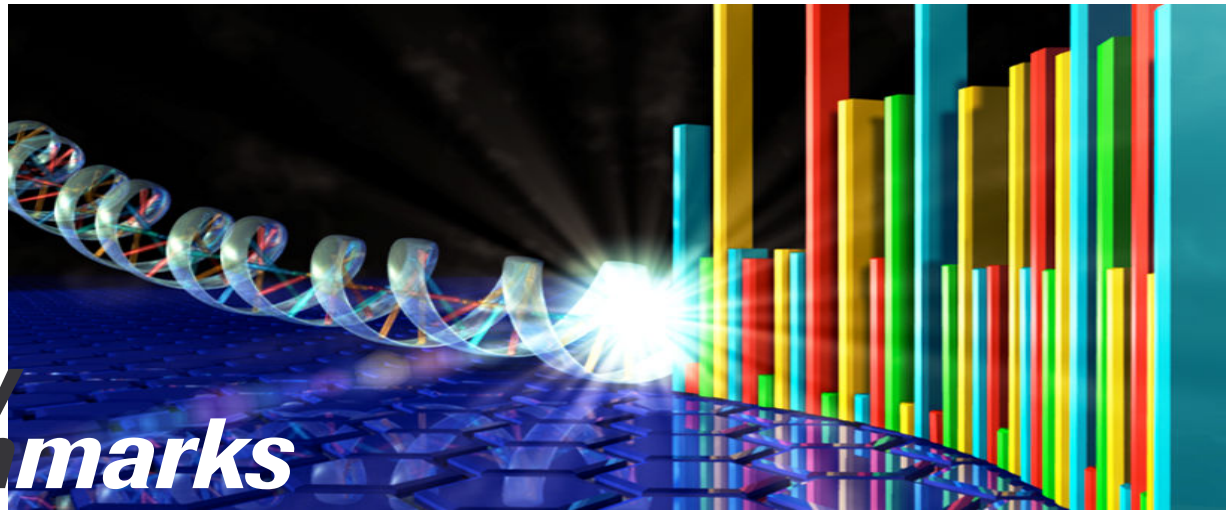
- New right click mouse options
- New Keyboard shortcuts

gsRunBrowser – Tooltips



- New Mouse over tooltips

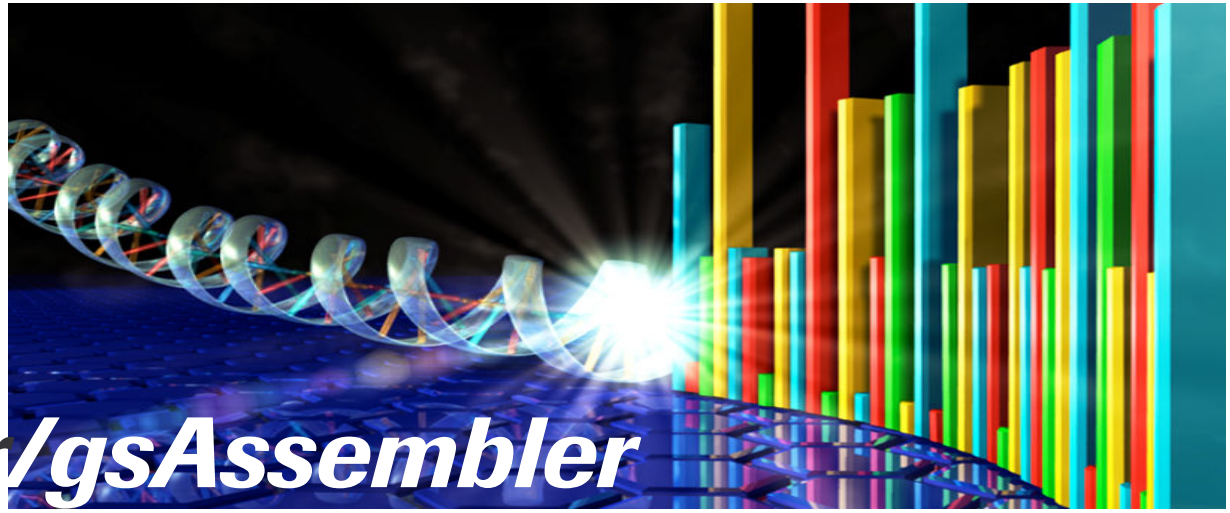
New Benchmarks



New Benchmarks

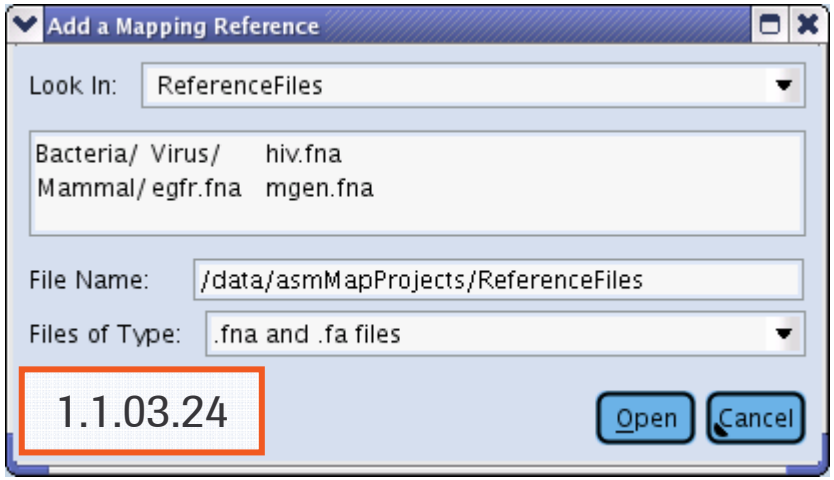
GS Run Processor and Metrics Files

Metrics File	Benchmark	Value
454QualityFilterMetrics.txt	totalKeyPass/totalRawWell	>90%
454QualityFilterMetrics.txt	totalPassFiltering/TotalKeyPass	>65%
454BaseCallerMetrics.txt	Average Read Length	>300bp
454BaseCallerMetrics.txt	Total Bases for 2 region PTP	400 million bases



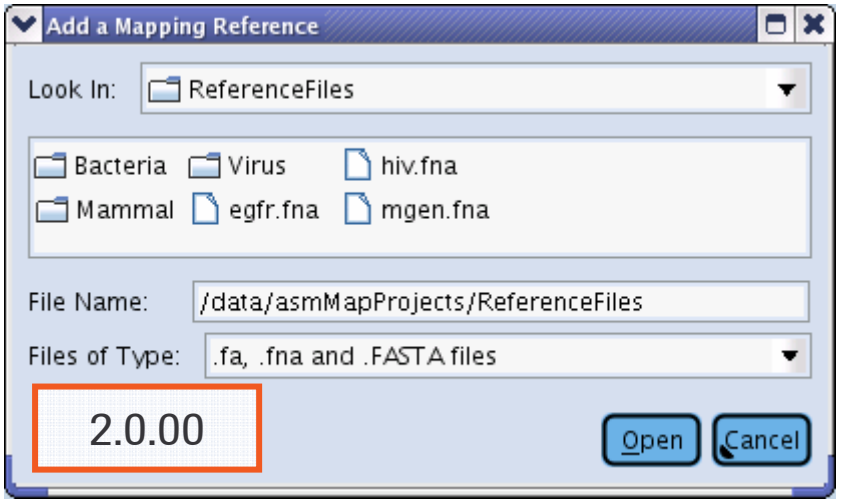
gsMapper/gsAssembler

Icons Now Present in all File Dialogs



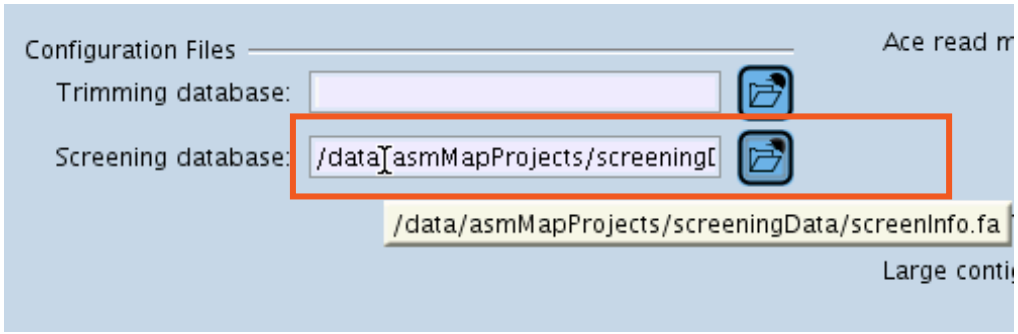
- All FileChooser dialogs have been augmented in 2.0.00 with folder and file icons to make them easier to read and use

- 2.0.00 FileChooser consult historic, user-specific values for both gsMapper and gsAssembler to jump into the last opened directory

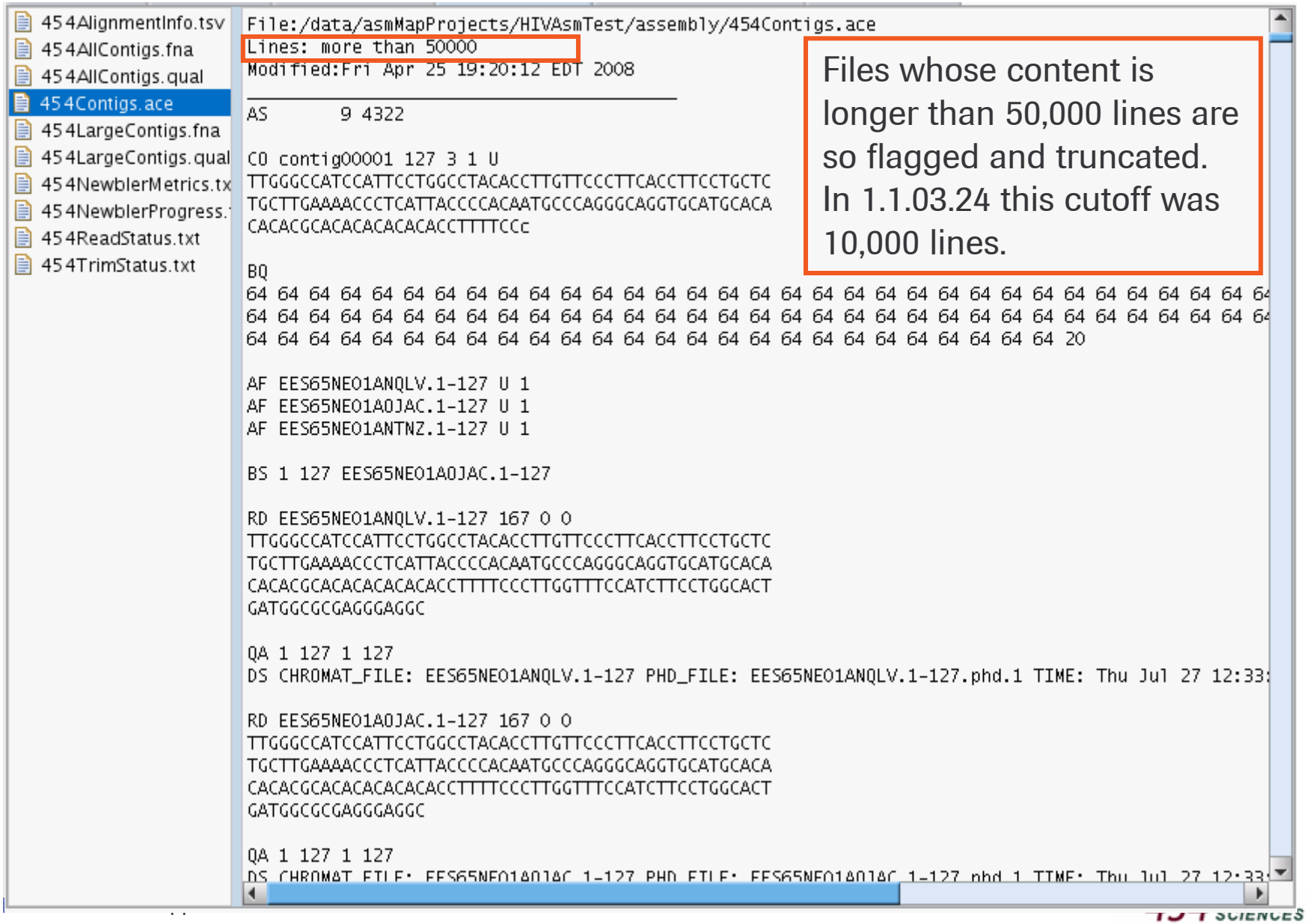


gsAssembler/gsMapper Parameter File Tooltips

- Tooltips have been added to display the full path if you hover over the text area.

