USER GUIDE

applied biosystems®
by life technologies™

SeqScape® Software 3

Publication Part Number 4474242 Rev. A
Revision Date February 2012

life technologies™

For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.

Information in this document is subject to change without notice.

LIFE TECHNOLOGIES AND/OR ITS AFFILIATE(S) DISCLAIM ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THOSE OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LIFE TECHNOLOGIES AND/OR ITS AFFILIATE(S) BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, OR UNDER ANY STATUTE OR ON ANY OTHER BASIS FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING BUT NOT LIMITED TO THE USE THEREOF.

NOTICE TO PURCHASER: DISCLAIMER OF LICENSE

The purchase of this product conveys to the purchaser the limited, non-transferable right to use the purchased amount of the product only to perform internal research for the sole benefit of the purchaser. No right to resell this product or any of its components is conveyed expressly, by implication, or by estoppel. This product is for internal research purposes only and is not for use in commercial applications of any kind, including, without limitation, quality control and commercial services such as reporting the results of purchaser's activities for a fee or other form of consideration. For information on obtaining additional rights, please contact outlicensing@lifetech.com or Out Licensing, Life Technologies, 5791 Van Allen Way, Carlsbad, California 92008.

TRADEMARKS

The trademarks mentioned herein are the property of Life Technologies Corporation and/or its affiliate(s) or their respective owners.

Microsoft Windows is a registered trademark of Microsoft Corporation.

Contents

Preface

Chapter 1 Introduction to SeqScape® Software

Chapter 2 Getting Started with SeqScape® Software

Chapter 7 Viewing the Results

Preface

How to Use This Guide

Purpose of This Guide

The Applied Biosystems® *SeqScape® Software 3 User Guide* provides step-by-step instructions to use this software.

Audience

This guide is intended for novice and experienced analysts and scientists who are doing resequencing.

Assumptions

This manual uses conventions and terminology that assume a working knowledge of the Windows® operating system, the Internet, and Web-based browsers.

What You Should Know Before Getting Started

To make the best use of SeqScape® Software 3 and the documentation, be sure you are familiar with:

- Microsoft® Windows 7 Professional, SP1(32-bit) operating system
- The Internet and Web browser terminology
- DNA sequence detection and analysis methods
- DNA and amino acid coding conventions

Conventions Used in This Guide

Text Conventions

This guide uses the following text conventions:

- **Bold** indicates user action. For example:
  Type **0**, then press **Enter** for each of the remaining fields.
- Titles of documents and CDs are shown in italics. For example:
  SeqScape® Software Version 3 User Guide
- *Italic* text indicates new or important words and is also used for emphasis.
- A right arrow bracket (>) separates successive commands you select from a drop-down or shortcut menu. For example:
  Select **File > Open Project**.
  Right-click the sample row, then select **View Filter > View All Runs**.

File Naming Convention

Spaces and some alphanumeric characters are not valid for user names or file names. The characters that are illegal are:

\ / : * ? “ < > |

User Attention Words

Two user attention words appear in Applied Biosystems user documentation. Each word implies a particular level of observation or action as described below:

Note: Provides information that may be of interest or help but is not critical to the use of the product.

IMPORTANT! Provides information that is necessary for proper software operation.

Examples of the user attention words appear below:

Note: Names for Reference Segments are not editable.

IMPORTANT! Do not click OK until you have completed the RDG.

Safety Conventions

Safety Alert Words

Four safety alert words appear in Applied Biosystems user documentation at points in the document where you need to be aware of relevant hazards. Each alert word implies a particular level of observation or action, as defined below:

Definitions

IMPORTANT! – Indicates information that is necessary for proper instrument operation, accurate chemistry kit use, or safe use of a chemical.

– Indicates a potentially hazardous situation that, if not avoided, may result in minor or moderate injury. It may also be used to alert against unsafe practices.

– Indicates a potentially hazardous situation that, if not avoided, could result in death or serious injury.

– Indicates an imminently hazardous situation that, if not avoided, will result in death or serious injury. This signal word is to be limited to the most extreme situations.

Examples

The following examples show the use of IMPORTANT, CAUTION, and WARNING safety alert words:

IMPORTANT! The sample name, run folder name, and path name, *combined*, can contain no more than 250 characters.

MUSCULOSKELETAL AND REPETITIVE MOTION HAZARD. These hazards are caused by potential risk factors that include but are not limited to repetitive motion, awkward posture, forceful exertion, holding static unhealthy positions, contact pressure, and other workstation environmental factors.

Do not attempt to lift or move the computer or the monitor without the assistance of others. Depending on the weight of the computer and/or the monitor, moving them may require two or more people.

Workstation Safety

Correct ergonomic configuration of your workstation can reduce or prevent effects such as fatigue, pain, and strain. Minimize or eliminate these effects by configuring your workstation to promote neutral or relaxed working positions.

CAUTION **MUSCULOSKELETAL AND REPETITIVE MOTION HAZARD**. These hazards are caused by potential risk factors that include but are not limited to repetitive motion, awkward posture, forceful exertion, holding static unhealthy positions, contact pressure, and other workstation environmental factors.

To minimize musculoskeletal and repetitive motion risks:

- Use equipment that comfortably supports you in neutral working positions and allows adequate accessibility to the keyboard, monitor, and mouse.
- Position the keyboard, mouse, and monitor to promote relaxed body and head postures.

How to Obtain More Information

Related Documentation

The following related documents are shipped with the software:

- SeqScape® Online Help – Provides procedures for common tasks. Help is available from the Help menu in the main SeqScape® window, or by pressing F1.
- *SeqScape® Software Tutorial*
- *SeqScape® Software Quick Reference Card*
- *SeqScape® Frequently Asked Questions*

Portable document format (PDF) versions of the documents listed above are available on the SeqScape® software installation CD or by selecting **Programs** > **Applied Biosystems** > **SeqScape**. If you do not have Acrobat Reader installed on your computer, install it from the SeqScape® software CD so you can open the pdf files.

Note: For additional documentation, see "How to Obtain Support" on page xv.

How to Obtain Support

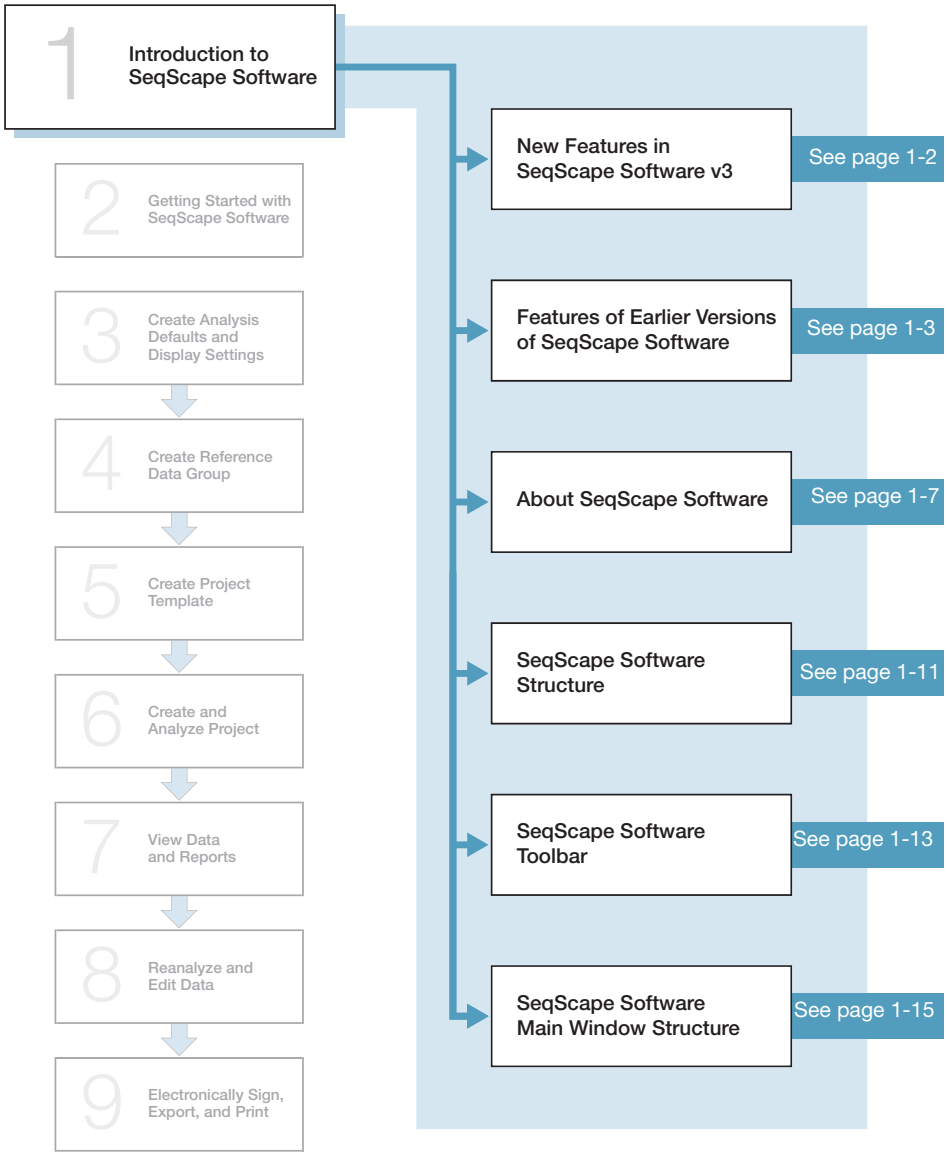
For the latest support information, go to:
lifetechnologies.com/support.

At the Support page, you can:

- Search through frequently asked questions (FAQs)
- Submit a question directly to Technical Support
- Order Life Technologies user documents, MSDSs, certificates of analysis, and other related documents
- Download PDF documents
- Obtain information about customer training
- Download software updates and patches

In addition, the Support page provides access to worldwide telephone and fax numbers to contact Life Technologies Technical Support and Sales facilities.

1 Introduction to SeqScape® Software



New Features in SeqScape® Software 3

SeqScape® Software 3 includes the following new features and enhancements:

- Runs on the Windows 7 operating system.
- KB™ Basecaller 1.4.1 supports basecalling data generated on the two new 3500 POP6BDTv1.1 run modules found in 3500 Data Collection v2.0. RapidSeq50_POP6 at 19.6 kV generates sequences of 450 bases or longer in 65 minutes or less, and FastSeq50_POP6 at 16.9 kV generates 600 bases in 90 minutes or less.
- KB™ Basecaller v1.4.1 supports basecalling data generated with the BigDye® Direct Cycle Sequencing kit with shortened workflow and improved 5' resolution on POP-7™ on the 3500/3500xl, 3730/3730xl, and the 3130/3130xl.
- Supports .ab1 data files generated from 3500/3500xl Genetic Analyzer Data Collection v1.0 and v2.0, 3130/3130xl Genetic Analyzer, and 3730/3730xl DNA Analyzer Data Collection v3.0, v3.1.1, and v4.0, and 310 with Data Collection v3.1. Files generated on older instrument platforms have limited support.

Features of Earlier Versions of SeqScape® Software

Features of v2.7

SeqScape® Software Version 2.7 includes the following features and enhancements:

- Supports data files generated on 3500/3500xl Data Collection v1.0
- Supports POP-6™ polymer under 3730/3730xl Data Collection
- Integrates KB™ Basecaller v1.4.1

Features of v2.5

SeqScape® Software Version 2.5 includes the following features and enhancements:

- Support for sample files generated by the Applied Biosystems® 3130/3130xl Genetic Analyzers.
- An optional electronic signature feature that allows you to review and electronically sign-off an audit trail (see page 9-3).
- The ability to replace "?" with another character when exporting a consensus sequence (see page 9-14).
- Integration with VariantSEQr® Resequencing System to provide an easy and accurate data analysis solution (see page 7-19).
- The ability to allow a sample that has a sample-level heterozygous indel mutation (HIM) to assemble by skipping the filtering step in the analysis pipeline. Assembling the consensus sequence without filtering enables a specimen-level HIM to be detected in the Mutation Report (see page 8-7).
- Support for KB™ Basecaller v1.2 – The KB™ Basecaller, which calls pure or mixed bases with quality values, has been updated to correctly call the bases generated by the 3130/3130xl instruments (see the *KB™ Basecaller v1.2 FAQ*, PN 4362968).
- Additional KB™ Basecaller basecalling status indicators in the Analysis QC Report (see page 7-27).
- The ability to label the specimen review status (see page 7-8).
- The ability to automatically export projects after auto-analysis (see page B-9).

- The ability to selectively delete layers from the Reference Data Group (see page 6-33).

Features of v2.1

The following features and enhancements were introduced in SeqScape® software v2.1:

- Optimization
  - Faster project load and import
  - Analysis pipeline time is 33% faster
- Data analysis
  - **Automatic analysis** – Integration for automatic analysis with Applied Biosystems® 3730/3730*xl* Data Collection Software version 2.0 and 3100/3100-*Avant* Data Collection Software version 2.0
  - **M13 primer trimming** – Trims the M13 universal primer sequence from the PCR product
  - **3′ clear range trimming** – Trims PCR primer sequences
  - **New Basecallers** – KB™ Basecaller basecallers identify pure and mixed bases, then assign quality values for data generated on 310, 3100, and 3100-Avant systems and Applied Biosystems® 3730 and 3730xl systems
- Microsoft® Windows® XP OS compatibility
- Reports
  - **NT Mutations report** – Shows the amino acid effect of each mutation and hyperlinks link the nucleotide variant with the amino acid variant (and vice versa).
  - **New Genotyping report** – Identifies the genotype at one or more positions of all the specimens in the project and provides coverage of the samples against the expected coverage based on the targeted sequence.
  - **Specimen Statistics report** – Provides the position of the clear range on the reference of the samples on the segment
  - **Amino Acid report** – Provides the option of displaying the amino acid full name or three-letter symbol
- Specimen heterozygous indel mutation (HIM) identification
  - Mutations report lists HIMs

- Amino acid alignment can be coupled to the nucleotide alignment
- Sequence Collector/BioLIMS software is no longer supported

Features of v2.0

The following features were introduced in SeqScape® Software v2.0:

- **Extended Reference Data Group** – SeqScape® Software v2.0 contains an extended reference data group (RDG). The RDG contains a known reference sequence and any known nucleotide or amino acid variants. The RDG available in this new version of the software enables analysis of simple or complex projects.

  The Reference Sequence within the RDG can be a:

  – Contiguous reference sequence with a single reading frame.

  – Contiguous reference sequence with multiple reading frames.

  – Reference sequence constructed from several reference segments. Each segment can come from different locations in the genome.

  The reference sequence can contain features such as exons, introns, splice junctions, primer-binding sites, and promoter regions.
- **Heterozygous insertion/deletion mutation** – SeqScape® identifies potential instances of this variant which often require manual review by trained personnel.
- **Library searching** – You can compare each consensus sequence to a sequence library to identify the closest match genotype, allele or haplotype.
- **Enhanced reports** – You can customize reports. Each variant in the report is hyperlinked to the sequence data, providing rapid transition from results to data. The results reports eliminate the need to manually record results. You can automatically sort and reorganize any report.
- **Password protection and audit trail** – The software protects your data by providing password protection, automatic lockout when the software is inactive, and three levels of access control. An audit trail records each manual insertion, deletion, or base modification, with reasons for each change.

- **Integration automation** – The software uses an improved process for setting up samples for automated analysis on Applied Biosystems® 3730/3730*xl* instruments.
- **New Basecallers** – The KB™ Basecaller basecaller is an algorithm that identifies mixed or pure bases and generates sample quality values. The ABI basecaller is an algorithm used in sequencing analysis software.
- **Basecalling with ABI basecaller only is no longer available** – In SeqScape® Software v1.1, you can choose to basecall data with ABI basecaller or ABI basecaller with TraceTuner™ Software. In SeqScape® Software v2.0, you do not have the option to basecall with ABI basecaller only. The new options are:
  - Basecall with ABI basecaller and TraceTuner™ software (automatic)
  - Basecall with KB™ Basecaller
- **Implicit Reference is no longer available** – In software v1.1, you can have an empty RDG and use the first specimen as your implicit reference sequence, but this is no longer available in SeqScape® Software v2.0. However, you can create an RDG and add an .ab1 sample file as a reference sequence.

About SeqScape® Software

Genetic Analyzer Applications

SeqScape® software is one of a suite of Applied Biosystems® Genetic Analyzer software applications designed to control an instrument, collect data, and manage automated analysis. This suite of data collection and analysis software systems includes:

- GeneMapper® Software – Performs genotyping using fragment analysis methods.
- Sequencing Analysis Software – Displays, analyzes, edits, and prints sequencing files.
- Variant Reporter Software - Performs sequence comparisons for variant identification in resequencing projects.

SeqScape® Software Applications

Common resequencing applications include:

- SNP discovery and validation
- Mutation analysis and heterozygote identification
- Sequence confirmation for mutagenesis or clone-construct confirmation studies
- Identification of genotype, allele, and haplotype from a library of known sequences
- VariantSEQr® Resequencing System

Resequencing Data with SeqScape® Software

SeqScape® software allows analysis of resequenced data, comparing consensus sequences to a known reference sequence and optionally searching against a sequence library.

For example, a simple project might contain one contiguous reference sequence in a single reading frame, with no known nucleotide or amino acid variant information. SeqScape® software compares a consensus sequence to this reference sequence, identifying any differences.

A more complex project might include a reference sequence constructed from several reference segments representing multiple exons and introns. You can use SeqScape® software to:

- Build unique sequence layers composed of different groupings of reference sequence features.
- Compare consensus sequences to each unique layer.

- Identify differences.
- Compare the sequence to a library of sequences to identify the closest match.

Data Sources for Resequencing Projects

You can create projects in SeqScape® software using sequencing data generated from the following systems:

- 310 Genetic Analyzer
- 377 DNA Sequencer – Support in SeqScape® Software 3 is limited to files generated by the 377 DNA Sequencer. Sample files can be displayed, edited, post-processed, printed and exported. Sample files may not be re-basecalled.
- 3100-*Avant* Genetic Analyzer
- 3100 Genetic Analyzer
- 3700 DNA Analyzer
- Applied Biosystems® 3730 DNA Analyzer
- Applied Biosystems® 3730*xl* DNA Analyzer
- Applied Biosystems® 3130 Genetic Analyzer
- Applied Biosystems® 3130*xl* Genetic Analyzer
- Applied Biosystems® 3500 Genetic Analyzer
- Applied Biosystems® 3500*xl* Genetic Analyzer

Each project can contain:

- Unanalyzed sample files (.ab1)
- Previously basecalled sample files (.ab1)
- Text sequences (.seq or .fsta)
- Aligned consensus sequences

A single project can contain sample files from one or a mixture of instrument platforms. The software analyzes the data, displays several views of the analyzed project, and reports on results for quality control and data review.

Levels of Automated Analysis

SeqScape® software performs two levels of analysis:

- It identifies variants, positions that differ from the reference sequence, and classifies those variants as known or unknown.

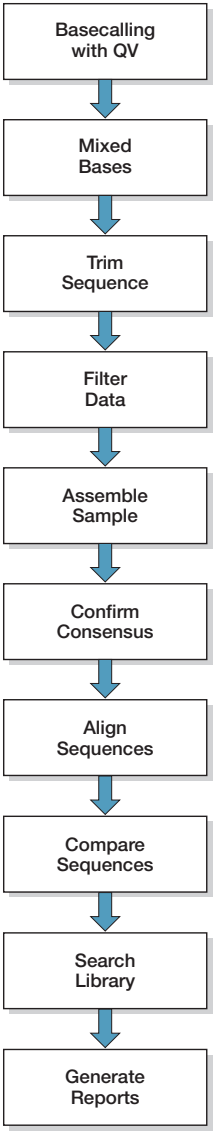
- It searches a library of alleles or haplotypes to identify the alleles that most closely match the sample.

What the Software Does

After you add a reference sequence, a library, and sample files, SeqScape® software performs two levels of analysis:

- Identification of nucleotide and amino acid variants. The software identifies positions that differ from the reference sequence and classifies those variants as known or unknown variants.
- Identification of genotypes, alleles, or haplotypes from a library. In addition to identification of variants, the software searches a library of genotypes, alleles, or haplotypes and identifies the alleles that most closely match each consensus sequence.

How the Software Performs Analysis



SeqScape® Software Analysis Pipeline

You provide the following information to the system before analysis:

- A reference sequence (backbone) made up of one or more reference segments and any known nucleotide variant information or amino acid variant information. (SeqScape® software uses the backbone to classify all polymorphic positions as known variants or unknown variants.)
- An allele library (a set of sequences for the alleles or haplotypes).

Using the reference sequence, variants, allele library, and software settings, you create a reusable project template. With this template and the sequencing samples, SeqScape® software:

- Performs (in order) basecalling, quality value assignment, and mixed base identification.
- Trims low-quality bases from each sequence.
- Identifies poor-quality samples and removes them from further analysis.
- Assembles the remaining samples against the reference sequence and generates a specimen consensus sequence.
- Reviews the basecalling quality values and the sample assembly to confirm, improve, and assign quality values to the consensus sequence.
- Identifies variants by aligning specimen sequences to the reference sequence and comparing the specimen consensus sequences to the reference sequence.
- Generates 10 detailed reports.

Note: If you link a library to a project, the software also automatically searches the library to find the closest match to each consensus sequence.

When the analysis is complete, the software generates a project file that contains sample files, a consensus sequence for each specimen, and 10 reports. You can print and export your results.

SeqScape® Software Structure

SeqScape® software is organized around two main windows:

- SeqScape® Manager window, from which you enter and manage the information necessary to perform analyses
- Project window, from which you manage the results of analyses

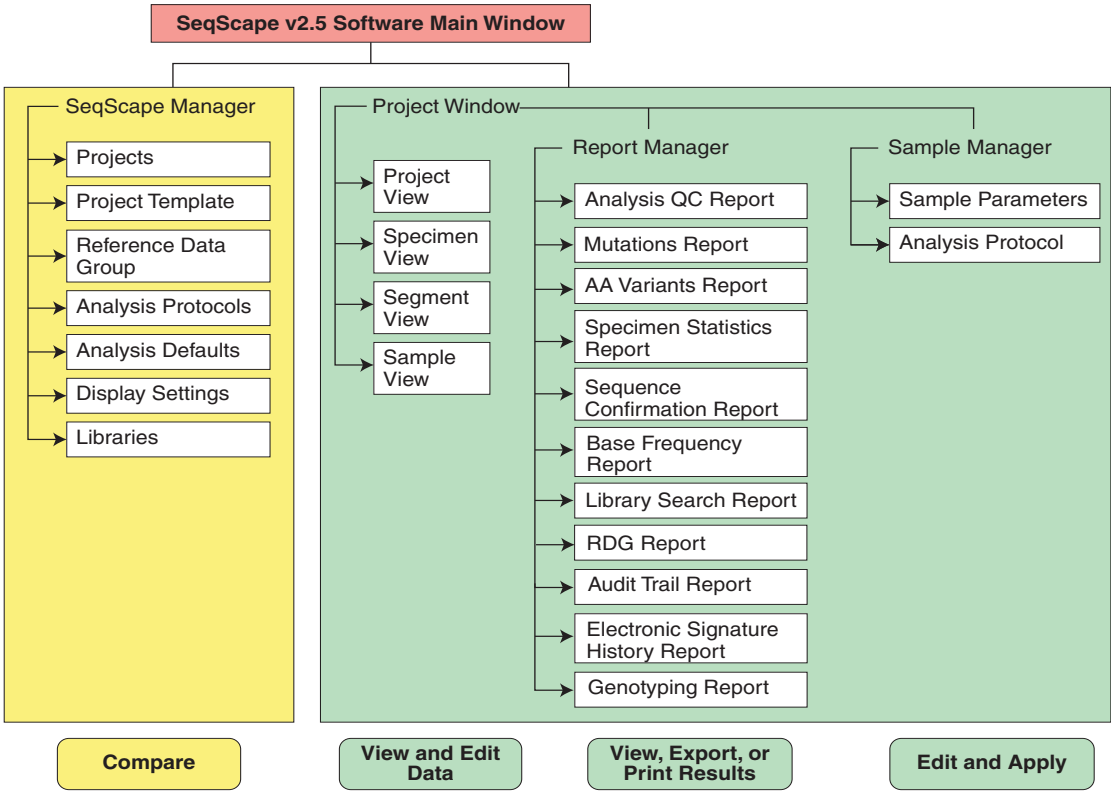


Figure 1-1 SeqScape® Software structure

SeqScape® Manager Window

In the SeqScape® Manager window, you configure projects by creating project templates. The project templates can be reused in multiple projects and can be exported to be shared with other researchers. The project templates contain:

- Reference sequence information
- Analysis settings (including analysis protocols)
- Display settings

Project Window

In the Project window, you can view your data in the following ways:

View	Description
Project View	Shows the reference sequence, each specimen consensus sequence, and electropherogram snippets for each sample file in each specimen. The Expanded Nucleotide View shows all the nucleotides. The Collapsed Nucleotide View shows only variants of the nucleotides. The Expanded Amino Acid View shows all the amino acids. Characters (NT or AA) that are the same as the reference are shown as dots. The Character/Dots button switches to show or hide the view. The Identification pane, which shows the library search results, appears at the bottom of the Project view.
Specimen View	Shows the clear range and orientation of each sample and how the samples line up to the reference sequence, and the overview pane with active ROIs.
Segment View	A table of sample information. Clicking a row in the table shows the corresponding sample sequence below. The Layout tab shows the direction of each sample within the segment. The Assembly tab shows samples aligned to the consensus sequence. An overview pane represents forward and reverse sequences, variants, and ROIs. Electropherograms can be displayed for one or all sequences.
Sample View	Shows pertinent information for the sample, which includes annotation, sequence, electropherogram, and raw data.

Refer to Appendix A, "Sample and Consensus Quality Values," for detailed descriptions of the Project Window views.

SeqScape® Software Toolbars

The SeqScape® software toolbars display buttons for software functions that you are likely to use often. Refer to the next two figures for the names, descriptions, and keyboard shortcuts for each button. The buttons in the top row, Figure 1-2, are processing tools.

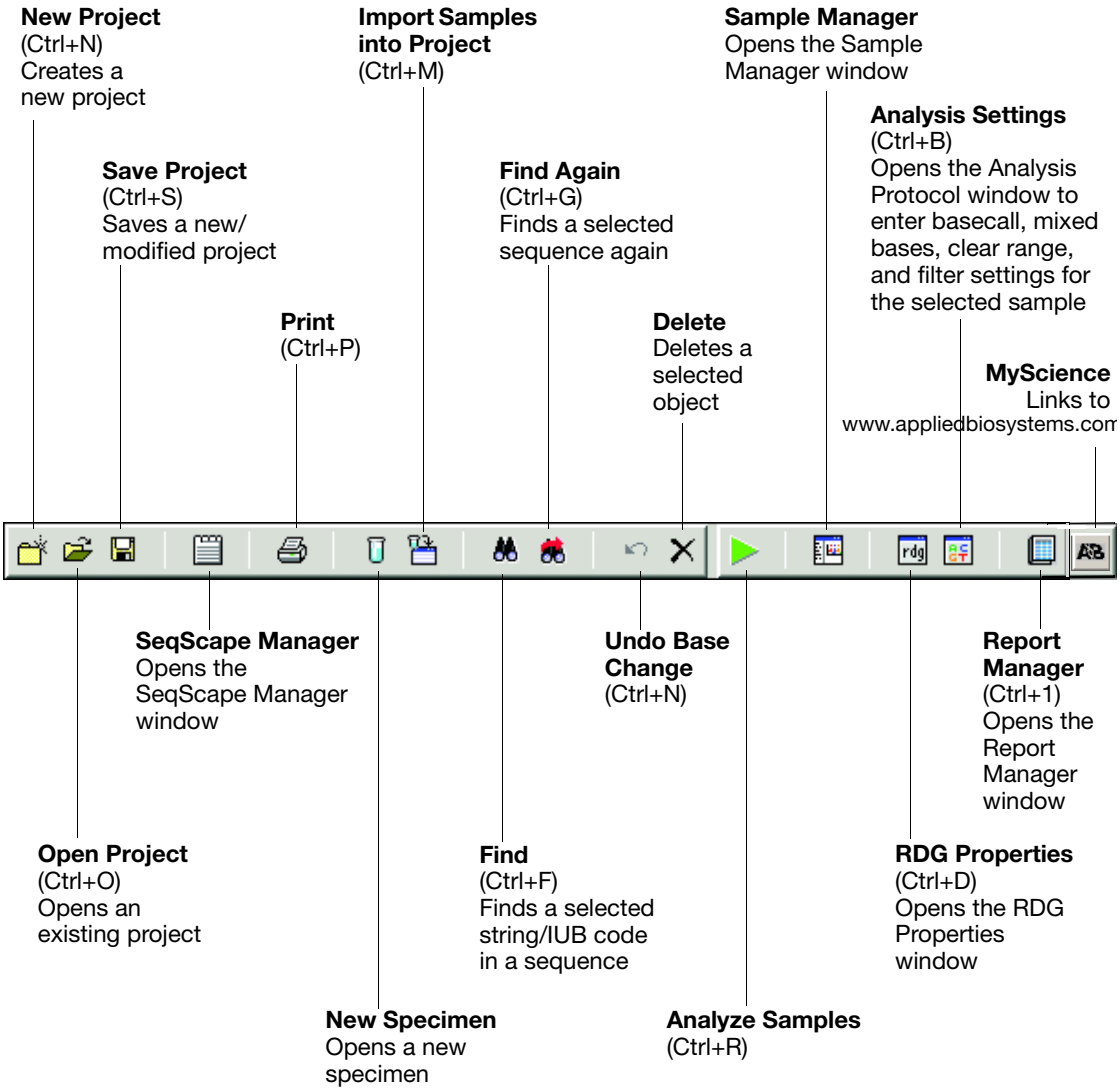


Figure 1-2 Main Toolbar

The buttons in the second row, Figure 1-3, are viewing options for the projects you create.

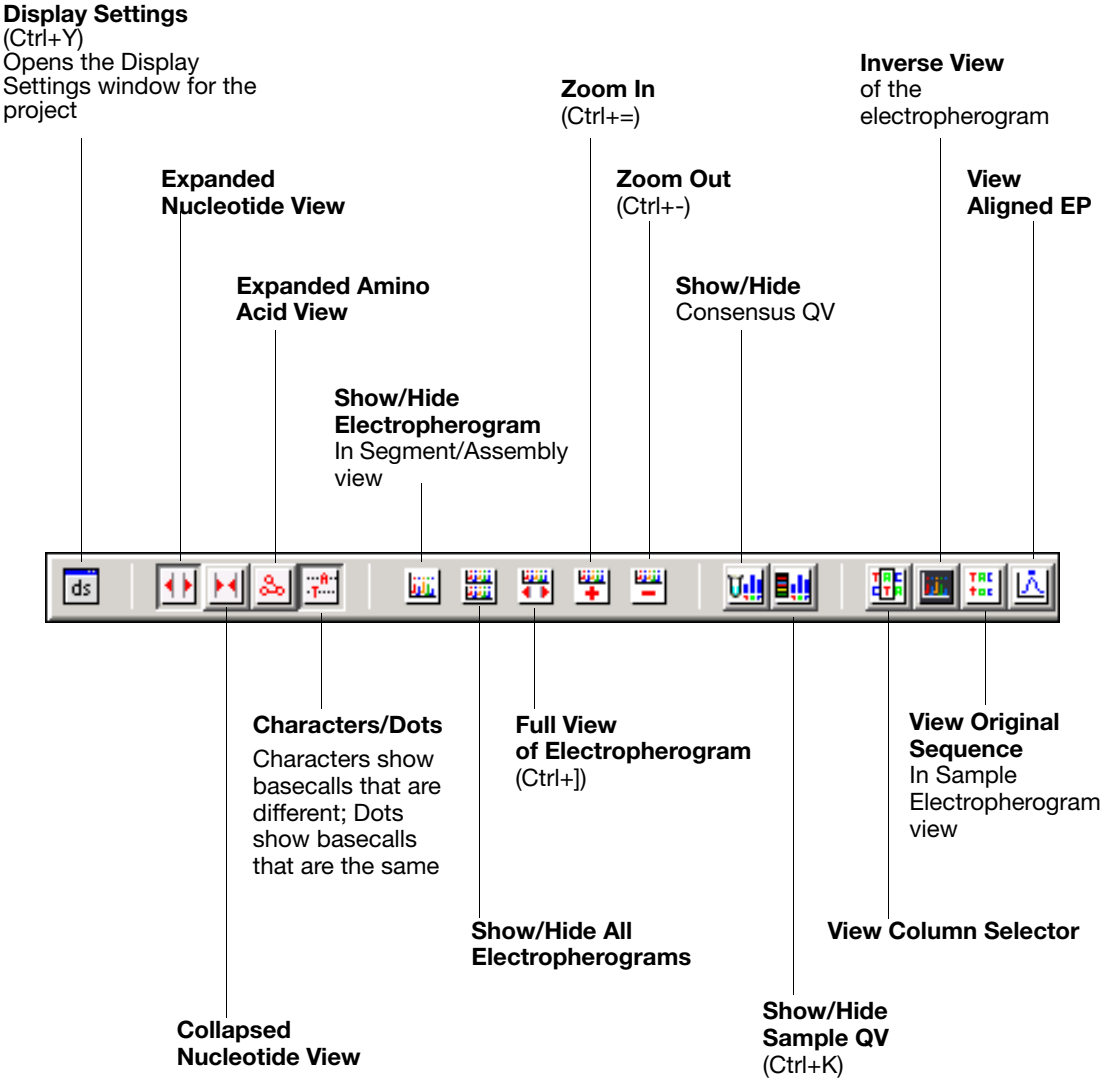
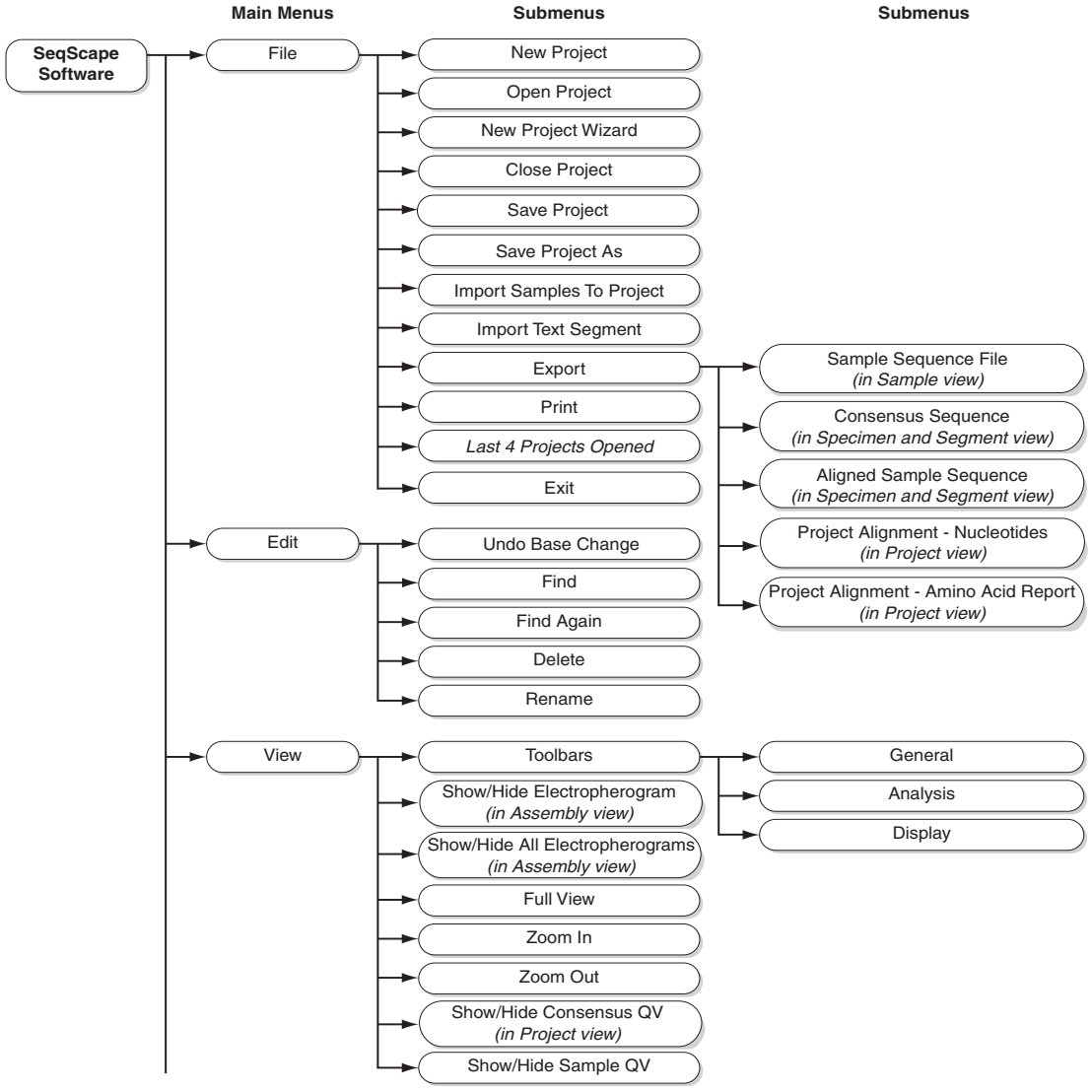


Figure 1-3 Viewing toolbar

SeqScape® Software Main Window Structure

Figure 1-4 shows the menu structure of the main SeqScape® window.