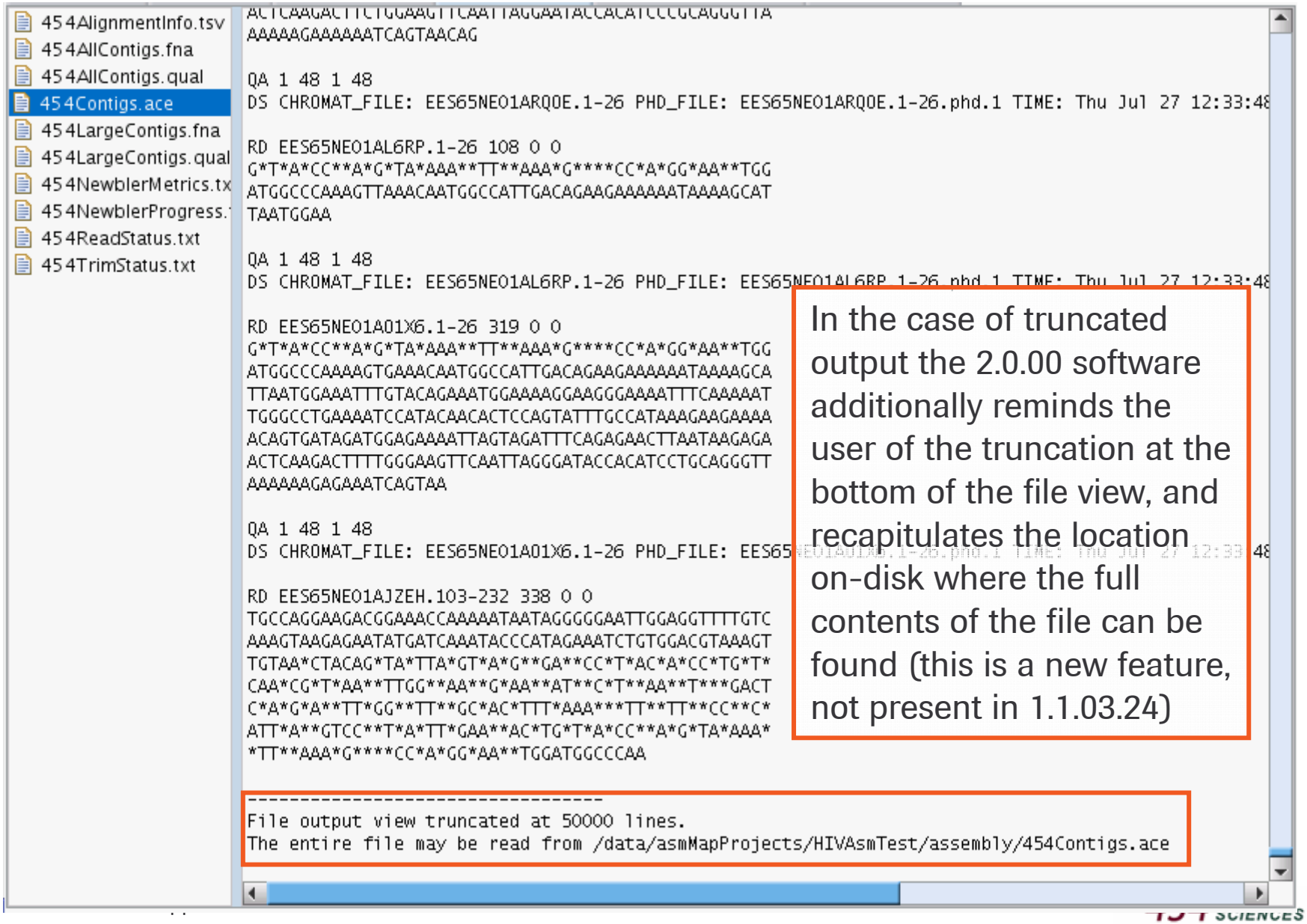


gsAssembler/gsMapper Result File Tab

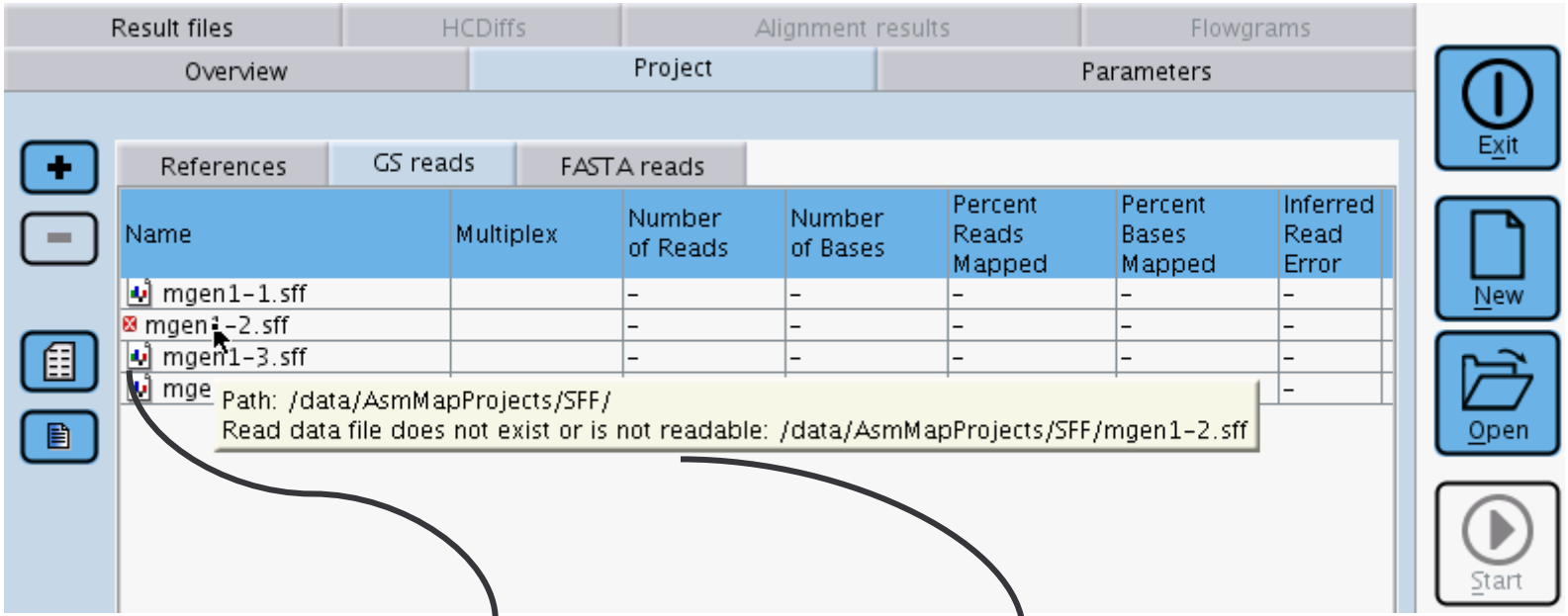


Files whose content is longer than 50,000 lines are so flagged and truncated. In 1.1.03.24 this cutoff was 10,000 lines.

gsAssembler/gsMapper Result File Tab



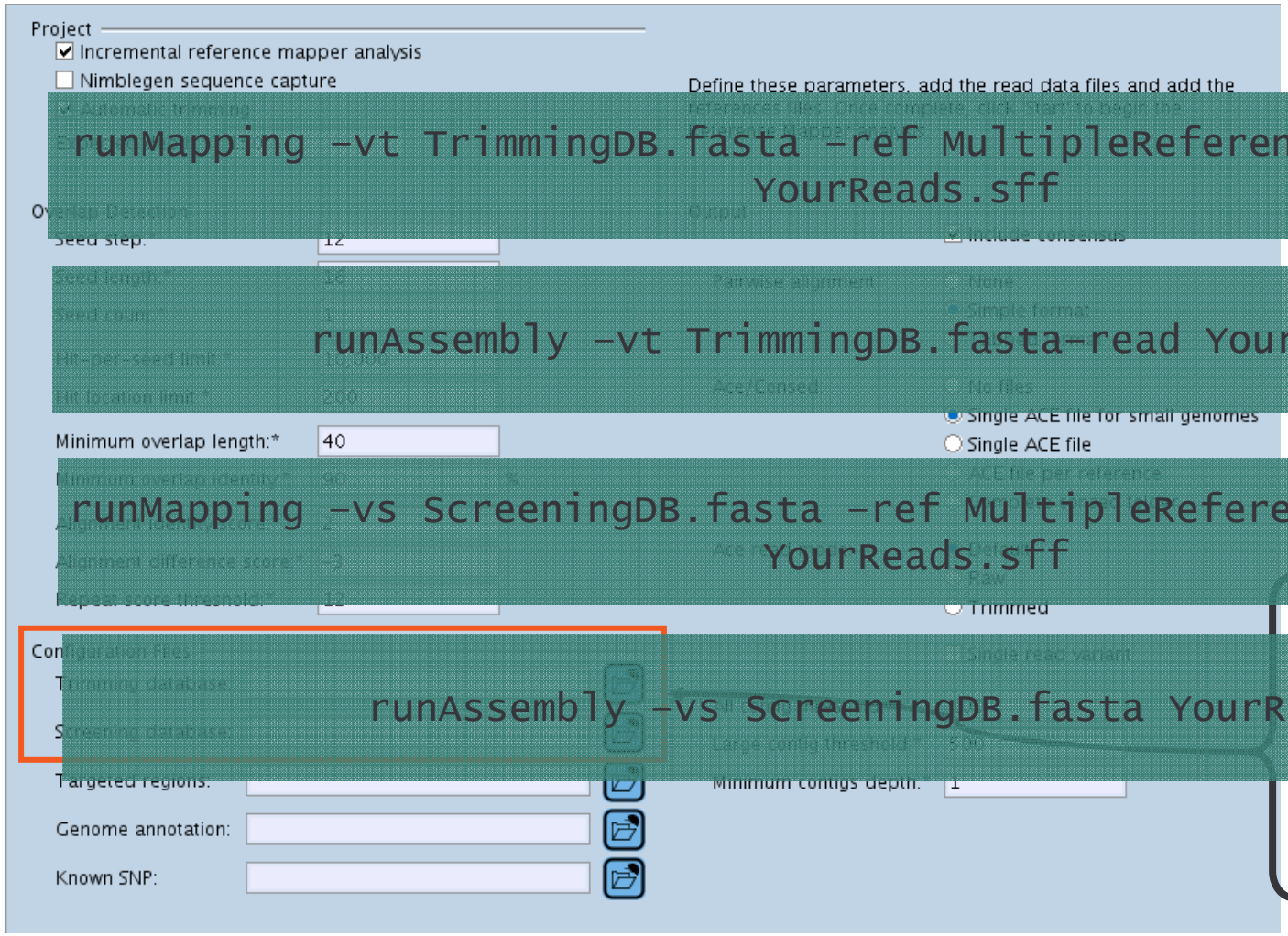
Visible Validation of Project Input Files



- Red “x” appears adjacent to Reference (Mapper), GS read or FASTA read (Mapper + Assembler) files that fail validation

- Mousing over the red “x” or anywhere in the cell with the red “x” provides a tooltip explaining the problem.