Continued on next page...

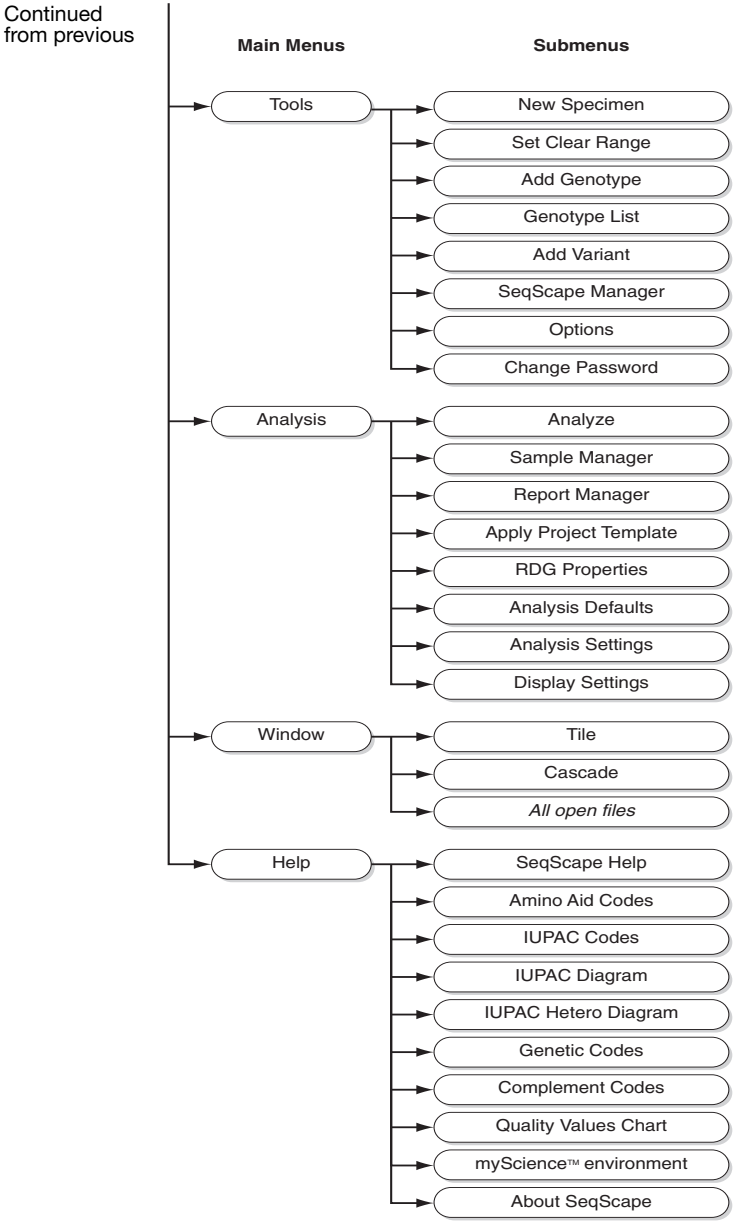
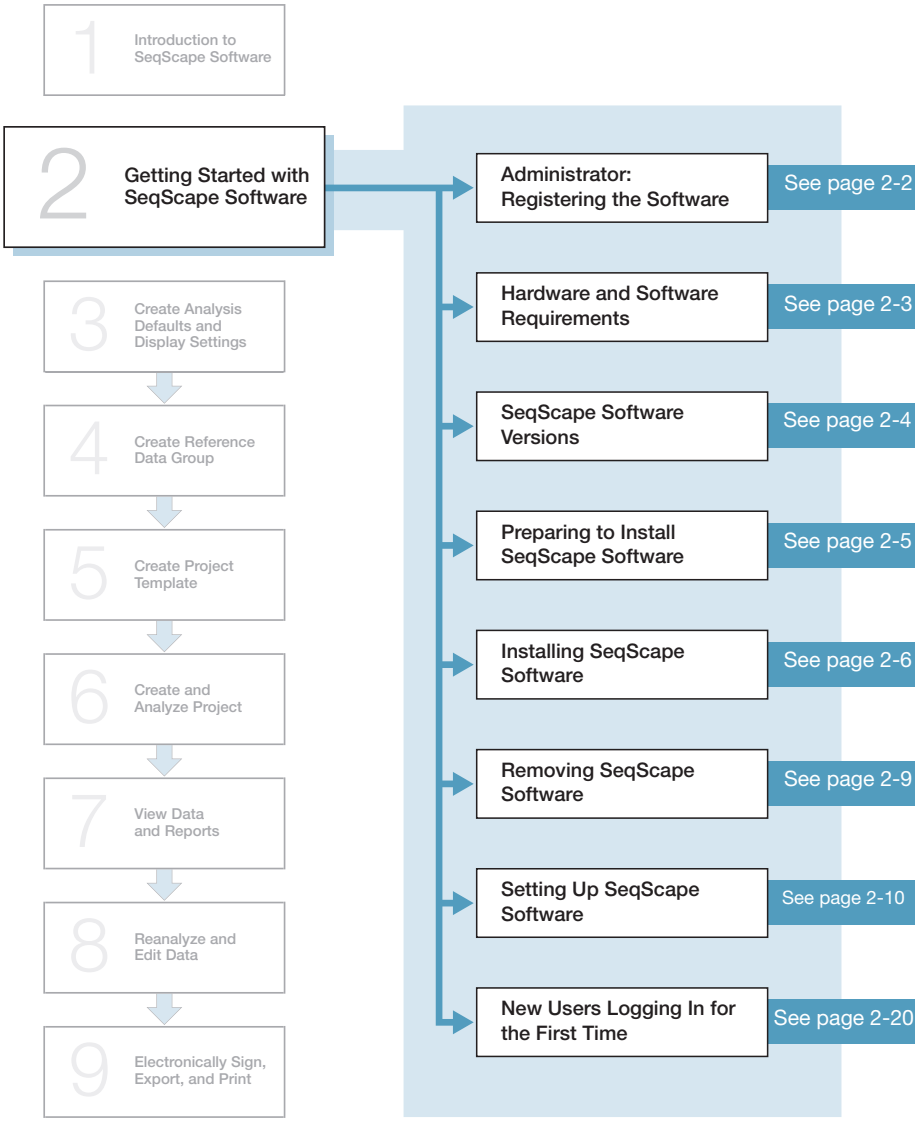


Figure 1-4 Main SeqScape® window menus

2 Getting Started with SeqScape® Software



Administrator: Registering the Software

This chapter provides information you need to know before installing and using the SeqScape® Software Version 3. The administrator must follow the procedures in this section, "Administrator: Registering the Software," through "Setting Up SeqScape® Software" on page 2-10.

License and Warranty

Before you begin, read Appendix H, "Software Warranty Information," which explains your rights and responsibilities regarding the SeqScape software.

During the installation process of the software, you must accept the terms and conditions of the Software License Agreement before the software can be installed.

Registering Your Software

To register your copy of SeqScape® software, complete the registration card (included in this software package) and return it to Life Technologies.

Registering the software enables Life Technologies to send you notification of software updates and any other future information that may be specific to SeqScape® software owners.

IMPORTANT! Your product registration number is on the registration card. Be sure to record the number here before you return the registration card.

Registration Number:

Hardware and Software Requirements

Minimum System Requirements

SeqScape® software can be installed on a computer that meets the minimum requirements summarized in Table 2-1 below.

In general, the more memory and the more processing power the system has, the better its performance.

Table 2-1 Minimum system requirements

System Component	Minimum Requirements
CPU	2 GHz or faster Intel processor The software does not run on computers with a Xeon chip set.
CD-ROM drive	DVD/CD-compatible optical drive.
Operating system	Microsoft® Windows 7 Professional (SP1), (32-bit)
RAM	2 GB
Monitor	1280 x 1024 resolution with full-screen display. Use the Windows 7 default theme.
Disk space	1 GB of free space for the application. Storage requirements depend primarily on the quantity of data to be generated and stored. It is common to store many SeqScape® software project files on the analysis computer. Install SeqScape® software on a partition with enough space for the projects and their files because SeqScape® software stores data files in the area where the program is installed.

Hard Drive Partitions

The installer uses the following location for SeqScape® software files:

C:\Applied Biosystems\SeqScape

The drive letter is determined by the following conditions:

Table 2-2 Drive letter determination:

If the computer …	The installer selects drive …
is not connected to a genetic analyzer	C (default), or D (if C drive is not available)
has data collection software that is connected to a genetic analyzer	C

SeqScape® Software Versions

There are two available versions of SeqScape® Software 3:

Software Version	Intended for a system that ...	Existing Data Status
SeqScape® Software 3 full release version	Has no SeqScape® software installed, or has the 30-day demo version installed	You must export any files created in the 30-day demo version before the software expires. You can then import any exported files into the full release version of SeqScape® Software 3. See "Upgrading from SeqScape® Software v2.x" on page 2-7
SeqScape® Software 3 30-day demo	Meets the minimum requirements	You must export any files created in the 30-day demo version before the software expires. You can then import any exported files into the full release version of SeqScape® Software 3.

Preparing to Install SeqScape® Software

Note: An administrator should install the software and use it for the first time. The administrator can set up the software for the analyst, scientist, or other administrator users.

To prepare for the installation:

1. Ensure that your system meets the minimum requirements (see "Hardware and Software Requirements" on page 2-3).

   Check that you have at least 1 GB of free disk space to accommodate SeqScape® software, and sufficient space for all projects and their sample files.

2. Temporarily turn off any virus-protection software.

3. Exit all programs except the Applied Biosystems® 3500/3500*xl*, 3730/3730*xl,* 3130/3130*xl*, or 3100/3100-*Avant* Data Collection software, if applicable.

IMPORTANT! To properly install SeqScape® Software 3 on a computer that is connected to an Applied Biosystems® 3500/3500*xl* or 3130/3130*xl* Genetic Analyzer or 3730/3730*xl* DNA Analyzer, **the data collection software must be running**. SeqScape® software does not register with the Data Service if the data collection software is not running.

Installing the Software

SeqScape® software can be installed for use with Administrative, Scientist, or Analyst privileges on computers that meet the minimum configuration requirements.

Installer Process

The SeqScape® Software 3 installer works as a clean installer only. It is not an upgrade installer.

If you have previously installed an older version of SeqScape® on your Windows 7 computer, you must uninstall the older version of SeqScape® before installation of SeqScape® Software 3 can proceed. This is to prevent overriding existing projects and data. See "Uninstalling Earlier Versions of SeqScape® Software" on page 2-9.

Installing the Full Version for the First Time

The administrator of the software installs the software and sets up new users.

Follow this procedure to install the full version of SeqScape® Software 3 only if you are installing on a computer that:

- Does not have a previous version of the software *or*
- Has a 30-day demo version of the software installed

IMPORTANT! If you are using a SeqScape® software 30-day demo, export your data **before** the end of the 30 days and **before** you load the full version of SeqScape® software. To ensure data are not lost:

a. Export your data objects: **Tools** > **SeqScape Manager** > **Files**, then click **Export**.

b. Export your User names and Authentication configuration settings: **Tools > Options > Users** or **Tools > Options > Authentication**. Select Files, then click **Export**.

c. Load the software as described in the procedure below.

d. Re-import your data.

To install the full version of SeqScape® software for the first time:

1. Insert the *SeqScape® Software 3* CD into the Windows 7 computer CD-ROM drive.

2. If the installer does not start automatically, double-click **setup.exe** on the CD.

3. When the InstallShield Wizard Complete window opens, click **Finish**.

After the software is installed, the administrator must log into the software for the first time. After the initial login, the software can be set up for additional users.

Upgrading from SeqScape® Software v2.x

To upgrade from 2.x versions of your software:

IMPORTANT! If you wish to use your data from previous versions of the SeqScape® software with SeqScape® Software 3, you must transfer the exported data to the Windows 7 computer. You can do this with a USB drive or other method.

1. Export your data objects: **Tools** > **SeqScape Manager** > **Files**, then click **Export**.

   Data objects to export are: Project, Project Template, Reference Data Group, Libraries, Analysis Defaults, Analysis Protocols, and Display Settings. You can select and export all Data Objects of one type in a single step.

2. Export your User Names and Authentication Configuration Settings: **Tools > Options > Users or Tools > Options > Authentication** > **Files**, then click **Export**.

3. Install SeqScape® Software 3 as described in "To install the full version of SeqScape® software for the first time:" on page 2-6.

4. Re-import your data objects: **Tools** > **SeqScape Manager** > **Import.**

5. Re-import your User Names and Authentication Configuration Settings: **Tools > Options > User or Authentication > Import**.

To import all the data objects at once from SeqScape® v2.7:

If you are upgrading from SeqScape® v2.7, you can import the DataStore exported from SeqScape® v2.7 and obtain all the data objects at once. Assuming that Windows 7 is installed on a separate computer, you will need to manually transfer the SeqScape® v2.7 DataStore folder from the old computer to the new computer.

1. Copy and store the DataStore folder from SeqScape® v2.7 found at:

   `C:\AppliedBiosystems\SeqScape\data\DataStore`

2. Install SeqScape® 3 on your Windows 7 computer as described in the installation directions above.

3. Rename the DataStore folder on the newly installed software found at:

   `C:\Applied Biosystems\SeqScape\data\DataStore`

   to

   `DataStore_as_installed`.

4. Copy the SeqScape® v2.7 DataStore folder to:

   `C:\Applied Biosystems\SeqScape\data\`

   This will generate:

   `C:\Applied Biosystems\SeqScape\data\DataStore`

   SeqScape® 3 will now use the DataStore from SeqScape® v2.7.

Existing Users

All existing users of an earlier version of SeqScape® software will have Analyst privileges. Only a user belonging to the Administrator group can change the user to Scientist or Analyst. A dialog box opens for users who existed in previous versions to set up their user profiles (name and password) when they try to use SeqScape® Software 3 for the first time.

Uninstalling Earlier Versions of SeqScape® Software

What the Uninstallation Process Does

To completely remove SeqScape® software from your computer, follow the procedure in this section. The uninstallation process:

- Deletes all folders and files installed by SeqScape® software. However, if you moved SeqScape® software folders or files from their original installed location, they may not be found and deleted by the uninstallation process.
- Does not delete any files or folders created by users. Any files that have been added to the application folders, such as those created when the applications are run, are not deleted by the uninstallation process.

Removing the Software

To uninstall SeqScape® software:

1. Select **Start** > **All Programs** > **Applied Biosystems** > **SeqScape** > **Uninstall SeqScape v*x.x***.

2. Continue to follow the instructions to uninstall the software.

   When the uninstallation is complete, all the software program files are removed. Your data files remain on the computer. The uninstaller does not delete any folders or files created after installation. If you want to delete any folders and files created after installation, you must remove them manually.

Setting Up SeqScape® Software

Before You Begin

When you start the software for the first time, you are prompted with a registration dialog box that creates an administrator account. Log in to SeqScape® software as **Admin** and enter the password you created.

To create new users, you must log in as **Admin**. Logging in with a user name allows SeqScape® software to track each user's interactions with each project.

For information on the privileges for each category of user using the software, refer to Appendix E, "User Privileges."

File-Naming Convention

Spaces and some alphanumeric characters are not valid for user names or file names. The invalid characters are:

\ / : * ? " < > |

An error message is displayed if you use any of these characters. You must remove the invalid character to continue.

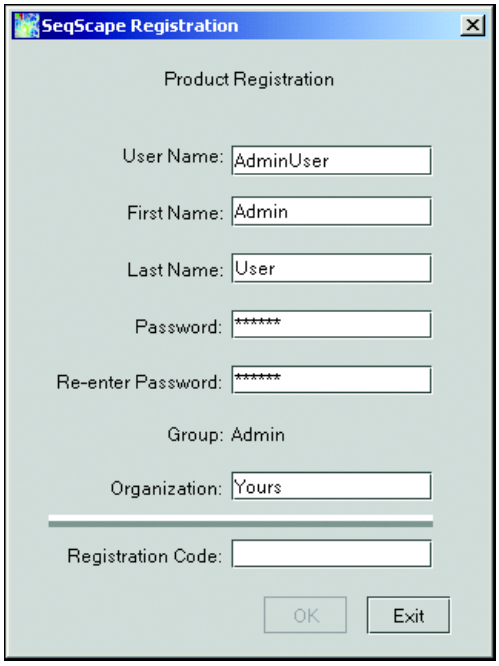
IMPORTANT! User names cannot be named seqscape_admin in this version of the software. If you have used this user name in a previous version of the software, you must change the user name to follow the file-naming convention shown above.

Starting SeqScape® Software

To start the software for the first time:

1. Double-click the SeqScape® software desktop shortcut.

   The SeqScape® Registration dialog box opens.



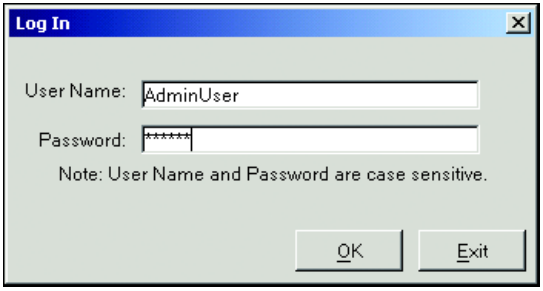
2. In the SeqScape® Registration dialog box, enter all the information in the text fields. The User Name and password must be 6 to 15 characters long.

   The first user created is automatically assigned Administrator privileges.

3. Enter the registration code from the registration card you received with your software.

4. Click **OK**.

   The splash screen appears while the program is loading, then the Log In dialog box opens.



5. Enter your user name and password again, then click **OK**.

   The License dialog box opens.

6. Read the license agreement then click **Accept**.

   The SeqScape® software main window opens.

Creating New Users

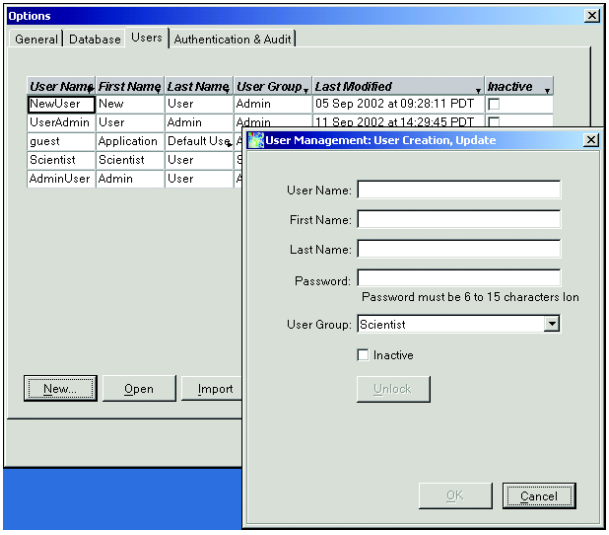
Because SeqScape® software tracks the projects and settings for each user, Life Technologies recommends that you create users for each individual who uses SeqScape® software on the computer. The Users tab allows exporting of user names and access privileges for these users.

IMPORTANT! The administrator is the only person who can set up and change the information in the Users tab. The selections in this tab are inactive for all other users.

To set up new users:

1. Select **Tools** > **Options** to open the Options dialog box.

2. In the Options dialog box, select the **Users** tab, then click **New**.



3. Fill in the appropriate user name, password, first and last names, then select the level of user from the User Group drop-down list.

   Note: Enter a User Name that contains only alphanumeric characters. This field must not contain any spaces or characters that do not conform with the Microsoft® Windows OS file system. Refer to "File-Naming Convention" on page 2-10.

   The new user appears in the list in the Users tab.

4. To set up users on multiple computers, Life Technologies recommends that the administrator:

   a. Create the list of users.

   b. Export the file.

   c. Install SeqScape® software on the other computers.

   d. Import the user file.

New users can log in after exiting and then restarting SeqScape® software.

Setting Up Authentication and Audit

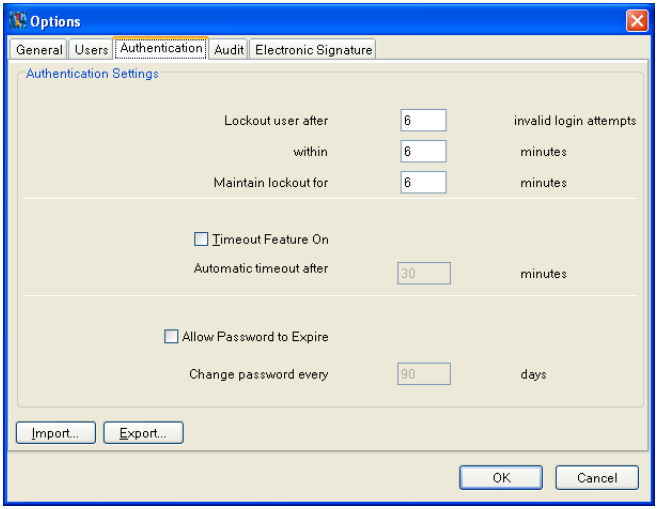
Users belonging to the Administrator group can change the default settings in the Authentication and Audit tabs for security features of the application.

Note: The Administrator is the only person who can set up and change the information in the Authentication and Audit tabs. The selections in the tabs are inactive for all other users.

The Authentication and Audit panes provide a way to track the changes in projects such as base change, variants, or processes you want to track. You must turn Audit Trail On for tracking to occur.

To set up authentication:

1. **Tools > Options > Authentication**



2. Lockout occurs when a user enters an incorrect password or user name the number of times you select for the **Lockout user after invalid login attempts** field. Enter the number or accept the default.

3. The **within minutes** field indicates that the user will be locked out if the maximum number of attempts occurs within the time entered in this field. Enter a number or accept the default.

4. The **Maintain lockout for minutes** field indicates the number of minutes that must elapse before the user can login again after being locked out of SeqScape® software. Enter the number of minutes or accept the defaults.

5. The **Automatic timeout after** field indicates the length of time the software can sit idle before logging out the user. Enter the number of minutes or accept the default. Select **Timeout Feature On** to activate this feature.

6. The **Change password every days** field indicates the number of days after which a user must enter a new password. Enter a number of days or accept the default. Select **Allow Password to Expire** to activate this feature.

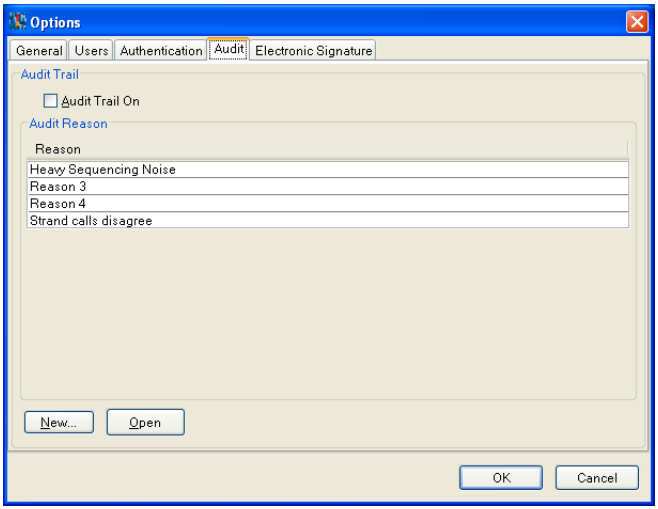
7. If desired, click **Export**, then navigate to export the authentication settings to another computer. The **Import** button allows authentication settings to be imported from another computer.

8. Click **OK** in the Options dialog box to save the authentication settings.

Note: An administrator can import or export Authentication configurations from one computer to another. For example, an administrator may want to set up authentication information for many users, then select all the files and export them to other systems using SeqScape® software.

To set up audit:

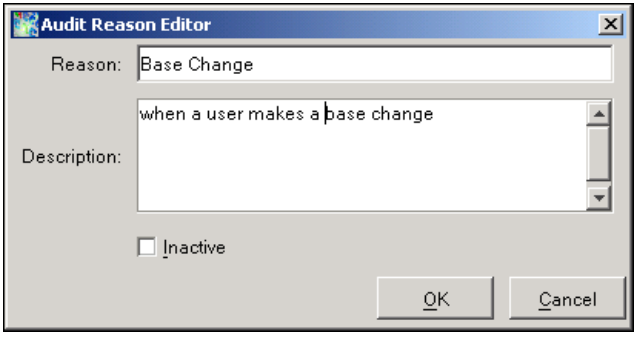
1. **Tools > Options > Audit**



Note: In the Audit Trail pane, select the **Audit Trail On** check box to have a dialog box open whenever a specified reason for recording an audit trail event occurs.

2. In the Audit Reason pane, enter reasons to provide an audit trail.

   a. To add a new reason to audit, click **New**.



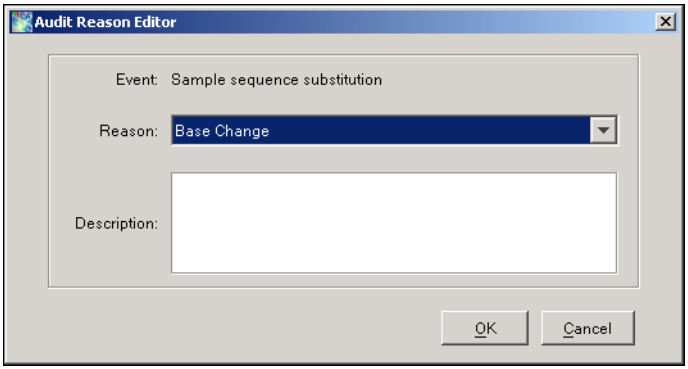
   b. In the Reason field, type a reason for a change to the project to identify, for example, a base change, or a variant that is imported.

   c. Enter a description of the reason, if desired.

d. Click **OK**. The first reason appears in the Reason list in the Options dialog box.

e. To open one of the existing reasons, select the reason and click on **Open** to read the description.

Note: Whenever a change is made in any of the project views, the Audit Reason Editor dialog box opens, allowing you to select the reason for the change from the drop-down list.



3. Click **OK** in the Options dialog box to save the audit settings.

Changing User Information

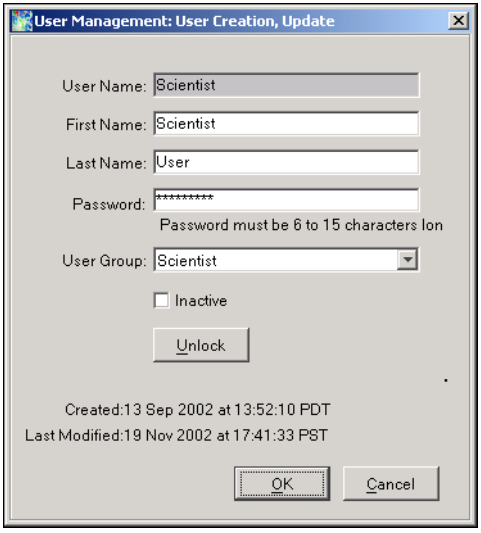
You can change the default settings for a user you are setting up.

IMPORTANT! The Administrator is the only person who can set up and change the information in the Users tab. The selections in this tab are inactive for all other users.

To change any of the information for a user:

1. In the Options dialog box, select the **Users** tab.

2. Double-click the name of the user in the list to open the User Management dialog box.



3. Change or correct the user information, then click **OK**.

4. If desired, click **Export** in the Options dialog box to export the application configuration settings and/or settings for a single user or multiple users in a zipped .ctf format.

5. Enter the path for exporting files in the Export User dialog box, then click **Export**.

6. Click **OK** to close the Options dialog box.

Note: This process can be used by the first administrator to set up additional users or another administrator. User settings can be imported or exported from one computer to another. For example, an administrator can set up user information for many users, then select all the user files and export them to other systems using SeqScape® software.

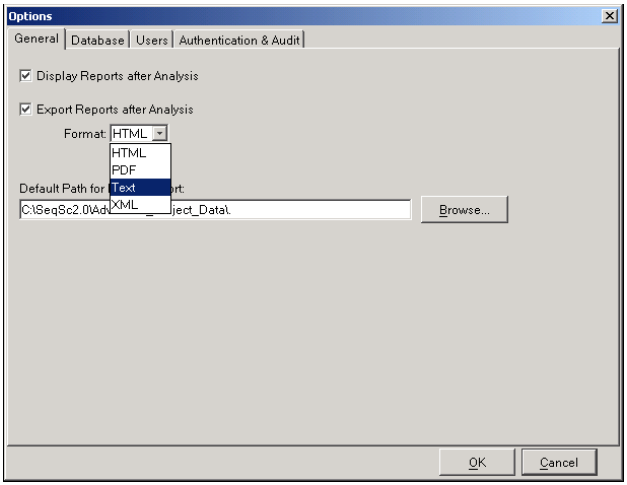
Setting Up the Default Directory

The default directory should be set up for users for importing and exporting data files. If the directory path is not set up, the default directory opens to C:\.

To set up the default directory path:

1. In the SeqScape® software main window, select **Tools** > **Options**.

2. In the General tab, select the appropriate check boxes for your setup, if desired.
   a. Select the **Display Reports after Analysis** check box.
   b. Select the **Export Reports after Analysis** check box, if desired, then select from the **Format** drop-down list the format in which to export reports.



3. Click **Browse**, then navigate to the default directory for storing files.

4. Click **Open**. The exported reports are stored in the directory you select as the default.

5. Click **OK** to save the directory path, then close the dialog box.

New Users Logging In for the First Time

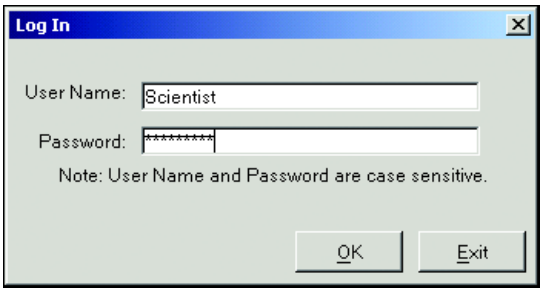
When New Users Log In

After the installation and setup are complete, new users can log in to the software.

To log in to SeqScape® software:

1. Start the software by double-clicking the SeqScape® software desktop shortcut  .
2. In the Log In dialog box, enter your user name and password, then click **OK**.

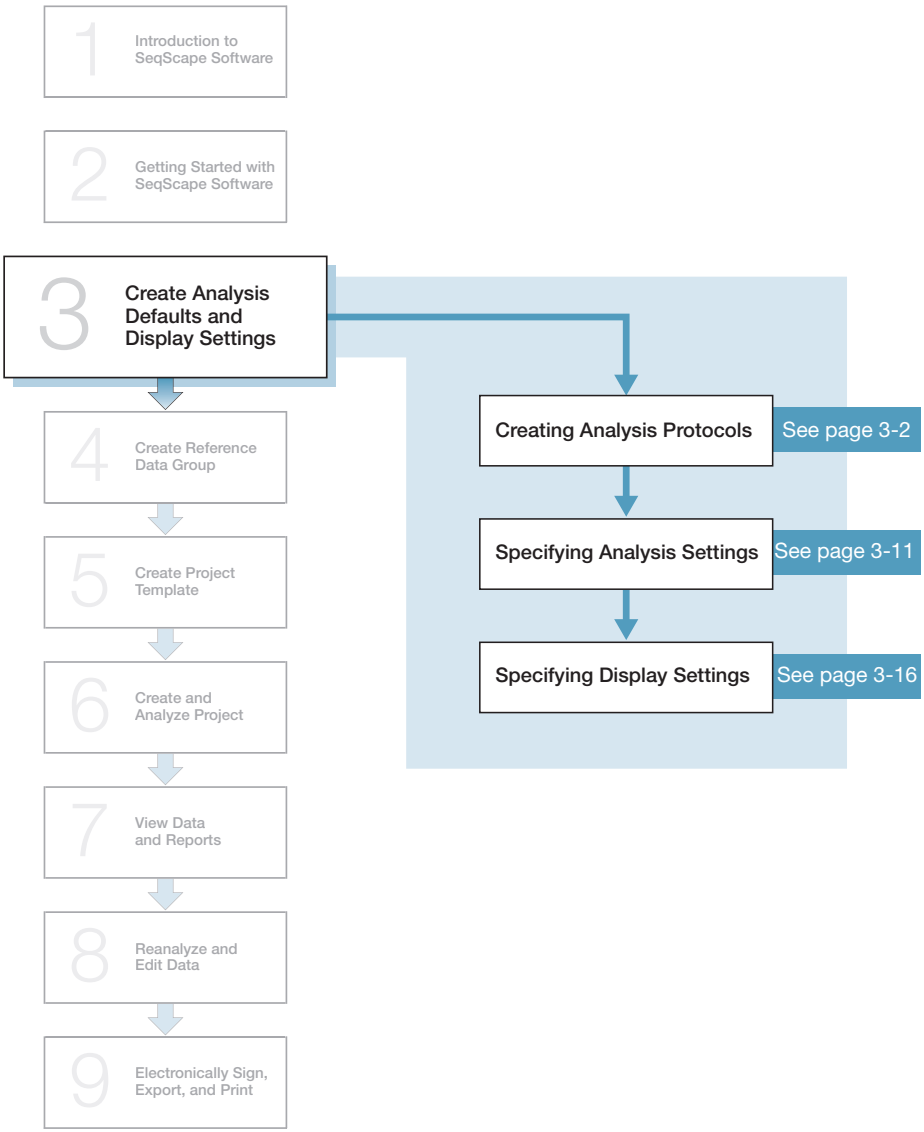
   IMPORTANT! If you have forgotten your user name or password, contact your administrator.



SeqScape® software is ready to use.

Note: All existing users of an earlier version of SeqScape® software have Analyst privileges. Only a user belonging to the Administrator group can change the user to Scientist or Analyst. Users who existed in previous versions are asked to set up their user profiles (name and password) when they try to use SeqScape® Software 3 for the first time.

3 Creating Analysis Defaults and Display Settings



Creating an Analysis Protocol

An analysis protocol in SeqScape® Software 3 specifies the analysis conditions to be applied to your samples. You can specify the analysis protocol settings for one or more samples. You must select an analysis protocol before selecting analysis defaults. The analysis protocol settings include:

- Basecalling
- Mixed bases
- Clear range
- Filtering

Opening the Analysis Protocol Editor

To open the Analysis Protocol Editor:

1. Select **Tools** > **SeqScape Manager**.
2. Select the **Analysis Protocols** tab, then select the project in the list for which you want to change the settings.
3. Click **Properties**.

General Settings

The **General** tab (Figure 3-1) displays general information on the analysis protocol, for example, the name, creation date, and modification date.

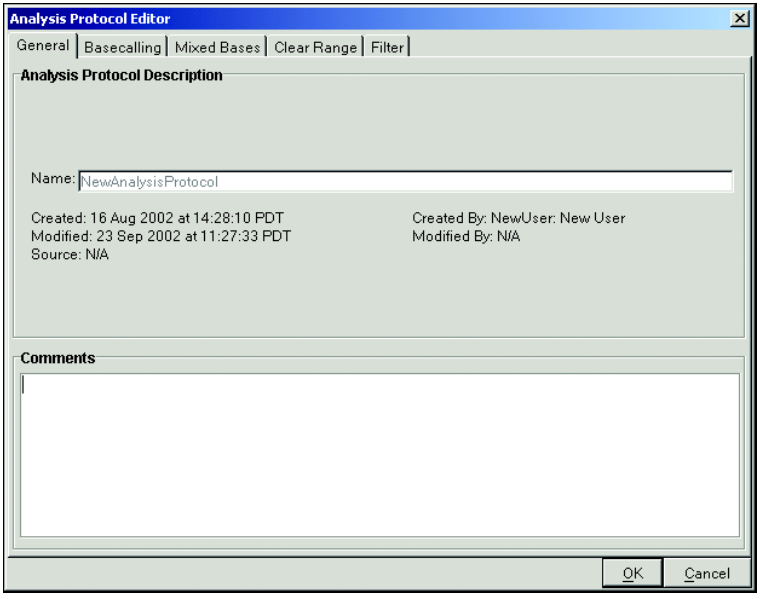


Figure 3-1 Analysis Protocol Editor General tab

Specifying the Basecall Settings

The Basecalling tab has settings for how the software calls bases. The basecaller you select is determined by the instrument and chemistry you are using. For details on basecalling files and dye primer set selections, see Appendix B, "Basecallers and DyeSet/Primer Files."