Vector Trimming/Screening Database




```
runMapping -vt TrimmingDB.fasta -ref MultipleReferenceFiles.fasta -read
YourReads.sff
```

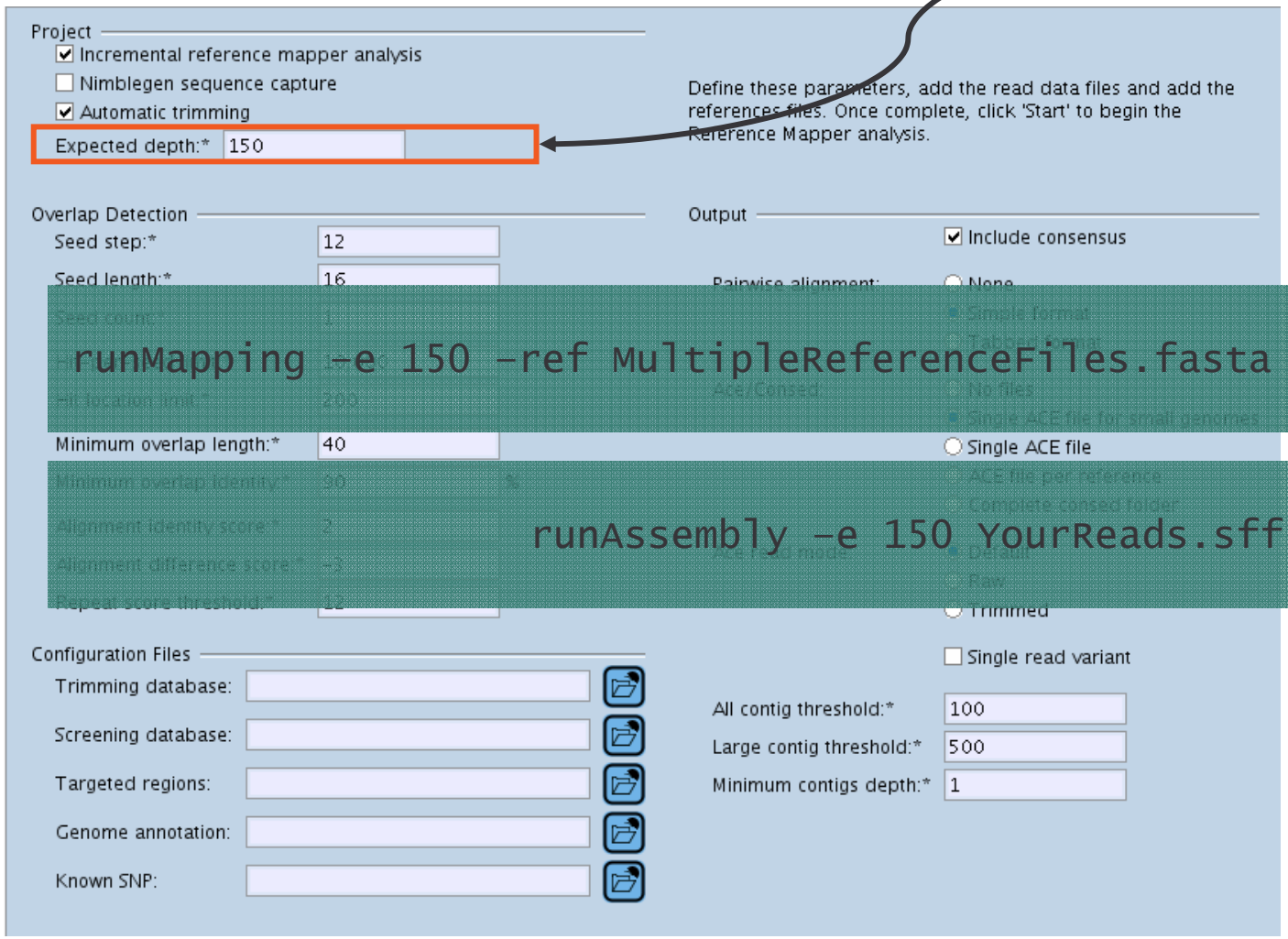
```
runAssembly -vt TrimmingDB.fasta-read YourReads.sff
```

```
runMapping -vs ScreeningDB.fasta -ref MultipleReferenceFiles.fasta -read
YourReads.sff
```

```
runAssembly -vs ScreeningDB.fasta YourReads.sff
```

- 454 reads are trimmed at the ends by the trimming database
- 454 reads are removed by the screening database

Expected Depth

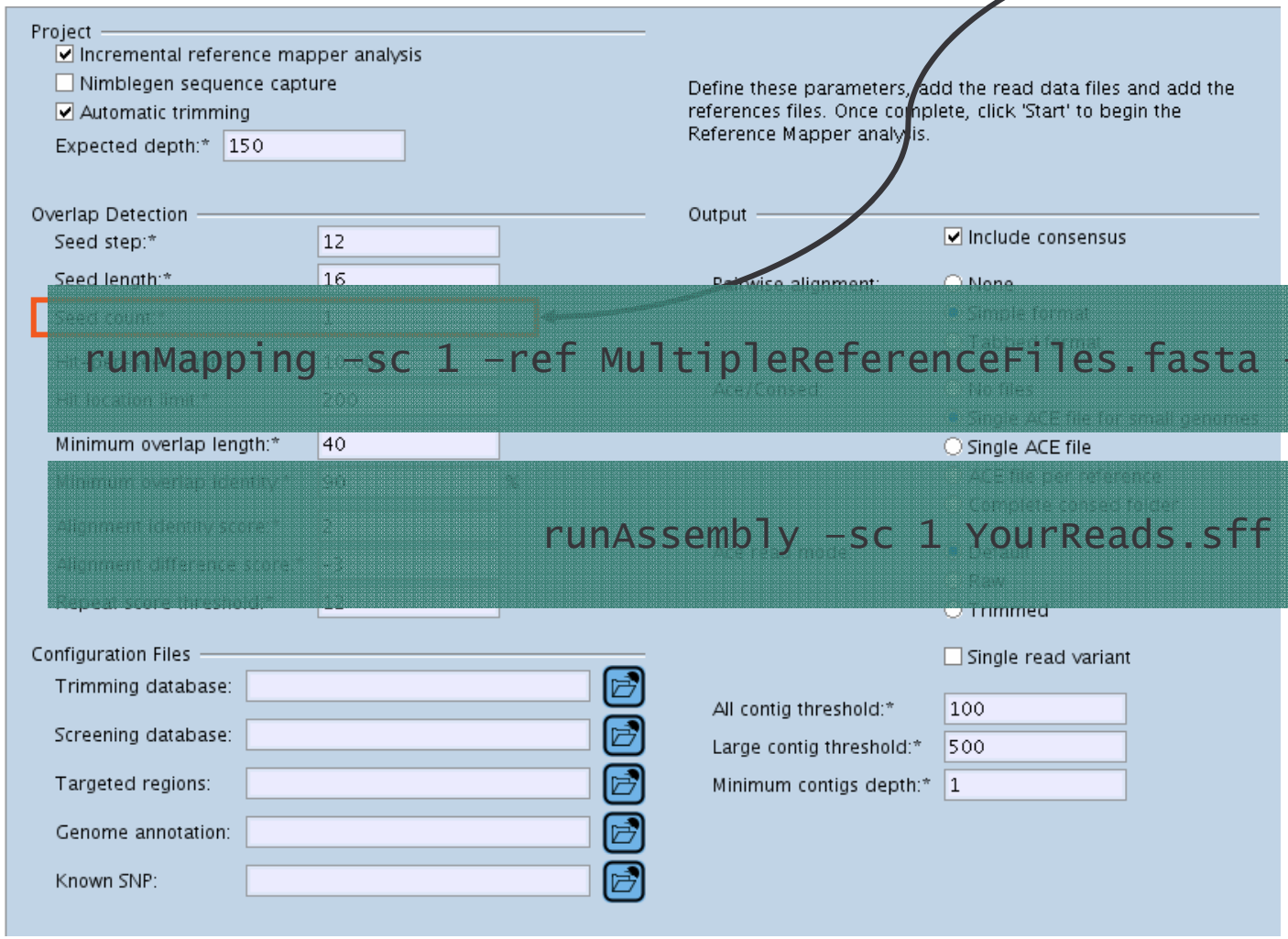



```
runMapping -e 150 -ref MultipleReferenceFiles.fasta -read YourReads.sff
```

```
runAssembly -e 150 YourReads.sff
```

- Algorithm adjustment to filter out random-chance events at higher coverage (for example 150 fold coverage)

Seed Count

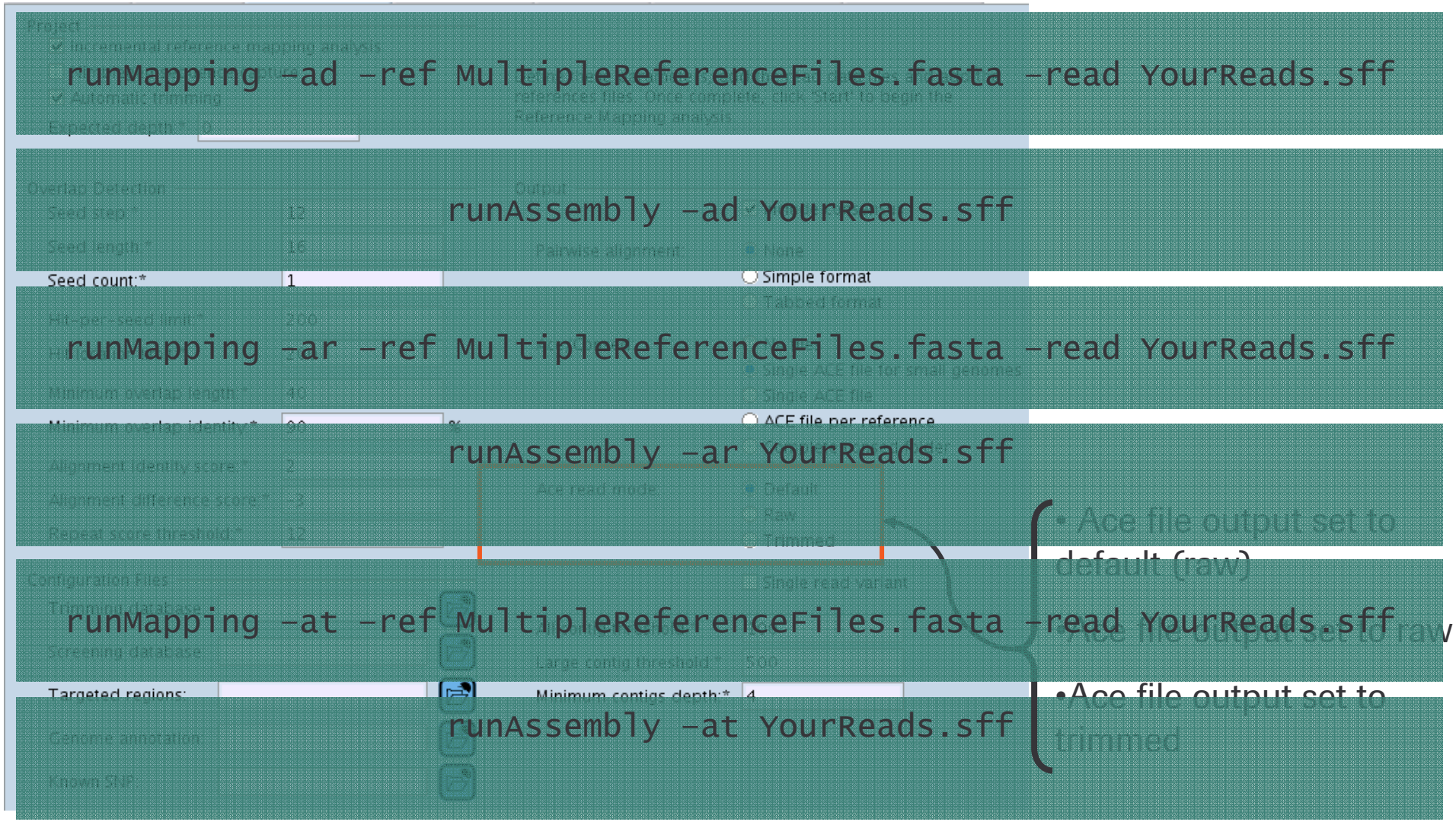



```
runMapping -sc 1 -ref MultipleReferenceFiles.fasta -read YourReads.sff
```

```
runAssembly -sc 1 YourReads.sff
```

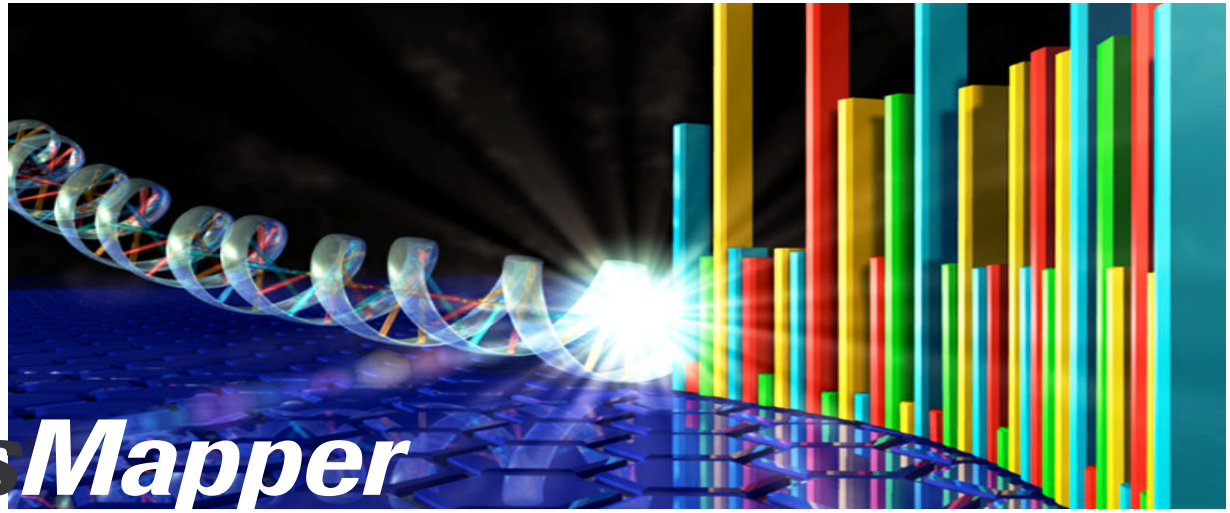
- Number of seeds required in a window before an extension is made

Ace File Read Mode

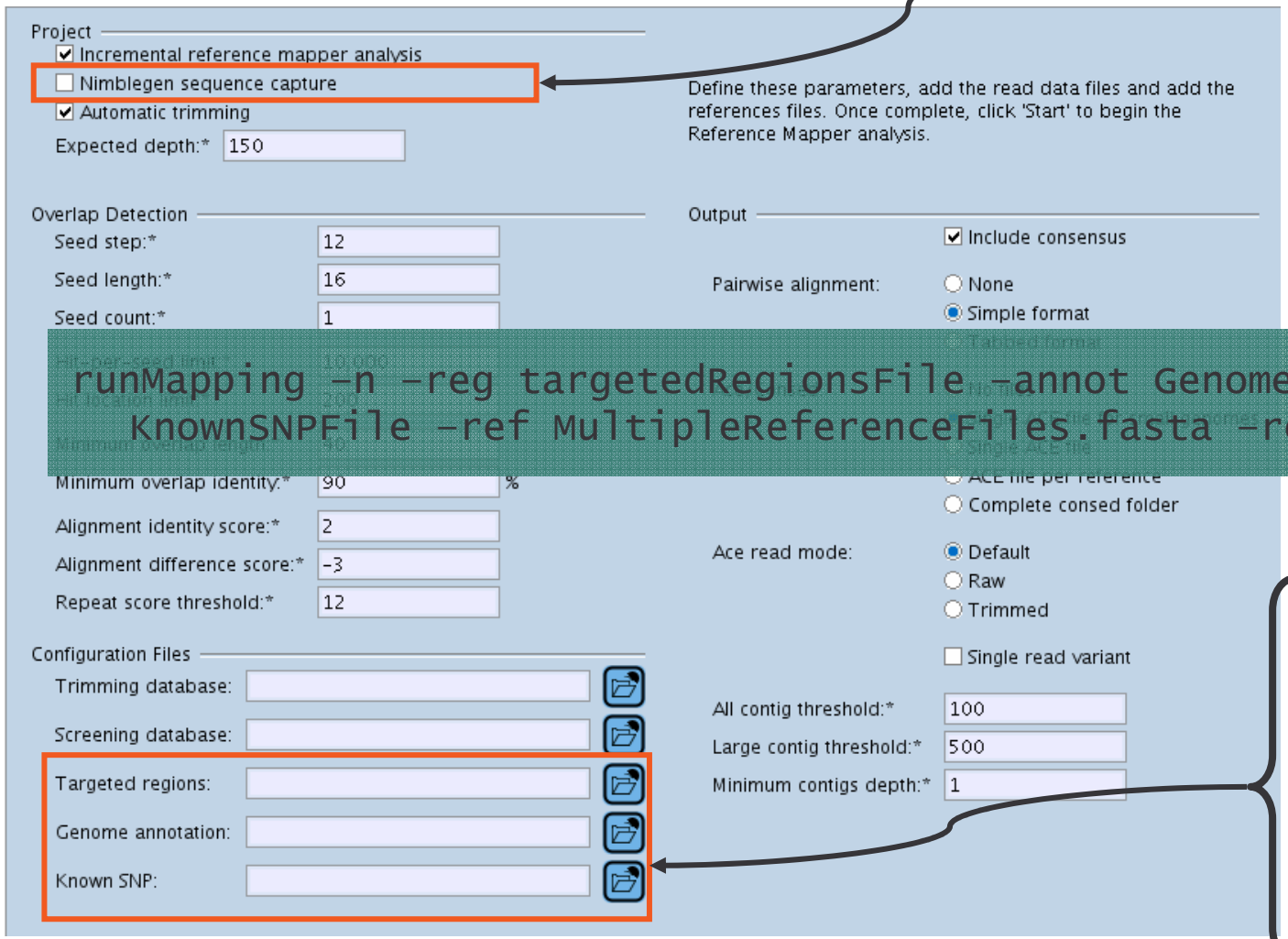


Additional Output Highlights

- Automatic quality score based trimming of 454 reads (`-trim –notrim`)
- `454TrimStatus.txt` reports the trim status of each read
  - "TooShort" is added as an additional read status (trimmed shorter than 50 bases)
- `454PairStatus.txt` is created whenever paired-end reads are part of the dataset
  - One line summary for each pair
- Each paired-end file is now treated as a distinct library in order to compute mean distances separately


gsMapper

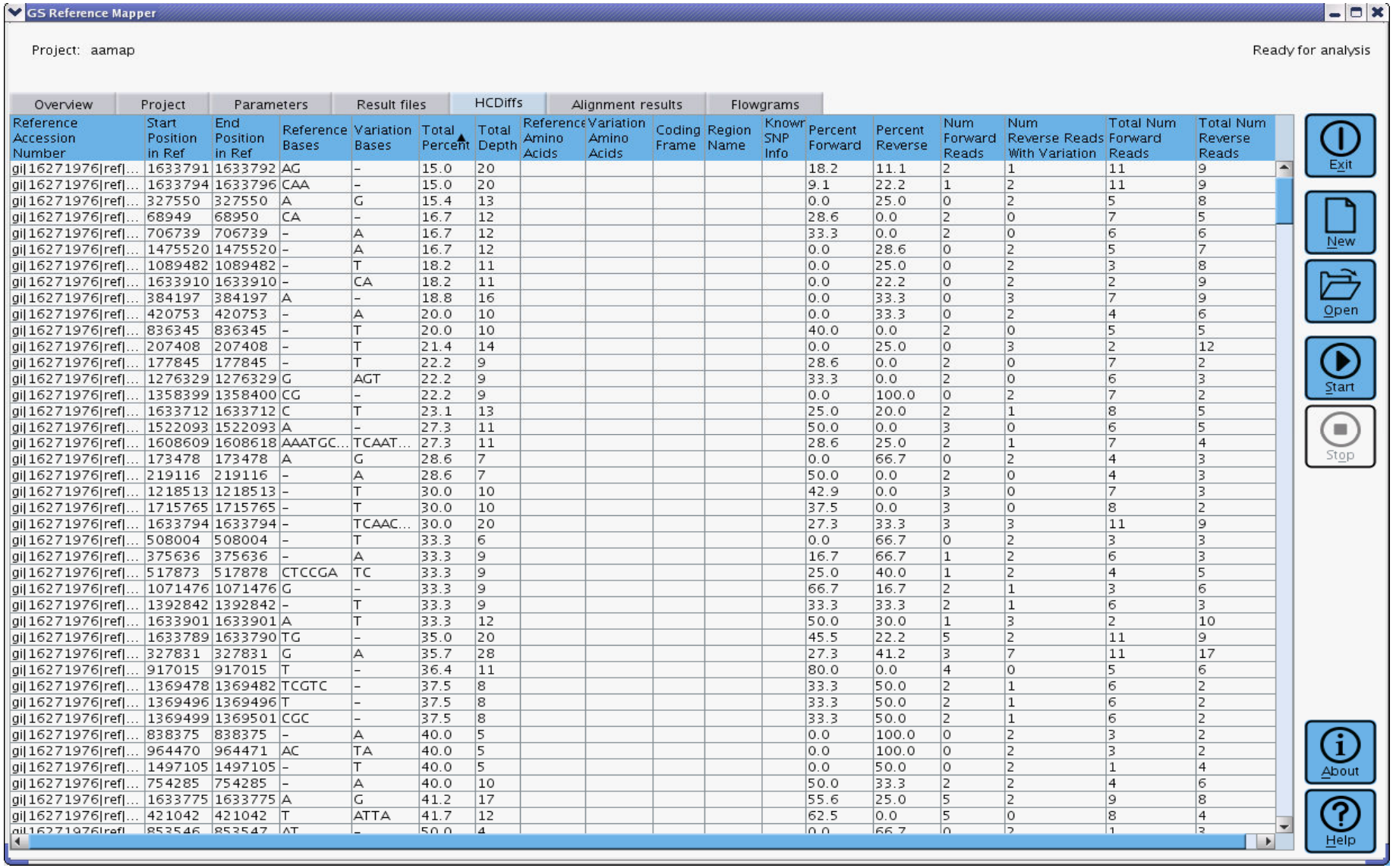
Mapper – Nimblegen sequence capture




```
runMapping -n -reg targetedRegionsFile -annot GenomeAnnotationFile -snp
KnownSNPFile -ref MultipleReferenceFiles.fasta -read YourReads.sff
```