1. In the Analysis Protocol Editor, select the **Basecalling** tab to view the basecalling settings (Figure 3-2).

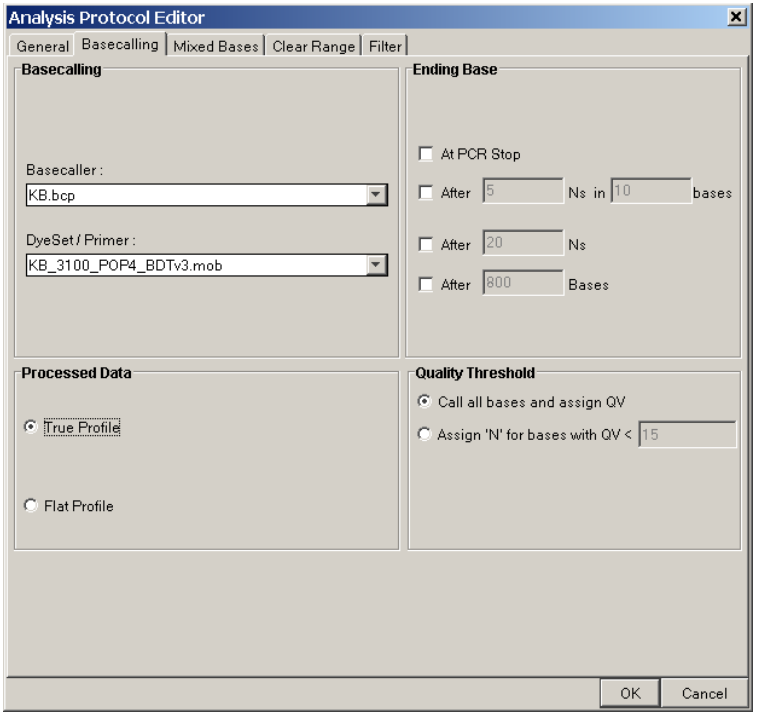


Figure 3-2 Analysis Protocol Editor Basecalling tab

2. Select the appropriate basecaller dedicated to your instrument. For more information, refer to Appendix B, "Basecallers and DyeSet/Primer Files."

3. Select the DyeSet/Primer settings (.mob files), for the instrument you are using. For more information, refer to Appendix B, "Basecallers and DyeSet/Primer Files."

4. If you have short PCR products, you should end basecalling at the end of the PCR product. In this case, select the **At PCR Stop** check box.

5. You can also stop basecalling after a specified number of ambiguities, or Ns, or after a certain number of bases. Enter your changes to the settings.

6. For KB™ Basecaller only, select how you want to display the data:

   True Profile – Displays data as processed traces scaled uniformly so that the average height of peaks in the region of strongest signal is about equal to a fixed value. The profile of the processed traces is very similar to that of the raw traces.

   Flat Profile – Displays data as processed traces scaled semi-locally so that the average height of peaks in any region is about equal to a fixed value. The profile of the processed traces is flat on an intermediate scale (> about 40 bases).

7. For KB™ Basecaller only, set the quality threshold.

   Do not assign 'N' to Basecalls– Use this setting to assign a base to every position, as well as the QV.

   Assign 'N' for bases with QV < ▫ – Use this setting to assign Ns to bases with QVs less than the set point. The QV will be grayed out.

   IMPORTANT! **Life Technologies recommends that you use the KB™ Basecaller to perform your analysis.** The KB™ Basecaller basecaller was introduced with SeqScape® software v2.0 and Sequencing Analysis software v5.0. We will continue to improve and develop this algorithm. The ABI basecaller is an older algorithm that will be removed from future versions of the software.

Specifying the Mixed Bases Settings

In the **Mixed Bases** tab, you can select **Use Mixed Base Identification** to generates calls following the international standard IUB code for heterozygous positions.

To specify the mixed bases settings:

1. In the Analysis Protocol Editor, select the **Mixed Bases** tab.

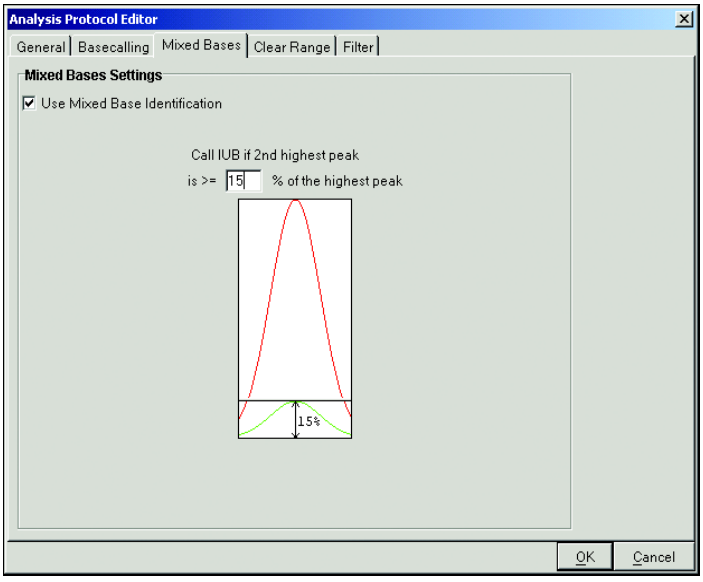


Figure 3-3 Analysis Protocol Editor Mixed Bases tab

2. Select the **Use Mixed Base Identification** check box to generate calls according to the international standard IUB code for two base heterozygous positions R, Y, K, M, S, and W. Mixed bases identification occurs only if the height of the second highest peak is greater than or equal to a percentage of the main peak height.

Note: The secondary peak threshold is only one of many criteria used by the KB™ Basecaller to call mixed bases. Reaching this secondary peak threshold is a necessary but not sufficient condition for arriving at a mixed base determination.

3. Set the level according to sample type, reaction kit, purification reaction, and expected or acceptable percentage. Enter the threshold for calling a mixed base for the % value of the primary peak.

   IMPORTANT! **Life Technologies recommends that you do not enter a value less than the 15% default value.** If you decrease the default percentage to detect low-percentage mixed bases, the background signal may be higher and interfere with mixed-base detection. Be aware of this condition.

Specifying Clear Range

Clear Range is the region of the sequence that remains after excluding the low-quality or error-prone sequence at the 5′ and 3′ ends and the M13 primer sequence, if applicable. You can specify a range as a default. Life Technologies recommends that you always select **Use reference trimming**.

You can apply all or a subset of the Clear Range Methods algorithms. Each is applied in order from top to bottom, with the clear range method never being lengthened based on the settings in subsequent algorithms. The result is that the smallest clear range is used. If you want to preserve the existing clear range in a sample when reapplying analysis protocol settings to a sample, do not select any of the Clear Range methods.

IMPORTANT! Although you can create a protocol without selecting a clear range method, Life Technologies recommends that you select at least one clear range method for reference trimming.

The Clear Range tab enables you to set the part of the sequence that you consider to be good quality. Good quality means that the sequence has the fewest errors and ambiguities, and offers good base calling and spacing.

To set the way the clear range is determined:

1. In the Analysis Protocol Editor, select the **Clear Range** tab (Figure 3-4).

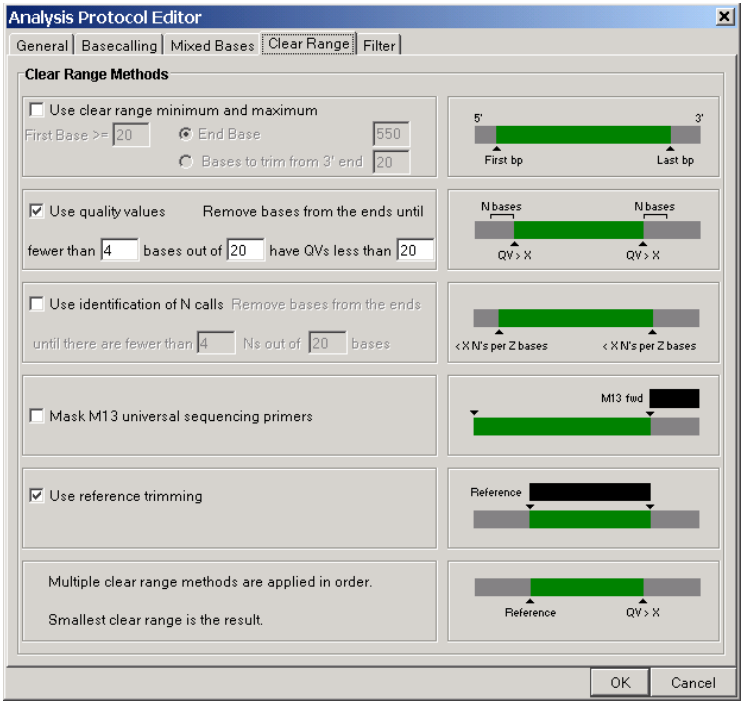


Figure 3-4 Analysis Protocol Editor Clear Range tab

Because SeqScape® software generates quality values for each base, you can choose to use a region of sequence where a certain number of bases reach an appropriate quality value.

2. Select **Use clear range minimum and maximum,** then set the minimum first base and maximum end base or the number of bases to trim from the 3′ end of the clear range.

3. Select **Use quality values** to remove bases until there are < X number of bases per Z number of bases with QV < Y. This sets a window with a specified number of allowed low-quality bases.

4. Select **Use identification of N calls** to remove bases until there are < X number of Ns per Y number of bases. This sets a window with a specified number of allowed ambiguous base calls (Ns).

5. Select **Mask M13 universal sequencing primers** to exclude the M13 primer sequence from the clear range.

6. Select **Use reference trimming** to have the samples automatically trimmed to contain only sequences that align to the reference.

Specifying the Filter Settings

The **Filter** tab sets the criteria for rejecting sequences if they do not meet minimum standards.

IMPORTANT! Sequences not meeting the filter settings are not assembled.

To select the filter settings:

1. In the Analysis Protocol Editor, select the **Filter** tab (Figure 3-5).

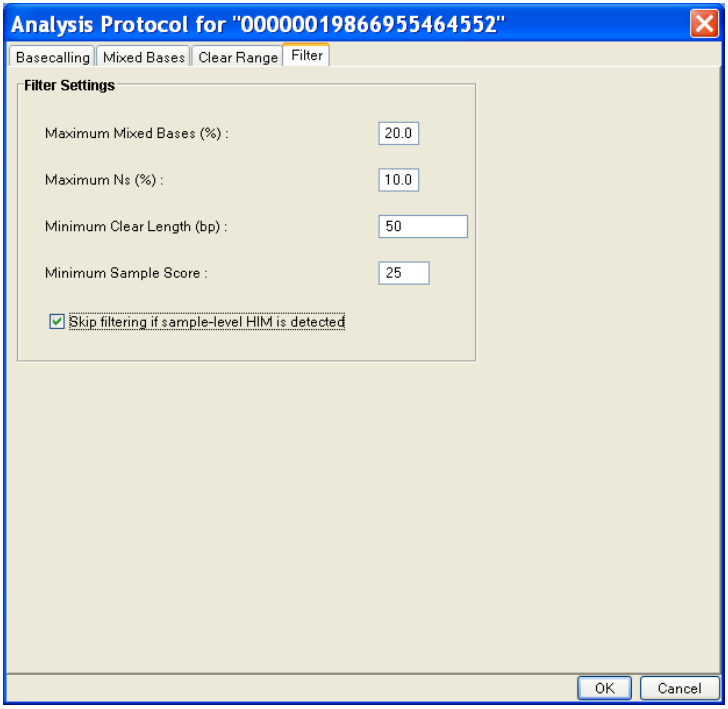


Figure 3-5 Analysis Protocol Editor Filter tab

2. Specify the values for the filter Settings, using the description of the settings in Table 3-1 as a guide

Table 3-1 Filter Parameter Descriptions

Parameter	Description
Maximum Mixed Bases (%)	Total maximum percentage of mixed bases that can occur in the clear range of a sample file. Any more than this number causes the sample to fail analysis. Use the maximum percentage of mixed bases to look for frame shift.
Maximum Ns (%)	Total maximum percentage of Ns that can occur in the clear range of a sample file. Any more than this number causes the sample to fail analysis. Use the maximum percentage of ambiguities (N) and the minimum length settings to ensure that you are working with enough data for further analysis.
Minimum Clear Length (bp)	Minimum length of bases required in the clear range of a sample file. Any less than this number causes the sample to fail analysis. Use the maximum percentage of ambiguities (N) and the minimum length settings to ensure that you are working with enough data for further analysis.
Minimum Sample Score	Minimum quality value score (average of all sample QVs in the clear range) that is acceptable. The range is 1–50 (see "Sample Quality Values" on page A-3. Use a minimum sample score to ensure that the quality of the sequences is high. A setting of 20 indicates that the data are accepted if the mean quality value of all bases in the clear range is 20 or greater. This corresponds to a 1-to-100, or 1%, error rate.

Completing the Analysis Protocol

When the analysis protocol is complete, click **OK** to save the new settings. If you do not want to save the new settings, click **Cancel** to save the previous settings.

Note: To implement the changes, you must click OK to save them and then run the analysis.

Specifying the Analysis Settings

To accommodate sample variability and to ensure the quality of your results, you can modify the settings used to analyze a sample and then reapply them to other samples. The Analysis Settings are set in the **SeqScape Manager > Analysis Defaults** Data Object tab.

You can save changes to the analysis defaults and display settings contained in a project, and you can also save them in SeqScape® Manager to be used in a project template.

The procedures in the following sections describe selecting the analysis settings for a set of samples. The Analysis Settings are defined and saved within each Analysis Default, found in **Tools > SeqScape Manager > Analysis Defaults**.

For information on reapplying a new project template, see "Reanalyzing a Project Using a Different Project Template" on page 6-21.

Gap and Extension Penalties

Project Settings

The Project Settings are found at **Tools > SeqScape Manager > Analysis Defaults**. Open an Analysis Default object by double-clicking on its name. In the Edit Analysis Settings window the Project Tab will show the gap and extension penalties.

The settings for Gap Penalty apply for alignment of different specimen consensus sequences to each other and to the reference.

If you add gap and extension penalties, these settings introduce gaps into sequence alignments, allowing the alignment to be extended into regions where one sequence may have lost or gained characters not in the other gap penalty score (G+Ln). G is gap penalty, L is the length of gap, and n is the number of bases. A penalty is subtracted for each gap introduced into an alignment because gap increases uncertainty in an alignment.

Note: Life Technologies recommends that you use the default values. The default settings are optimized for the current algorithm.

Specimen Tab Settings

The settings for Gap Penalty and Extension Penalty apply to setting alignment of samples to the reference.

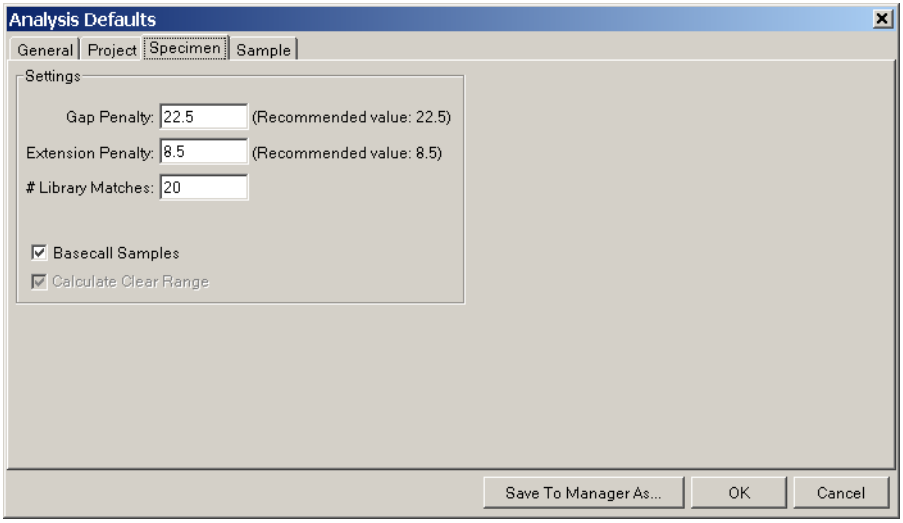


Figure 3-6 Analysis Defaults Specimen tab

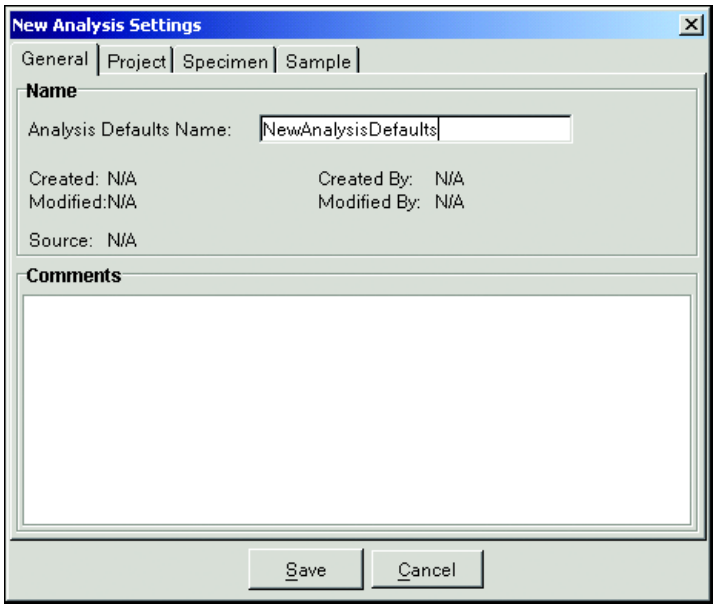
Setting Analysis Defaults

To create new Analysis Defaults:

1. In the SeqScape® Manager, select the **Analysis Defaults** tab, then click **New**.

2. In the General tab of the New Analysis Settings dialog box, enter an Analysis Defaults Name.

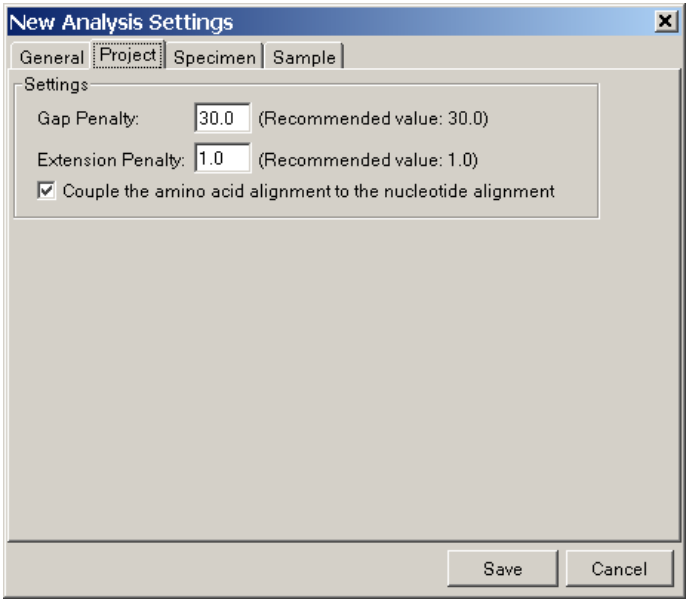
   Note: The name cannot contain spaces or characters that do not conform with the Windows file system. See "File-Naming Convention" on page 2-10.