- Nimblegen sequence capture primer is trimmed
- The targeted regions file from Nimblegen
- The Genome annotation file (UCSC GoldenPath)
- The dbSNP file (UCSC GoldenPath)

gsMapper – New HCDiffs Tab

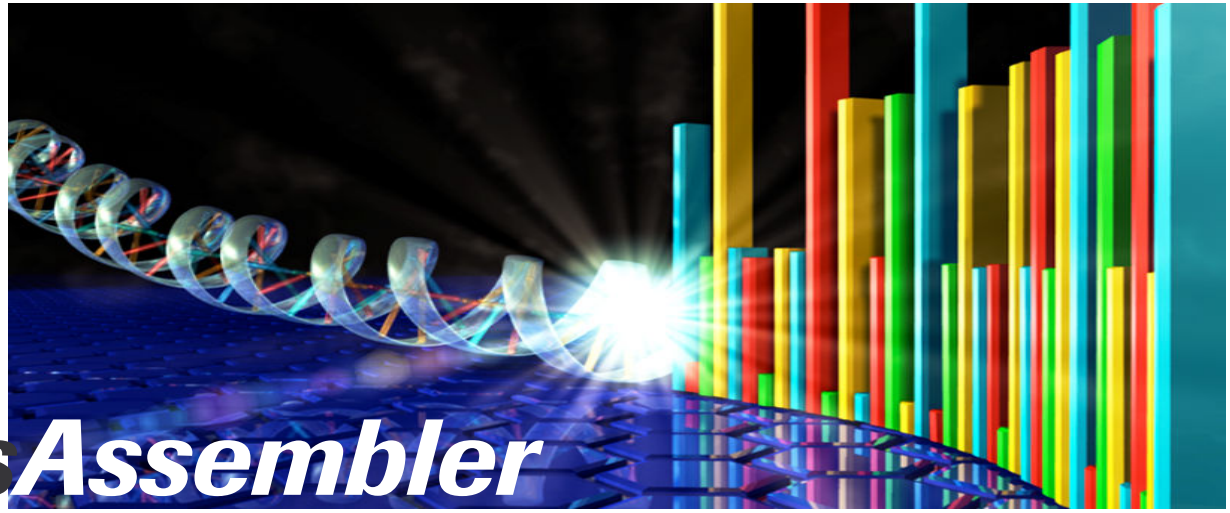


Reference Accession Number	Start Position in Ref	End Position in Ref	Reference Bases	Variation Bases	Total Percent	Total Depth	Reference Amino Acids	Variation Amino Acids	Coding Frame	Region Name	Known SNP Info	Percent Forward	Percent Reverse	Num Forward Reads	Num Reverse Reads With Variation	Total Num Forward Reads	Total Num Reverse Reads
gi|16271976|ref|...	1633791	1633792	AG	-	15.0	20						18.2	11.1	2	1	11	9
gi|16271976|ref|...	1633794	1633796	CAA	-	15.0	20						9.1	22.2	1	2	11	9
gi|16271976|ref|...	327550	327550	A	G	15.4	13						0.0	25.0	0	2	5	8
gi|16271976|ref|...	68949	68950	CA	-	16.7	12						28.6	0.0	2	0	7	5
gi|16271976|ref|...	706739	706739	-	A	16.7	12						33.3	0.0	2	0	6	6
gi|16271976|ref|...	1475520	1475520	-	A	16.7	12						0.0	28.6	0	2	5	7
gi|16271976|ref|...	1089482	1089482	-	T	18.2	11						0.0	25.0	0	2	3	8
gi|16271976|ref|...	1633910	1633910	-	CA	18.2	11						0.0	22.2	0	2	2	9
gi|16271976|ref|...	384197	384197	A	-	18.8	16						0.0	33.3	0	3	7	9
gi|16271976|ref|...	420753	420753	-	A	20.0	10						0.0	33.3	0	2	4	6
gi|16271976|ref|...	836345	836345	-	T	20.0	10						40.0	0.0	2	0	5	5
gi|16271976|ref|...	207408	207408	-	T	21.4	14						0.0	25.0	0	3	2	12
gi|16271976|ref|...	177845	177845	-	T	22.2	9						28.6	0.0	2	0	7	2
gi|16271976|ref|...	1276329	1276329	G	AGT	22.2	9						33.3	0.0	2	0	6	3
gi|16271976|ref|...	1358399	1358400	CG	-	22.2	9						0.0	100.0	0	2	7	2
gi|16271976|ref|...	1633712	1633712	C	T	23.1	13						25.0	20.0	2	1	8	5
gi|16271976|ref|...	1522093	1522093	A	-	27.3	11						50.0	0.0	3	0	6	5
gi|16271976|ref|...	1608609	1608618	AAATGC...	TCAAT...	27.3	11						28.6	25.0	2	1	7	4
gi|16271976|ref|...	173478	173478	A	G	28.6	7						0.0	66.7	0	2	4	3
gi|16271976|ref|...	219116	219116	-	A	28.6	7						50.0	0.0	2	0	4	3
gi|16271976|ref|...	1218513	1218513	-	T	30.0	10						42.9	0.0	3	0	7	3
gi|16271976|ref|...	1715765	1715765	-	T	30.0	10						37.5	0.0	3	0	8	2
gi|16271976|ref|...	1633794	1633794	-	TCAAC...	30.0	20						27.3	33.3	3	3	11	9
gi|16271976|ref|...	508004	508004	-	T	33.3	6						0.0	66.7	0	2	3	3
gi|16271976|ref|...	375636	375636	-	A	33.3	9						16.7	66.7	1	2	6	3
gi|16271976|ref|...	517873	517878	CTCCGA	TC	33.3	9						25.0	40.0	1	2	4	5
gi|16271976|ref|...	1071476	1071476	G	-	33.3	9						66.7	16.7	2	1	3	6
gi|16271976|ref|...	1392842	1392842	-	T	33.3	9						33.3	33.3	2	1	6	3
gi|16271976|ref|...	1633901	1633901	A	T	33.3	12						50.0	30.0	1	3	2	10
gi|16271976|ref|...	1633789	1633790	TG	-	35.0	20						45.5	22.2	5	2	11	9
gi|16271976|ref|...	327831	327831	G	A	35.7	28						27.3	41.2	3	7	11	17
gi|16271976|ref|...	917015	917015	T	-	36.4	11						80.0	0.0	4	0	5	6
gi|16271976|ref|...	1369478	1369482	TCGTC	-	37.5	8						33.3	50.0	2	1	6	2
gi|16271976|ref|...	1369496	1369496	T	-	37.5	8						33.3	50.0	2	1	6	2
gi|16271976|ref|...	1369499	1369501	CGC	-	37.5	8						33.3	50.0	2	1	6	2
gi|16271976|ref|...	838375	838375	-	A	40.0	5						0.0	100.0	0	2	3	2
gi|16271976|ref|...	964470	964471	AC	TA	40.0	5						0.0	100.0	0	2	3	2
gi|16271976|ref|...	1497105	1497105	-	T	40.0	5						0.0	50.0	0	2	1	4
gi|16271976|ref|...	754285	754285	-	A	40.0	10						50.0	33.3	2	2	4	6
gi|16271976|ref|...	1633775	1633775	A	G	41.2	17						55.6	25.0	5	2	9	8
gi|16271976|ref|...	421042	421042	T	ATTA	41.7	12						62.5	0.0	5	0	8	4
gi|16271976|ref|	853546	853547	AT	-	50.0	4						0.0	66.7	0	2	1	3

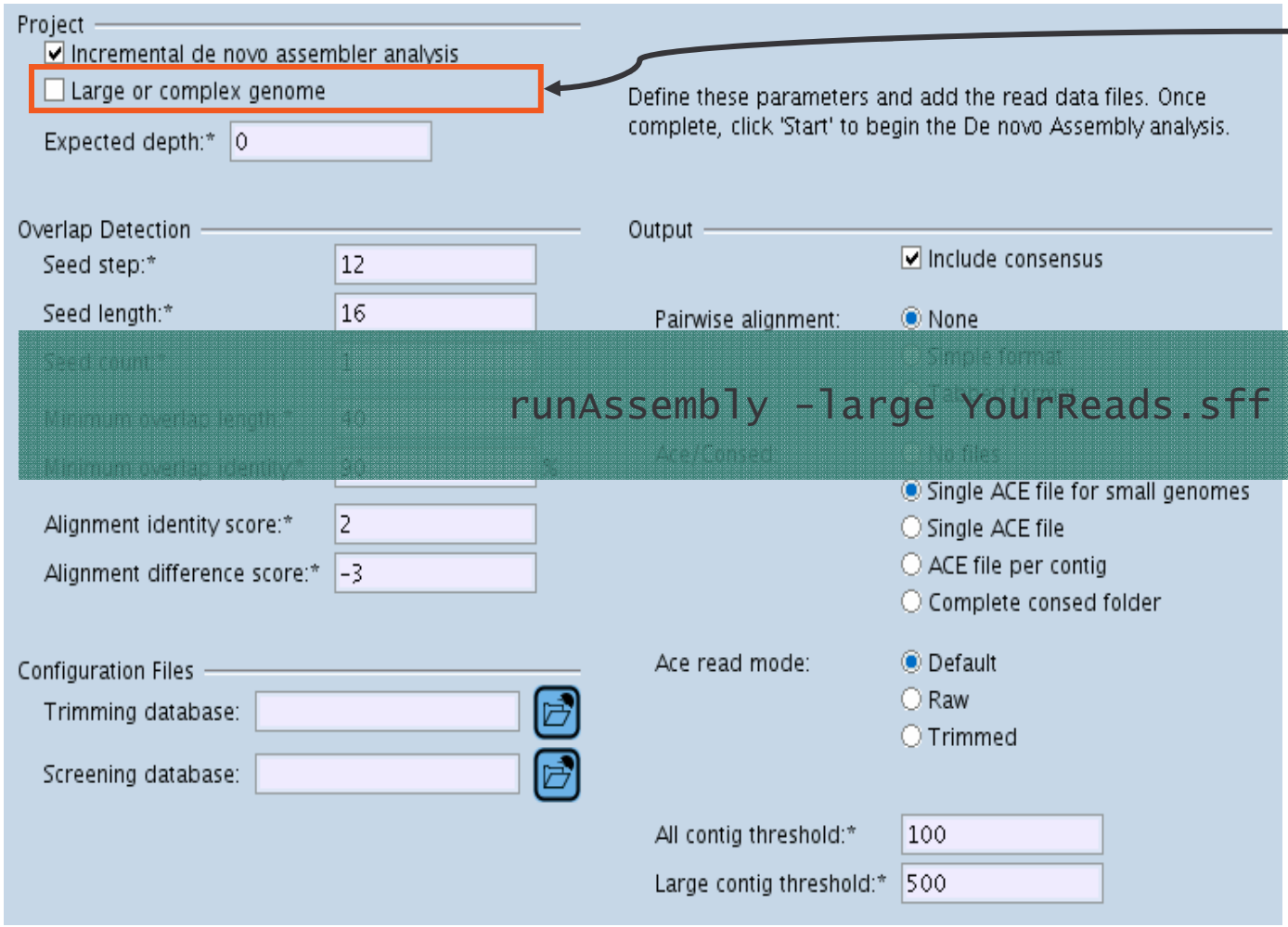
gsMapper – HCDiffs Tab Capabilities

- Parsed Contents of HCDiffs file with live link to Alignment Results
  - Right-click on any row and select the “Show Alignments” menu item to go directly to the given HCDiff variation in the “Alignment results” Tab view
  - The first base of the variation will appear leftmost in the view.