3. Enter any comments pertaining to the new analysis settings in the **Comments** box.

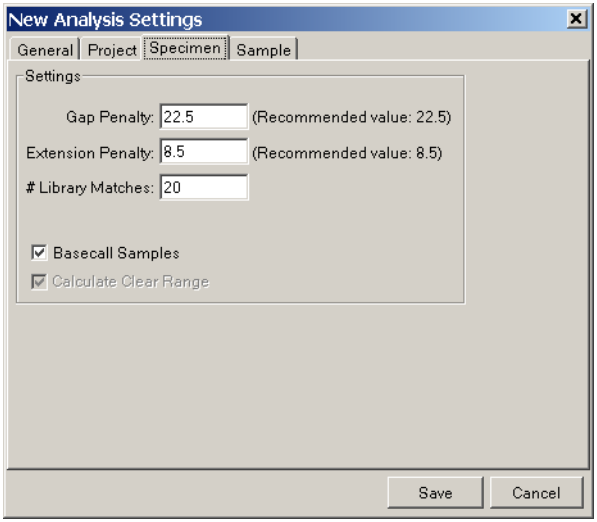
4. Select the **Project** tab and, if desired, change the Penalty Settings. The recommended gap penalty is 30 and the recommended extension penalty is 1.0.

   Note: The gap and extension penalties refer only to the alignment algorithms that are used to align the consensus sequences to the references and to each other. They do not affect the alignment of the samples to the reference for assembly.



5. Select the **Specimen** tab.

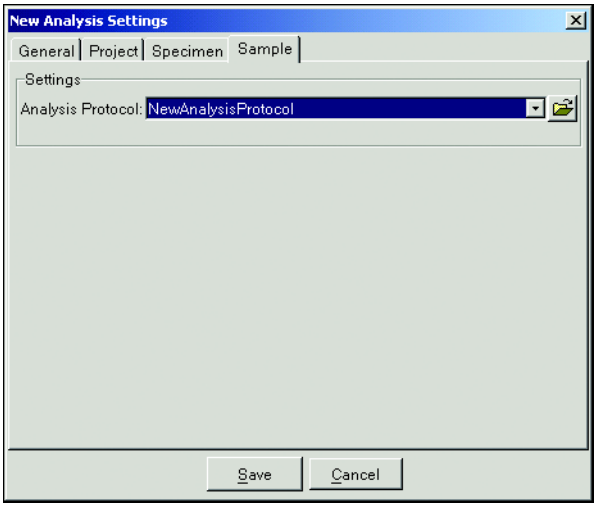
   The # Library Matches check box indicates the number of hits desired to match the library you select. The recommended gap penalty is 22.5 and the recommended extension penalty is 8.5.



   a. If desired, change the settings.

b. Select **Basecall Samples** to automatically calculate clear range and basecall samples. If you do not select Basecall Samples, the sample files are not basecalled, and it is assumed that you have previously basecalled and edited the data. When basecalling is skipped, the software proceeds to filtering and assembly in the analysis pipeline.

6. Select the **Sample** tab then select the analysis protocol you just created from the Analysis Protocol drop-down list.



7. Click **Save** to save the new settings for the project.

8. Click **Close** in the SeqScape® Manager dialog box.

Selecting the Analysis Default Settings for Individual Samples

Note: Changing the analysis defaults does not affect the analysis settings of samples that are already in the project.

To select the analysis settings for each sample individually:

1. Select the sample in the Project view.

2. Select **Analysis** > **Analysis Settings** to open the Analysis Protocol for that individual sample file.

3. Make relevant changes to the settings, then click **Save**.

Specifying Display Settings

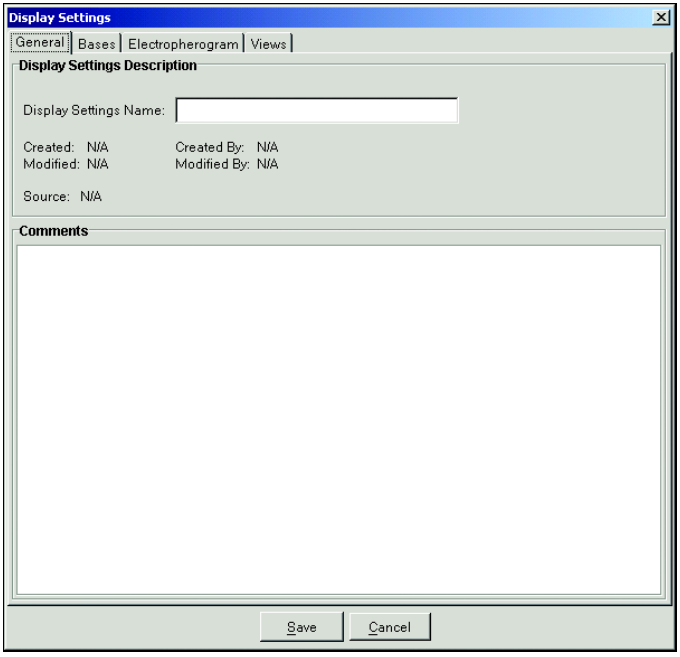
To accommodate personal preferences, SeqScape® software allows you to select the way results are displayed. The display settings can be modified and then reapplied to a project. The selected settings can also be saved in the SeqScape® Manager to be used in a project template. The display settings control:

- Font colors and style for bases
- Electropherogram display and axis scale
- Display views for variants
- Display views for nucleotide translation
- Quality value display and thresholds

To specify the display settings:

1. In the SeqScape® Manager, select the Display Settings tab, then click **New**.

   The Display Settings dialog box opens displaying the General tab.



2. Click the **Display Settings Name** field, then enter a name for the new display settings.

   IMPORTANT! The name cannot contain spaces or characters that do not conform with the Windows file system. Refer to "File-Naming Convention" on page 2-10.

3. Enter any comments you want to record for the sample.

4. Select the **Bases** tab.

   a. In the Base Font section, select your font preferences for the sequence bases, or use the defaults.

   b. In the Base Scale section, enter the frequency at which to display bases for the reference sequence in the Project view.

c. In the Base Colors section, select your color preferences for the sequence bases and electropherogram traces. To select a color, click the colored box (next to A:, G:, C:, and T:) to open the color chart, select a new color, then click **OK**.

d. In the Quality Values section, click the colored bars to open a color chart, then select the color, if necessary. To select the threshold values, drag the divider bars between the colors.

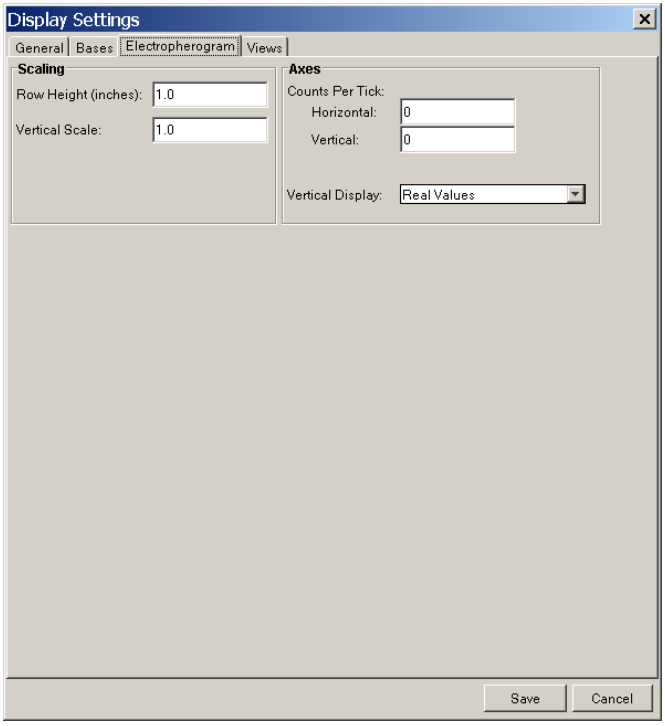
Note: The styles you specify here do not apply to variants.

For more information on quality values, see Chapter A, "Sample and Consensus Quality Values."

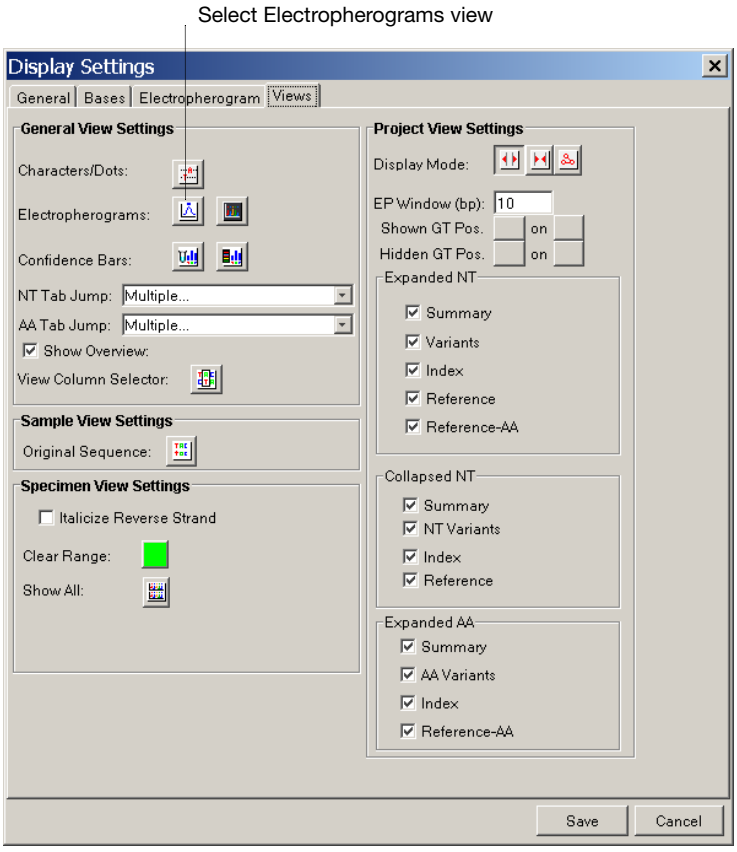
5. Select the **Electropherogram** tab, then:

a. Enter your Scaling and Axes preferences.

b. Select a Vertical Display setting (**Real Values** or **Relative**).



6. Select the **Electropherograms** view button, then enter your preferences for the Project and Specimen views.



a. In the General View Settings section, select the buttons for the displays you want turned on in the project. Select the **Views** tab.

Note: If you do not select the Electropherograms view button, peaks are not aligned when the Assembly view is printed.

Most of the buttons in the Views tab are the same as the viewing buttons on the lower row of the toolbar in the main SeqScape® window. Refer to "Viewing toolbar" on page 1-14.

b. In the drop-down lists, select how you want to tab through the data.

c. In the Sample View Settings section, select the icon if you want to see the original sequence displayed.

d. To differentiate forward and reverse sequences, in the Specimen View Settings section, select **Italicize Reverse Strand**.

e. In the Project View Settings section, enter the number of bases to be displayed for the Project view electropherogram snippets in the EP Window field (the minimum is 3).

7. Click **Save** to save the changes and close the dialog box.

   The new display settings are added to the SeqScape® Manager.

4 Creating a Reference Data Group

Section 4.1 Reference Data Group (RDG)

In This Section

Ways to Create a New Reference Data Group (RDG)

IMPORTANT! Only a user from the Administrator or Scientist group can set up a new RDG. Refer to Appendix E, "User Privileges," for a list of the privileges that apply to each user group.

You can create a new RDG by using the:

- RDG wizard
- SeqScape® Manager window to open a blank RDG

Follow the RDG wizard procedures below, if desired, to familiarize yourself with the windows of the RDG. Then, create subsequent RDGs by using the SeqScape® Manager. Refer to "Creating a New RDG Using SeqScape® Manager" on page 4-13.

About the Reference Data Group

The Reference Data Group defines the sequence to which SeqScape® Software Version 2.1 (and higher) compares the consensus segments to the reference sequence. It contains the reference sequence and reference-associated data. The reference sequence is the entire "backbone" sequence for the project, consisting of one or more reference segments separated by reference breaks.

The RDG contains all the gene/analysis-specific information consisting of:

- A reference sequence containing continuous or discontinuous sequences made up of one or more reference segments
- Nucleotide variants
- Amino acid variants
- Translation codon table
- Layers, which are units of analysis in any project, and regions of interest (ROIs) grouped together into layers for display and translation

- Associated allele libraries
- User-defined styles for identification of variants in the project

A reference segment is a contiguous segment of the reference sequence corresponding to a single contiguous DNA sequence. It is also a region of interest. The reference segment consists of:

- An analyzed sample file
- A text-only format, FASTA, or .seq file
- A GenBank format file

GenBank

GenBank Features

Every GenBank entry has a single contiguous sequence associated with it. This is also referred to as the source feature. This sequence is always numbered starting at 1.

Because of this, the sequence from a single GenBank entry translates into a single reference segment in the extended RDG. Numbering of the base ROI on this segment is set by default to start at 1.

Every GenBank entry has a feature table (Table 4-1). These features translate into regions of interest and layers in the extended RDG. In the following table, items in {} are qualifiers read for that feature key (for example, {gene} is the value of the `\gene` qualifier). If that qualifier does not exist, then "" is substituted.

Table 4-1 GenBank feature table

GenBank Feature	Extended RDG Equivalent
source	Skipped. The source feature corresponds to the region of interest associated with the whole reference segment that is automatically created.
exon	Region of interest is created, called {gene}_exon{number}. Translatable by default.
intron	Region of interest is created, called {gene}_intron{number}. Not translatable by default
gene	Region of interest is created, called {gene}_gene. Translatable by default.
CDS online platform	Layer is created, called ({gene}|{product})_CDS. If translatable regions of interest exist that correspond to this CDS, then those are used for building the layer. Otherwise, new regions of interest are created as required. New ROIs are called {layerName}_region1, {layerName}_region2, etc. Translation frame and orientation is taken from CDS qualifiers (`complement`() and `\codon_start`).
misc_feature	Region of interest is created called {note}. Not translatable by default.

Table 4-1 GenBank feature table

GenBank Feature	Extended RDG Equivalent
Unknown feature	Region of interest is created called {feature key}. Not translatable by default.

It is possible with this translation table to create many non-uniquely named ROIs (for example, if the entry had many variation features).

Downloading a GenBank File

To download a GenBank file from the Internet:

1. Open your web browser and enter the following URL:

   http://www.ncbi.nlm.nih.gov

2. In the **All Databases** pull-down menu, select **Nucleotide**.
3. Enter the name of the Nucleotide Sequence you want for the reference sequence in the box, then click **Search**.
4. After finding the desired sequence, select the check box to the left of the sequence name.
5. At the top of the page (next to Display Settings) select **Genbank (full)**. Then, near the top right, select **Send to**. Select **Destination > File > Format > Genbank (full) > Create File** and navigate to the location where you wish to save the file. After it is saved, the file, which can have a .gb, .fcgi, or.cgi extension, can then be imported into the RDG.

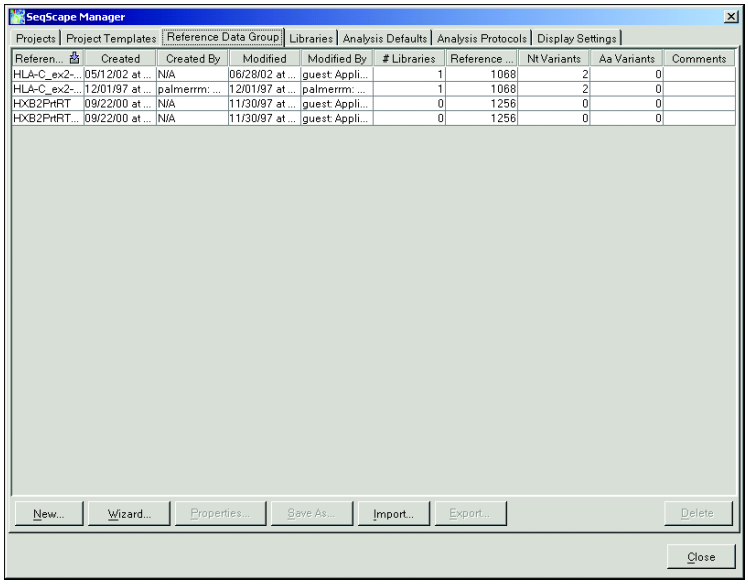
Section 4.2 Creating a New RDG Using the Wizard

Using the Wizard to Learn the Software

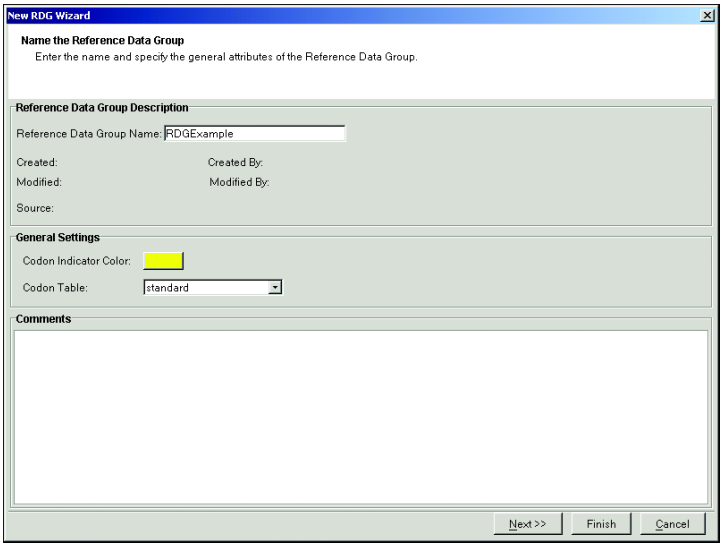
To create an RDG using the RDG wizard:

1. In the main SeqScape® window, select **Tools > SeqScape Manager**.

2. Select the **Reference Data Group** tab, then click **Wizard** at the bottom of the page.



3. Enter a name for the new RDG that conforms with the Windows file system. Refer to "File-Naming Convention" on page 2-10.