- Sortable Table
  - Click on header to sort by contents of column.
  - Click again on same header to reverse the sort (triangular indicator specifies which column was last used for sort operation and the ascending/descending direction of the sort)
  - Sort is “stable”: Subsequent sorts that have tied values will not re-order previously sorted rows, allowing the grouping of data via nested sorting operations by sorting different columns in succession.


gsAssembler

Assembler – Large option



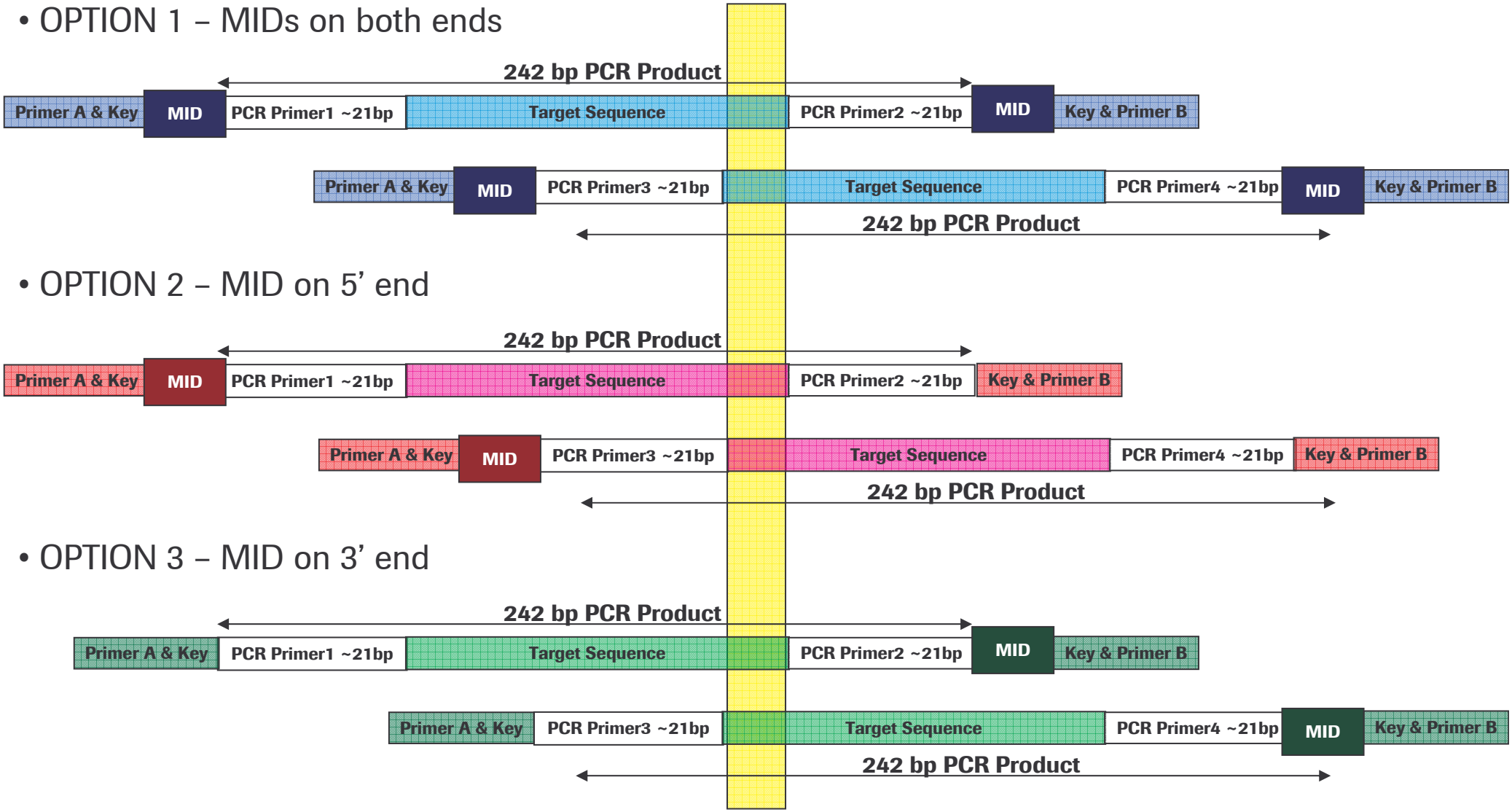

```
runAssembly -large YourReads.sff
```

- Algorithm adjustment to assemble large and/or complex genomes

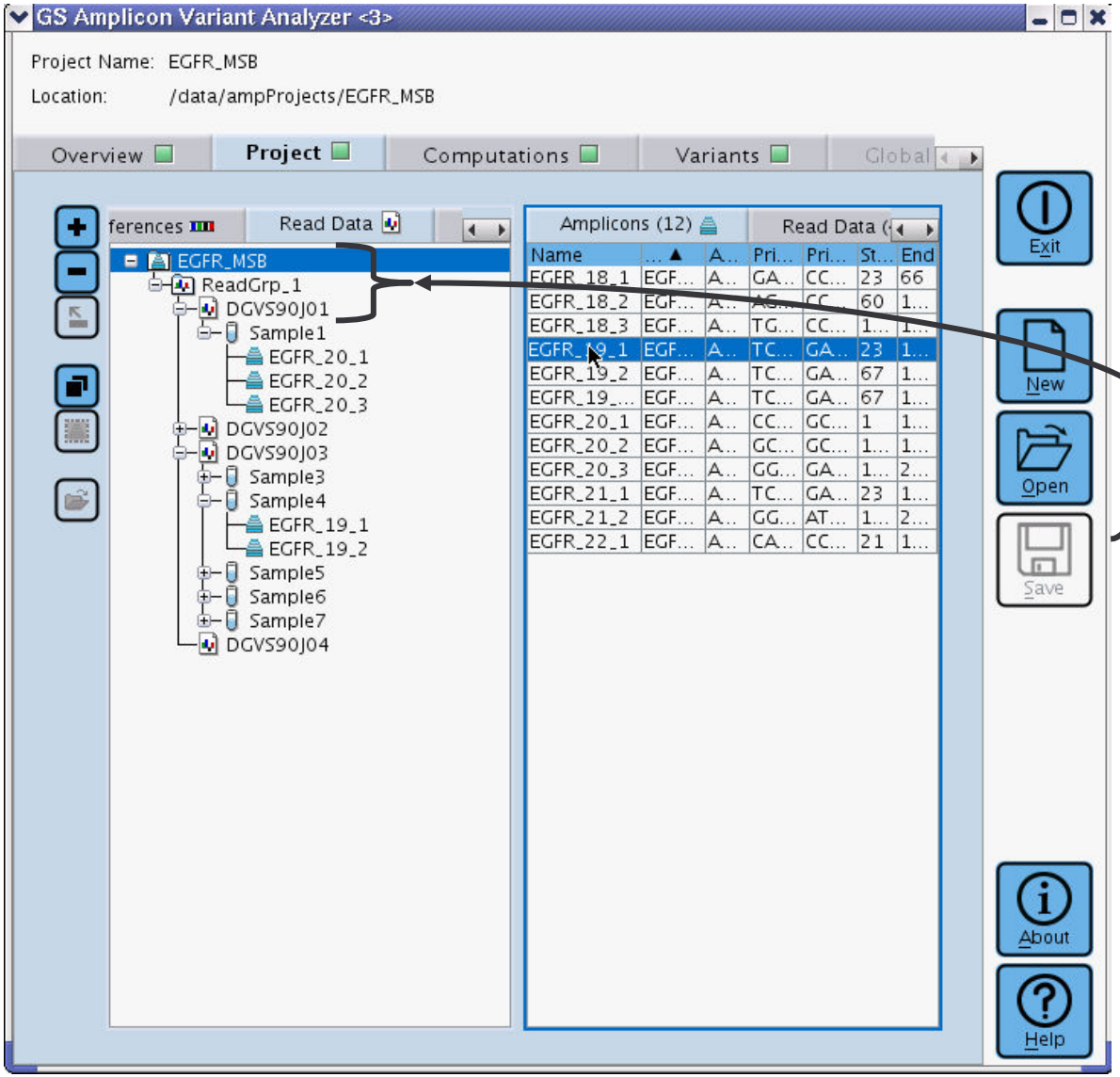
Amplicon Variant Analyzer

PCR product design options using MIDs

- OPTION 1 – MIDs on both ends
- OPTION 2 – MID on 5’ end
- OPTION 3 – MID on 3’ end



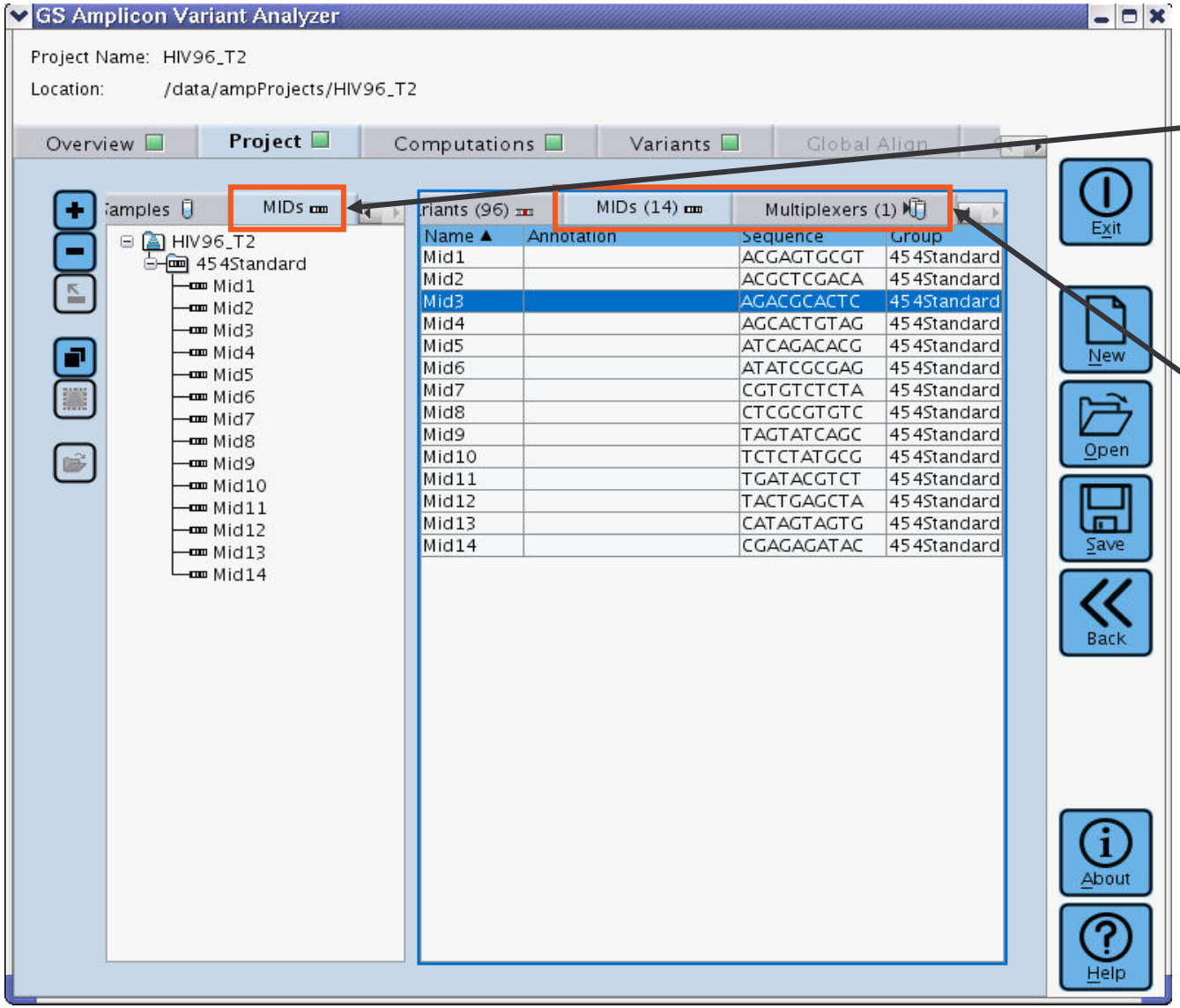
Project Tab: Dragging Amplicons To Read Data Tree



- Amplicons may now be dragged to the Project (root) node, any Read Data Group, or any Read Data node

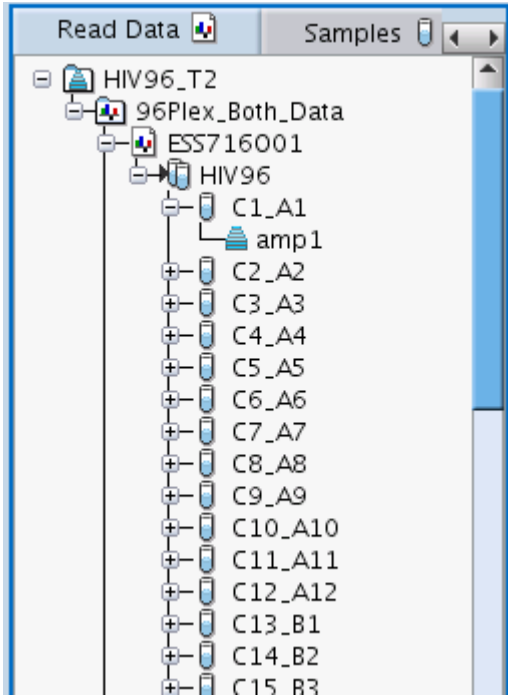
- The drop will only be accepted if there is at least one Read Data below the level of the drag that has at most one Sample or Multiplexer currently associated with it

Multiplexing Features – New Tree & Tables



- "MIDs" Tree enumerating the MIDs of the project in relationship to their (optional) MID Group
- The Table section contains two new tables for defining the MIDs and Multiplexers of the Project

Multiplexing Features – Tree Nodes

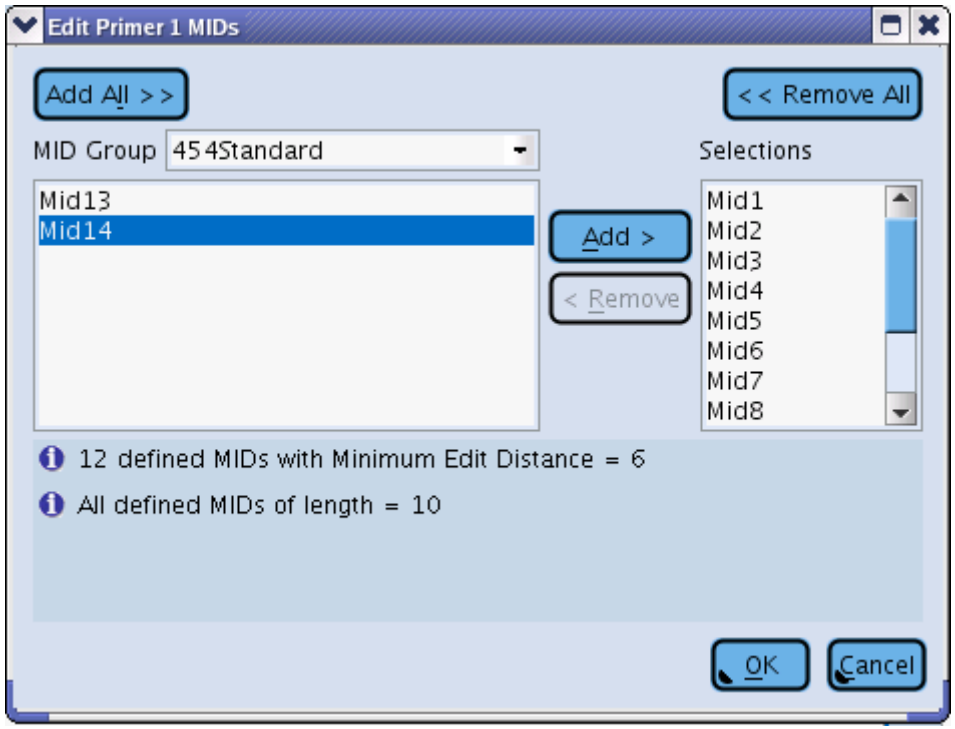


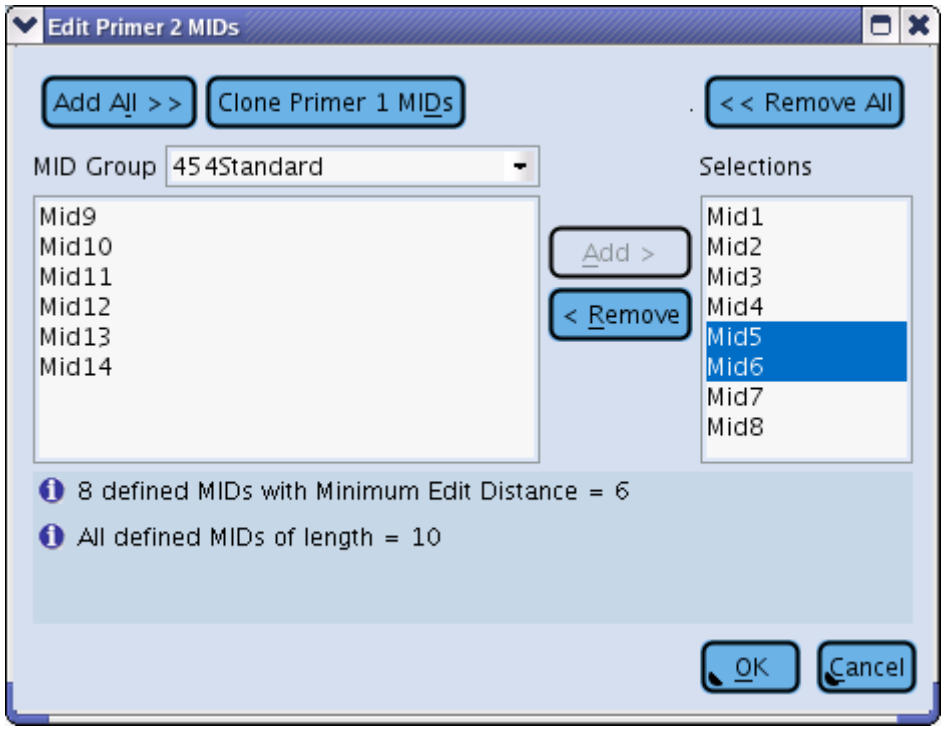
- For Read Data sets that contain MID-based multiplexed data, a new node representing the Multiplexer will appear in the Read Data Tree of the Project Tab: 

- Under this Multiplexer node will be the Samples encoded by MIDs of that Multiplexer, with the associated Amplicons appearing below those Samples.

Multiplexing Features – MID Association Editors

- By double clicking on the Primer1 MID and Primer2 MID columns one can bring up editors for associating previously defined MIDs with the Primer1 and Primer2 positions of the Multiplexer:



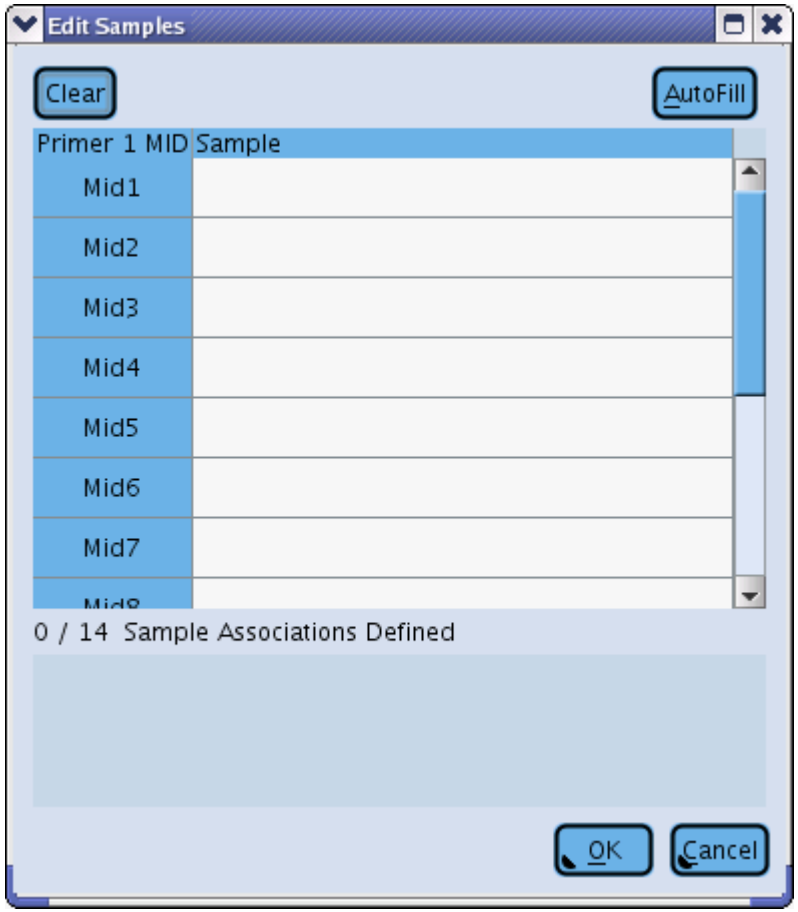


- These editors provide informative status messages regarding the mutual compatibility of the selected MIDs for use during multiplexing.