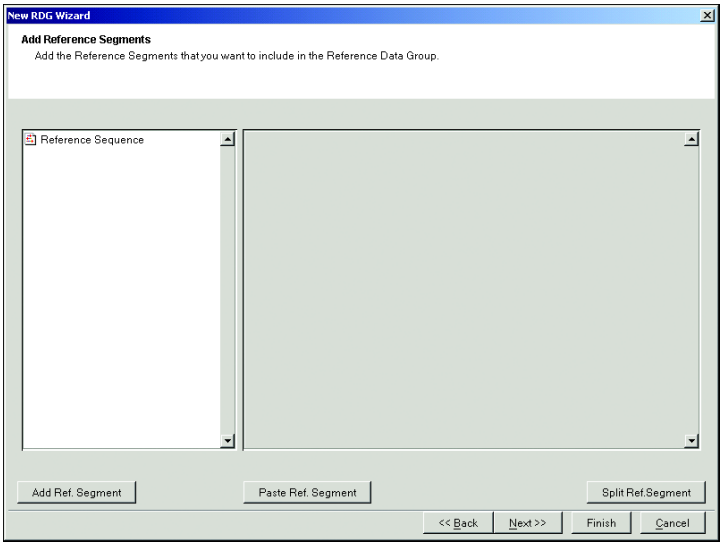


4. If desired, click the Codon Indicator Color button, then select a new color.

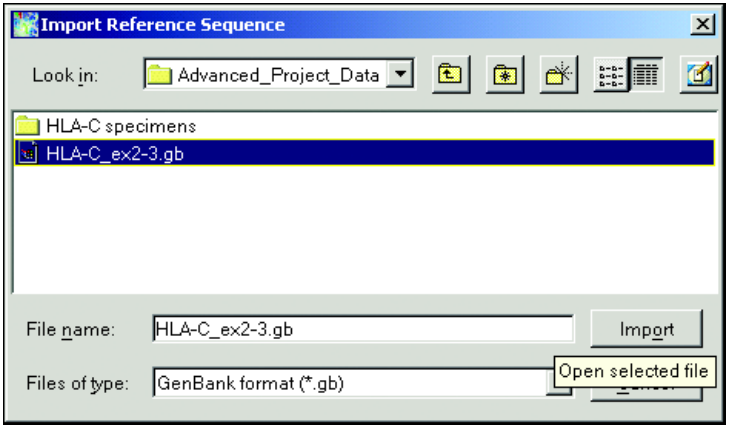
5. Select the Codon Table to use.

6. Click **Next**. To display the Reference Sequence pane. The Reference Sequence forms the backbone for comparison. It is made up of one or more reference segments.

7. Click **Add Ref. Segment** in the lower left to add a segment to the Reference Sequence. A reference segment is a single sequence imported from a text file or GenBank file.

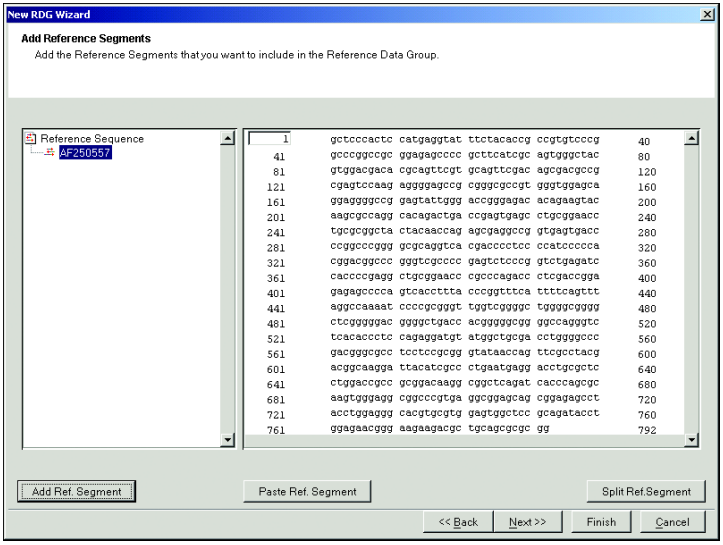


8. Navigate to the file containing the reference sequence that you have stored, such as a GenBank file (the file may have a .gb extension).



IMPORTANT! The window opens to the directory that was set up during installation of the software. If no default directory has been specified, the window opens to the C:\ drive. If you need to set up the default directory, select **Tools** > **Options** > **General**, then click **Browse** to locate the directory.

9. Click **Import**. The imported sequence appears in the right pane of the dialog box, as shown in the figure below.



10. Click **Next**.

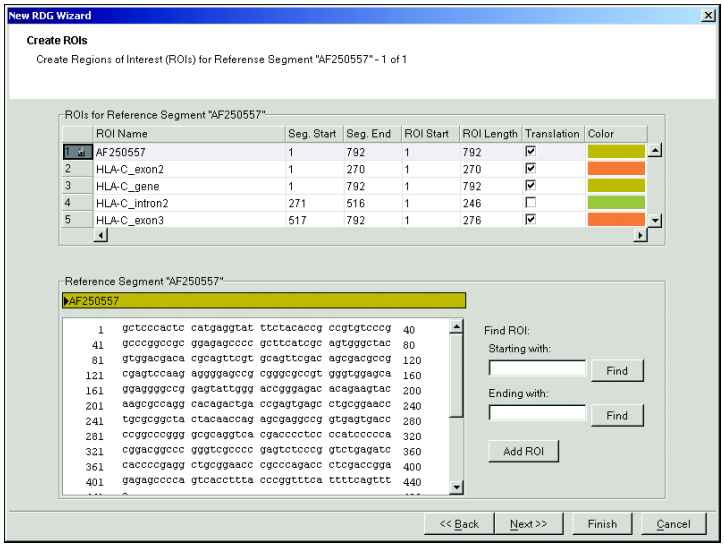
Note: For a procedure on using the Paste Ref. Segment button, refer to "Pasting a Reference Segment" on page 4-16; for a procedure on using the Split Ref. Segment button, refer to "Adding a Reference Break in a Sequence" on page 4-28.

The wizard continues the instructions to add a new layer and regions of interest (ROI) to that layer. An ROI is a region on a reference segment that defines exons, introns, splice junctions, and other features.

Setting Up the Reference Segment

To set up the reference segment:

1. Select the bases in the region of interest that you want to compare to the reference sequence (or backbone). In the Reference Segment pane, drag through the bases you want to select, or type the starting and ending bases under the Find ROI label.



2. Click **Add ROI** to add the segment to the ROI table in the ROI pane above the sequence. Add as many ROIs as desired.

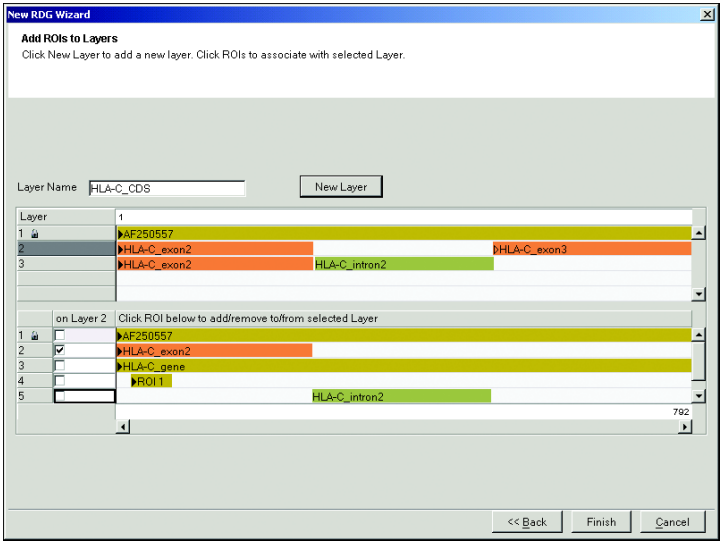
3. Click **Next**.

4. Follow the instructions to add layers and ROIs to layers. Layer 1 is always the reference sequence, which is generated by the software and is locked. Click **New Layer**, then name each layer that you add.

   IMPORTANT! To avoid confusion, give each layer that you add a unique name.

5. Click the new layer under the Layer label in the layer pane, then select the ROI **on Layer** check box in the ROI pane to associate it with the selected layer. Do this for each layer you create.

   IMPORTANT! In a layer, you cannot define ROIs that overlap one another.



6. Click **Finish**, or if you want to change any of the selections, click **Back**.

   The newly created RDG appears in the Reference Data Group list.

Section 4.3 Creating a New RDG Using SeqScape® Manager

In This Section

Before You Begin

You must have administrator or scientist privileges to create a new RDG using SeqScape® Manager.

Before creating a new RDG, make sure you:

- Download a GenBank file, a FASTA text file, or have a reference sequence that is stored on your computer
- Define on paper the ROIs, layers, and segments to compare to the reference sequence

Creating an RDG from SeqScape® Manager

Creating a Reference Data Group, requires that you:

- Import reference segments
- Create ROIs
- Create layers

To create a new RDG from the SeqScape® Manager:

1. In the main SeqScape® window, select **Tools** > **SeqScape Manager**, then select the **Reference Data Group** tab.
2. Click **New**.
3. In the General tab, enter a name in the Reference Data Group Name field.
4. Select a Codon table type and add comments, if desired.

5. Select the **ROI** tab.

IMPORTANT! Do not click OK. More steps are needed to set up the RDG.

Importing a Reference Segment

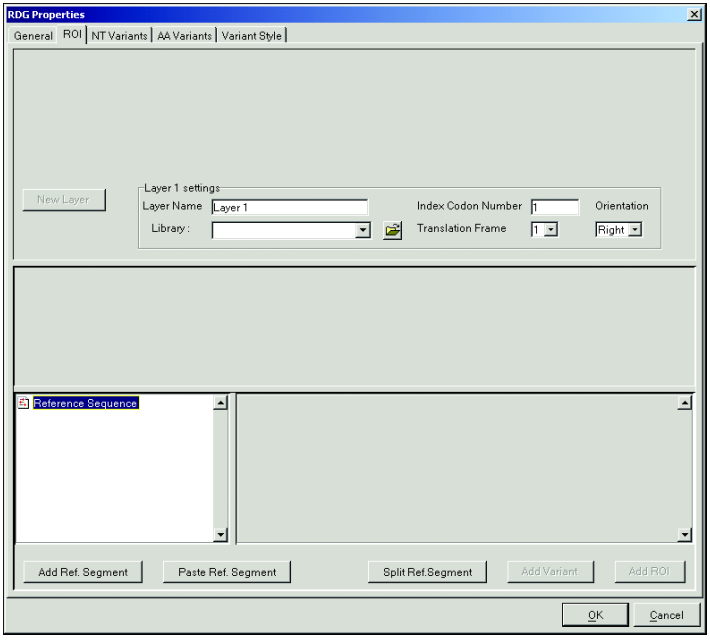
About the Reference Sequence

The reference sequence is made up of one or more reference segments that become a backbone or reference to which all other sequences or regions of interest are compared. After the reference sequence is imported into the RDG, it cannot be changed or edited.

To form the reference sequence, you need to import one or more segments.

To import a reference segment:

1. Select the **ROI** tab. The dialog box that opens shows Reference Sequence as a place holder in the lower left pane.



2. Click **Add Ref. Segment** in the lower left to add a segment to the reference sequence.

3. Navigate to the file containing the reference sequence. It can be a GenBank file or a file that you stored on your computer (the file may have a .gb extension).

4. Click **Import**.

   The reference sequence is on Layer 1, which is locked so it cannot be modified.

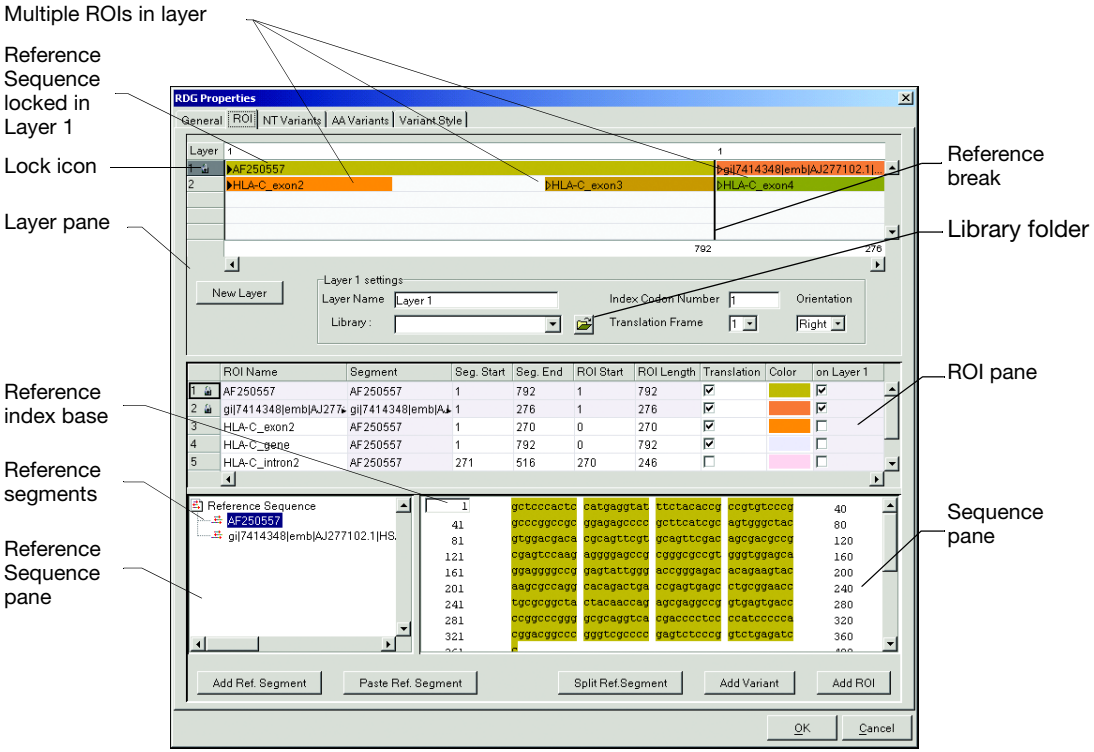


Figure 4-1 ROI tab in the RDG Properties dialog box

Defining Regions of Interest (ROI)

Defining an ROI

Each reference segment has its own locked ROI. On a sheet of paper, identify the ROIs you want to define, then use the information to define ROIs in the software. However, if you are using a GenBank file, the ROIs or features are already defined. You can add other ROIs where appropriate to your analysis.

To define an ROI:

1. In the ROI tab, select an empty layer or a layer where you want the ROI to appear.
2. Select a segment in the nucleotide sequence pane (by dragging through the region of interest), then click **Add ROI**.
3. Enter a name for the ROI under the ROI Name column in the ROI pane.
4. Define as many ROIs as appropriate by dragging through each region of interest, or by entering a number in the text box where the ROI should begin.

Pasting a Reference Segment

You can create or enter a sequence in a text editor or word processing program and copy the segment into the RDG at a later time.

To define a reference segment for copying and pasting:

1. Open a text file, then drag through the region of interest you want to use as a reference segment.
2. Select **Edit** > **Copy**.
3. In the RDG Properties ROI tab, click **Paste Ref. Segment** to use a reference segment that you copied to the clipboard. The copied reference appears in the Reference Sequence pane.
4. If you want to delete the copied reference segment, select it, press **Delete**, then click **OK** in the Confirmation dialog box.

Deleting an ROI, Layer, or Reference Segment

To delete an ROI, layer, or reference segment:

1. Select the ROI, layer, or segment.
2. Press **Delete**. Only unlocked rows can be deleted.

   IMPORTANT! After you delete an object, it cannot be undone.

Deleting a Reference Segment

When the RDG Properties window is open without being associated with a project, a reference segment can be deleted.

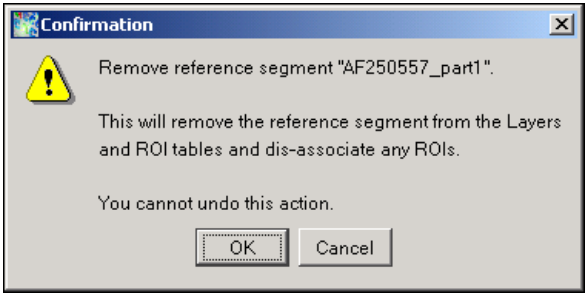
Note: The reference segments and Layer 1 cannot be deleted because they are in locked layers. They cannot be deleted when they are part of an existing open project.

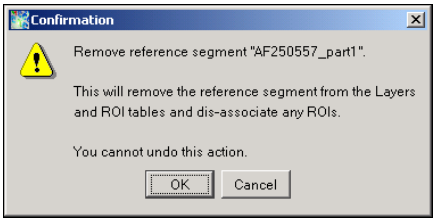
To delete a reference segment:

1. In SeqScape® Manager, select the **Reference Data Group** tab, then select the RDG in which you want to delete a reference segment.
2. Click **Properties**, then click the **ROI** tab.
3. In the Reference Sequence pane, select the reference segment you want to delete.

4. Right-click the selected segment, then select **Delete**.

   In the confirmation dialog box that opens, click **OK.** After you click OK, the delete cannot be undone.