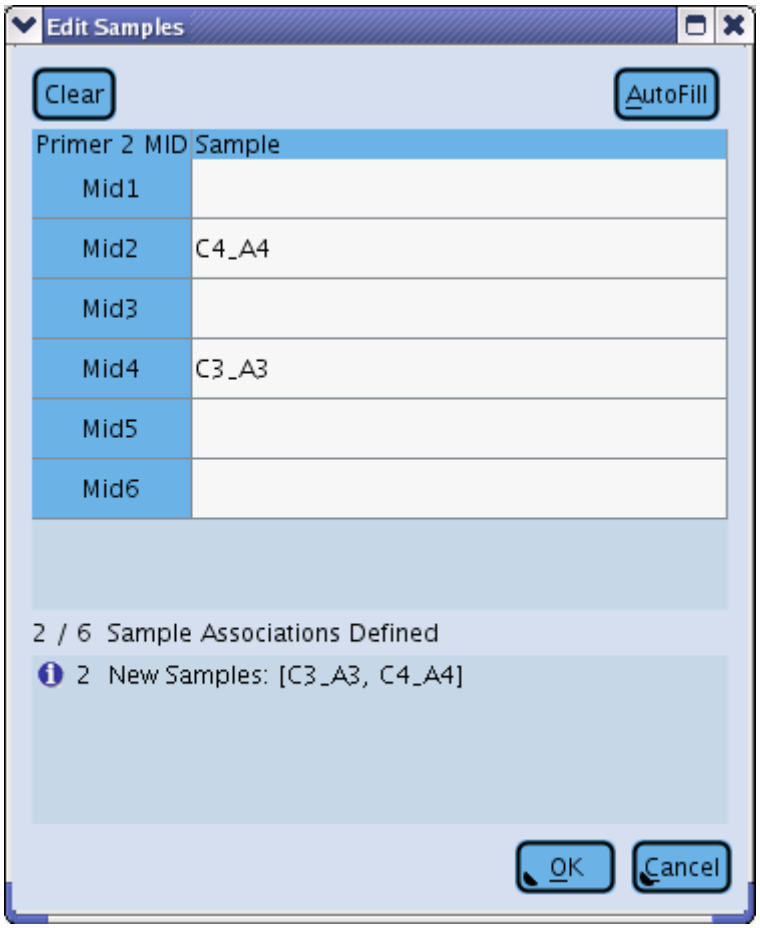
Multiplexing Features – Sample Editors

- By double clicking on the Samples column of the Multiplexer table one can define the relationship between combinations of Primer1 and Primer2 MIDs and Samples.
- Different editors are available depending on the type of Multiplexer Encoding, which is assignable via a dropdown editor on the Encoding column of the Multiplexer table, and will be one of the values:
  - **`Both`**: both MIDs are used to determine the Sample for each Read.
  - **`Either`**: there are MIDs at both ends of the sequence, but *either* one is sufficient to determine the Sample, with the MID in at the 5' end, in the read's orientation, being used for a given read
  - **`Primer1`**: there is only a MID sequence at the Primer1 end of the Read
  - **`Primer2`**: there is only a MID sequence at the Primer2 end of the Read.
- All editors provide informative status messages regarding any possible incompatibilities of the Multiplexer's associated MIDs, any incompletely specified or duplicate Sample mappings, or Samples that will be created new because their names don't match any Samples currently in the project.

Multiplexing Features – Sample Editors

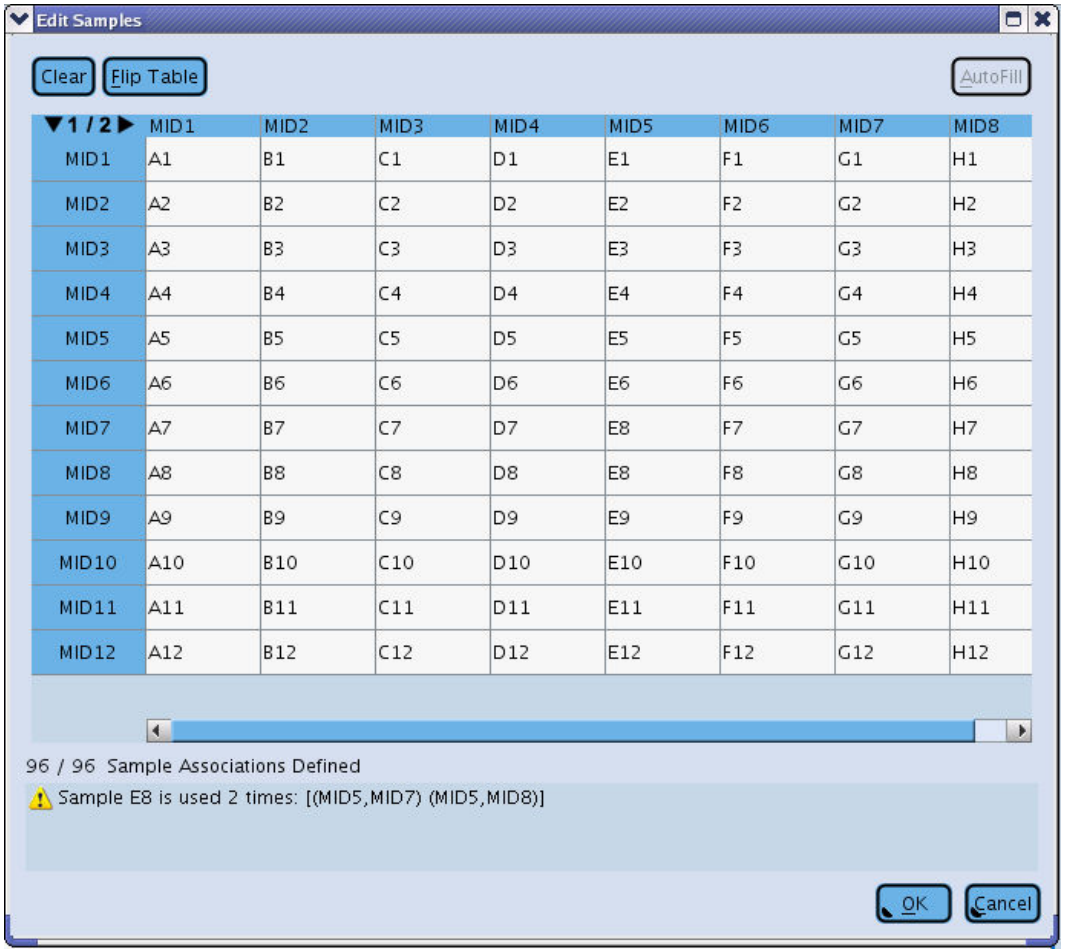


Primer1 MID Encoding Sample Editor

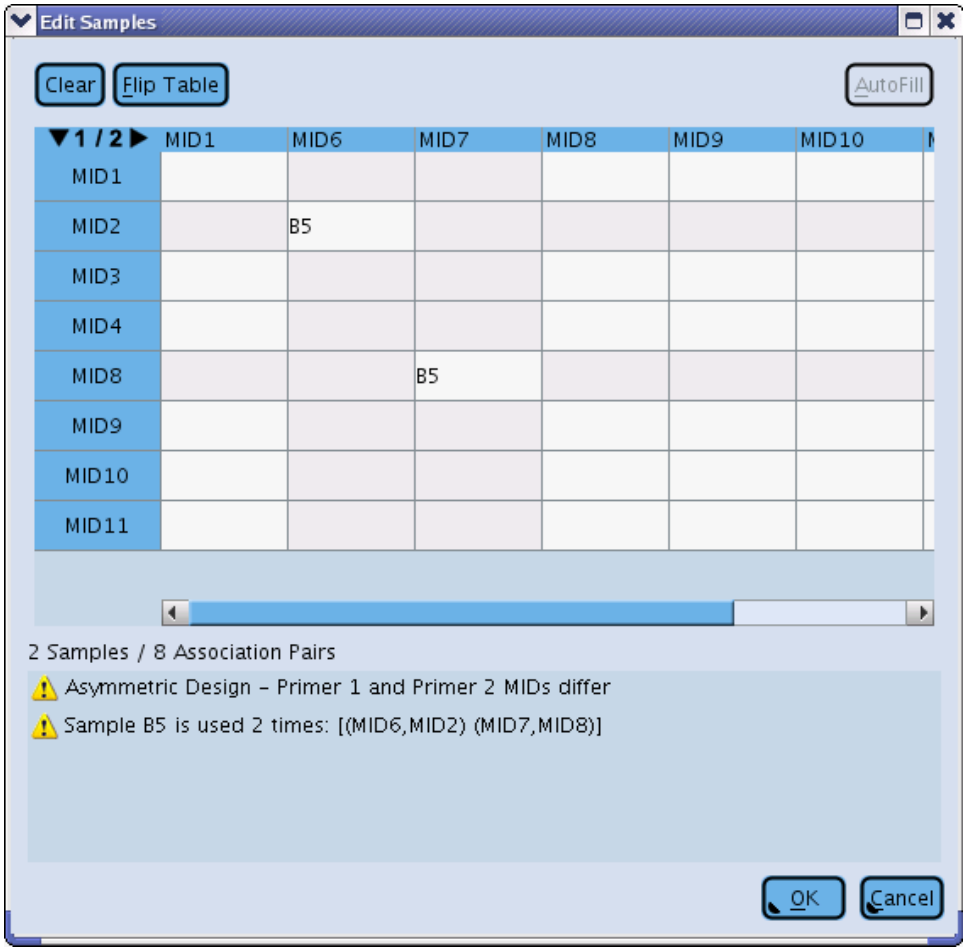


Primer2 MID Encoding Sample Editor

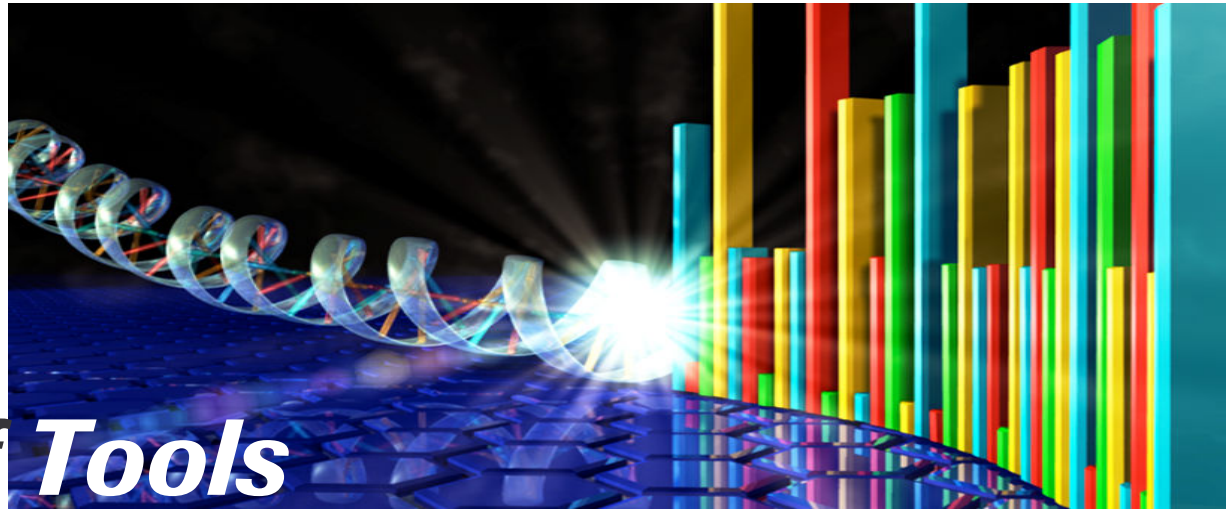
Multiplexing Features – Sample Editors



"Both" Encoding Sample Editor



"Either" Encoding Sample Editor



sff Tools

SFF Tools

- `sffcall` has been removed as it was only applicable to runs processed with 1.0.51 earlier
- `sffvolume` has been removed since the NCBI Trace Archive has shifted to the new Short Read Archive
- `sffrescore` has been added to rewrite SFF files with the new quality scores
- `sfffile` command now contains options to randomly pick a subset of reads (`-pick #`) or bases (`-pick #`)

Please contact Roche Applied Science Technical Support via:	
phone: 1-800-262-4911	e-mail: us.gssupport@roche.com

gsMapper Project Tab / References / Accession Index