ROI Tab Descriptions

Layer Pane Functions

The Layer pane in the ROI tab (see to Figure 4-1 on page 4-15) has the following functions:

- Layers – Shows the locked Reference Sequence in Layer 1 and the ROIs associated with each layer.
- New Layer button – Adds a new layer to the end of the layer table.
- Layer Number Settings – The settings of the selected layer. Each layer has its unique settings.
- Layer Name – The name of the layer, which can be edited.
- Library – Contains libraries to select if you are performing allele or haplotype identification. Before you select a library to associate, the Library field is blank. A library can be copied into the RDG, but is not associated until you select it from the Library drop-down list. After you select a library, the Library field shows the name of the library.
- Index Codon Number – The first amino acid number. This number is always in relation to the number of the first reference segment base, positive numbers only.
- Translation Frame – Sets the translation frame for the layer. The values are 1, 2, 3.
- Orientation – Sets the orientation of the layer, right (forward) or left (reverse).

The ROI Pane

The ROI pane has the following features:

- Clicking a row selects the ROI. When you select an ROI in the RDG, it selects and scrolls the reference segment and the associated sequence.
- Primary ROIs that are created when reference segments are imported are locked as indicated by the lock icon. These primary ROIs cannot be deleted from the ROI table, but can be deleted from the Reference Sequence navigation pane by right-clicking and selecting Delete.

Columns in the ROI Pane

The ROI pane in the middle of the RDG Properties dialog box has the following columns:

- **ROI numbers** – The number of the ROI. The Reference Sequence on Layer 1 is always locked. Reference *segments* that make up the Reference Sequence are also locked. Unlocked layers are below the reference segments and can be edited.
- **ROI Name** – Name of the ROI. ROI names that are not locked can be edited. The ROI Name must be unique.

  Note: Names for Reference Segments are not editable in the ROI pane. They can be edited in the Reference Sequence navigation pane by right-clicking and selecting Rename.
- **Segment** – Name of the segment to which the ROI is associated.
- **Seg. Start** – The nucleotide number in the Reference Sequence where the ROI begins.
- **Seg. End** – End of the ROI segment.
- **ROI Start** – The first nucleotide number you assign to this ROI. The number can be positive or negative.
- **ROI Length** – Length of the ROI. The value is automatically recalculated if you change the Segment Start or ROI Length values. Entering a number into this cell automatically recalculates the ROI Length value.
- **Translation** – Specifies whether or not the ROI is translated.
- **Color** – Shows the color of the ROI. Click to display the standard Color Picker dialog box if you want to select a different color for the ROI.

  Note: When an ROI is defined, a default color is applied to the ROI based on the name of the ROI.
- **On Layer (number)** – Check box. The label for this column changes based on the selected layer. If the check box is selected, the ROI appears on the selected layer. ROIs can be associated with multiple layers. However, ROIs cannot overlap on a layer. Therefore, the check box is disabled if the Start/Length range of the ROI overlaps with the range of an ROI already associated with the layer. This prevents you from overlapping ROIs on the Layer table. A dialog box appears if you try to select an overlapping ROI.

Creating a Library

About the Library

You must classify your library as a haploid or diploid library and determine how many library matches you would like to see for each consensus sequence. A library match is one allele or a pair of alleles that agree closely with each consensus sequence.

A haploid library contains sequences that have pure bases only (AGCT). When searching against a haploid library, SeqScape® software provides library matches, and each library match contains a pair of sequences (haplotypes) that best match the genotype of each consensus sequence.

A diploid library contains sequences composed of pure bases only, or pure bases and mixed bases. When searching against a diploid library, SeqScape® software provides library matches, and each match is a single sequence that best matches the genotype of each consensus sequence.

Refer to Appendix G, "Library and BLAST Searching."

Using Aligned FASTA Files

To use the library search feature, you must import an aligned multiple sequence FASTA file into the SeqScape® software. All sequences in the library must be of equal length. If some sequences are shorter than other sequences, you can use dashes (-) for missing bases.

Using a Tool to Align the Files

If you have a series of text sequences or electropherograms, you must create FASTA files, then use a tool to align those sequences and create a single multi-aligned FASTA file.

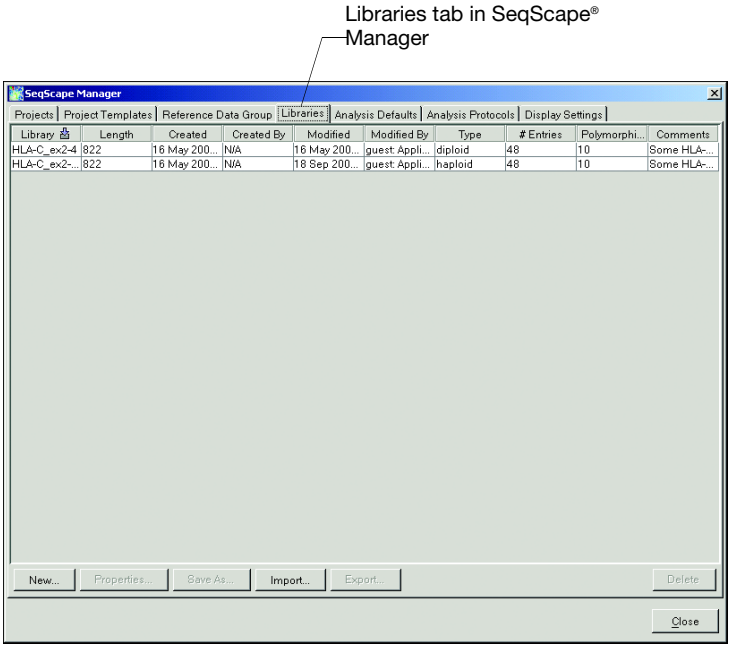
A common tool used to create aligned multiple-sequence FASTA files is Clustal X. For instructions on how to obtain and use Clustal X, see "About Creating a Multi-Aligned FASTA File" on page G-3.

Setting Up Your Library

Use this procedure to select the library before continuing with the procedure to create new layers.

To set up a library:

1. In the main SeqScape window, select **Tools** > **SeqScape Manager**, then select the **Libraries** tab.



2. Select **New**.

3. In the Library Editor General tab, enter a name for the new library, then select **Haploid** or **Diploid**.

4. Select the **Entries** tab, then click **Import**.

5. Import the aligned multiple sequence FASTA file, then click **OK**.

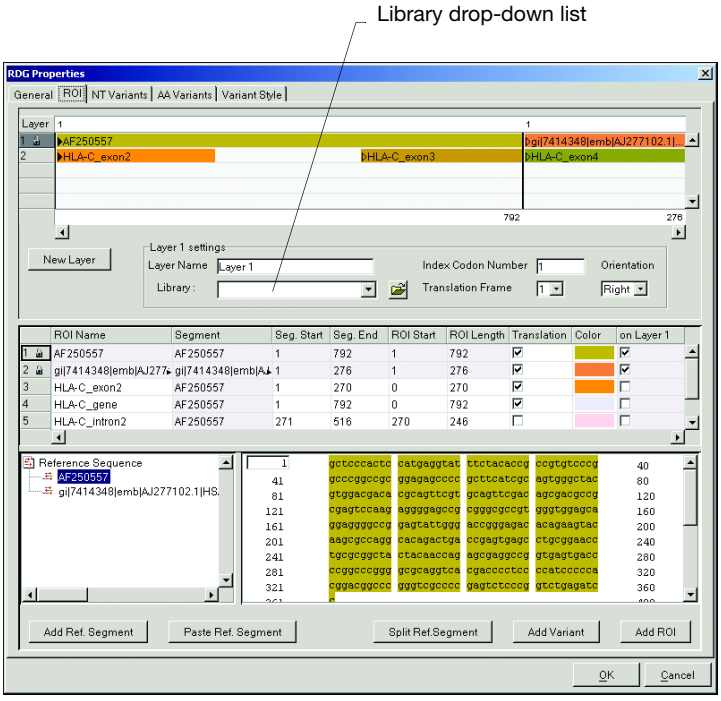
6. In SeqScape® Manager, select the **Reference Data Group** tab, then select the RDG that you want to link to the library

7. Click **Properties**, then select the **ROI** tab.

   Note: At this point, if you do not have layers in the RDG or you do not know how to create a layer, go to "Creating New Layers" on page 4-25. Otherwise, continue to the next step.

8. Select a layer in the Layer pane.

9. In the Library drop-down list, select the corresponding library that you created in steps 2 through 5.



Creating New Layers

Layers organize groups of related, nonoverlapping ROIs. By organizing ROIs into layers, results reviewing and library searching are faster and more focused. The Layer table shows the organization of ROIs into layers.

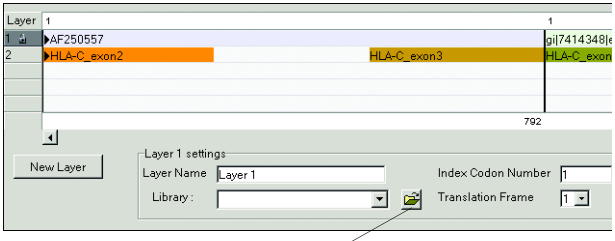
To create new layers:

1. In the RDG Properties dialog box, select the ROI tab, then in the Layer pane, click **New Layer** and enter a name in the Layer Name field.

2. Select a layer by clicking it under the Layer label in the Layer pane. If you need more information on libraries, refer to "Creating a Library" on page 4-21.

   a. Select a library from the Library drop-down list if you are performing allele or haplotype identification.

   b. Click the library folder icon to open the Library Editor and view the entries.



Note: The selected library in the Library drop-down list is associated with the layer in the Layer Name field.

The Library Editor opens as shown in the sample below.

c. In the Layer pane, enter the Index Codon Number.

d. Select the Translation Frame.

e. Select the Orientation.

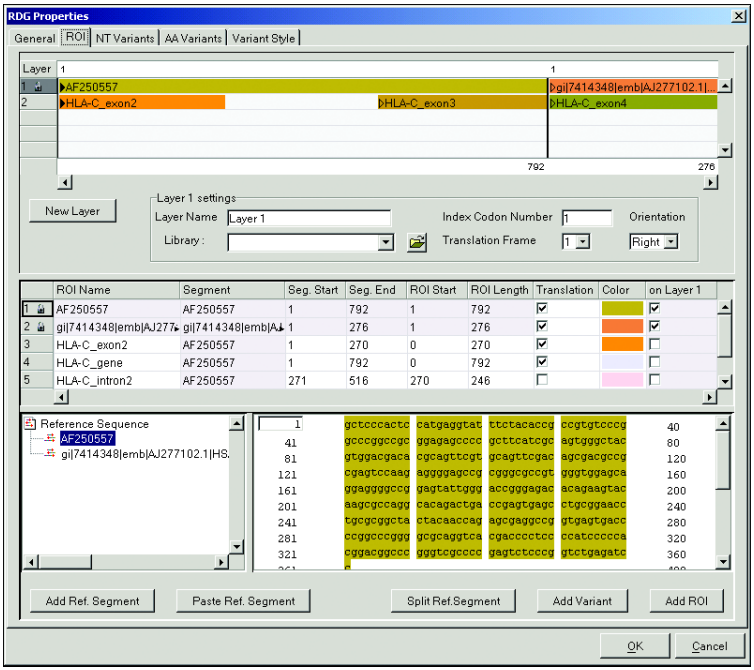
3. Select the appropriate Reference Segment in the Reference Sequence pane, select the sequence representing the ROI, then click **Add ROI**.

   The ROI is added to the ROI pane and to the selected layer.

4. Repeat the process to build layers containing all the ROIs and layers you previously recorded on paper.

5. You can edit the ROIs in the ROI pane by selecting the attributes, then editing them directly in the table.

6. To include an existing ROI on an unlocked layer, select or create the layer, then select the **On Layer** check box for the ROI.



IMPORTANT! If you want to add variants, follow the procedure on page 4-30. Do not click OK. If you do not want to add variants, go to the next step.

7. When you finish adding ROIs and layers, click **OK**. The new RDG appears in the Reference Data Group list.

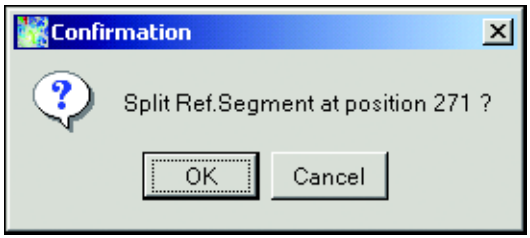
Adding a Reference Break in a Sequence

You can add a reference break in the Reference Sequence if you want to delete intervening reference sequences. When reference segments are split, the ROIs associated with the reference segment are also split.

To add a reference break:

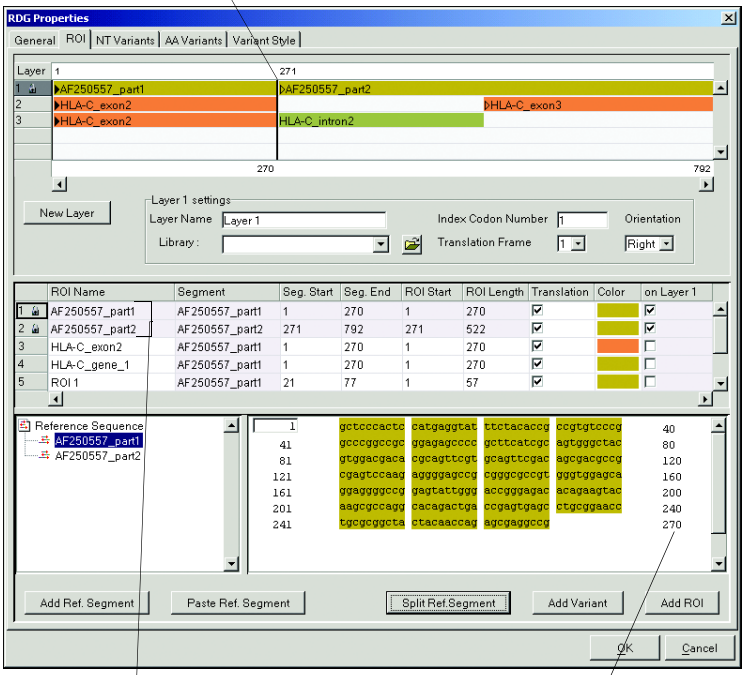
1. In the ROI tab Sequence pane, select the base position where you want a split to occur, then click **Split Ref. Segment**.

2. A confirmation dialog box opens showing the position for the reference break. Click **OK**.



There is a new reference break in the Reference Sequence in the Layer pane, as shown in the next screen shot. The Reference Sequence shown in the ROI Layer is now in two locked layers, one segment ending at position 270, and the second segment starting at position 271. The Sequence pane shows the first segment ending at position 270.

New reference break at position 271



Two segments in the Reference Sequence

First segment ending at position 270

Declaring Variants into an RDG

About NT Variants

The NT Variants tab in the RDG Properties dialog box lists the known nucleotide variants associated with a reference sequence. The entries you define in this tab are used to identify known and unknown variants in your projects.

You can enter NT variants by:

- Clicking Add Variant in the ROI tab, then entering the variant attributes in the New NT Variant dialog box.
- Creating a table of variants in a tab-delimited format, then saving the file and importing it into the NT variant file.

  One way of creating a table of variants is by using Microsoft® Excel. The columns in the Excel table must map to the columns in the NT Variants tab as shown below.

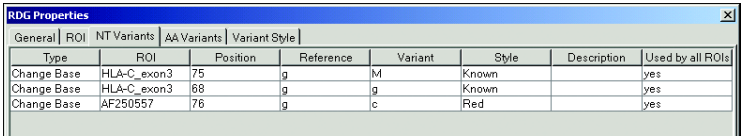


Figure 4-2 NT Variants tab showing table column names

- Importing an aligned FASTA file.
- Selecting a sequence within a reference segment, then clicking Add Variant. This procedure is described below.

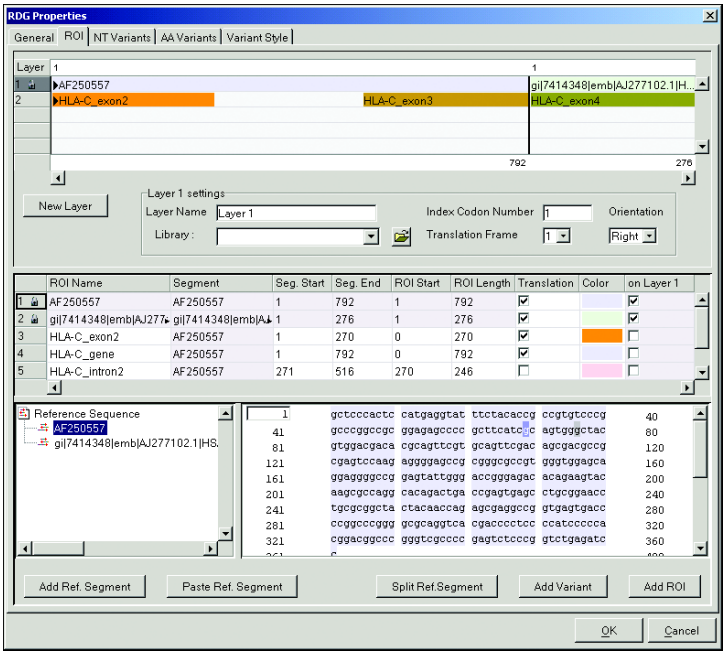
Creating New NT Variants

To create new NT variants:

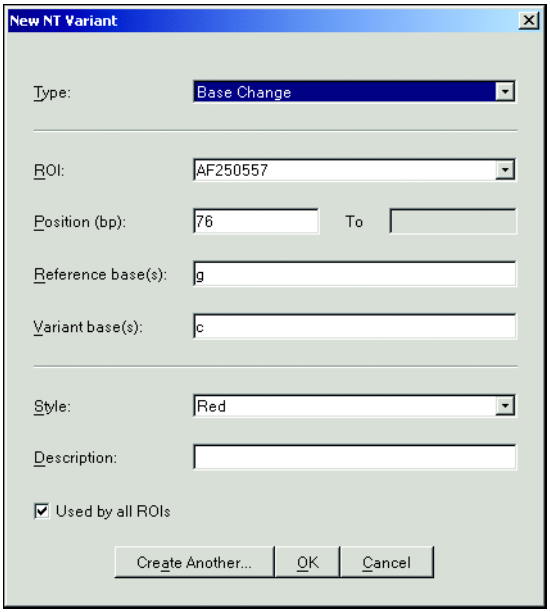
1. In the SeqScape® Manager, select the **Reference Data Group** tab, then click **Properties**.

2. Select the **ROI** tab.

3. Drag to select a sequence in the nucleotide sequence area of the tab, then click **Add Variant**.



4. In the New NT Variant dialog box, select the type of variant: **Insertion**, **Deletion**, or **Base Change**.



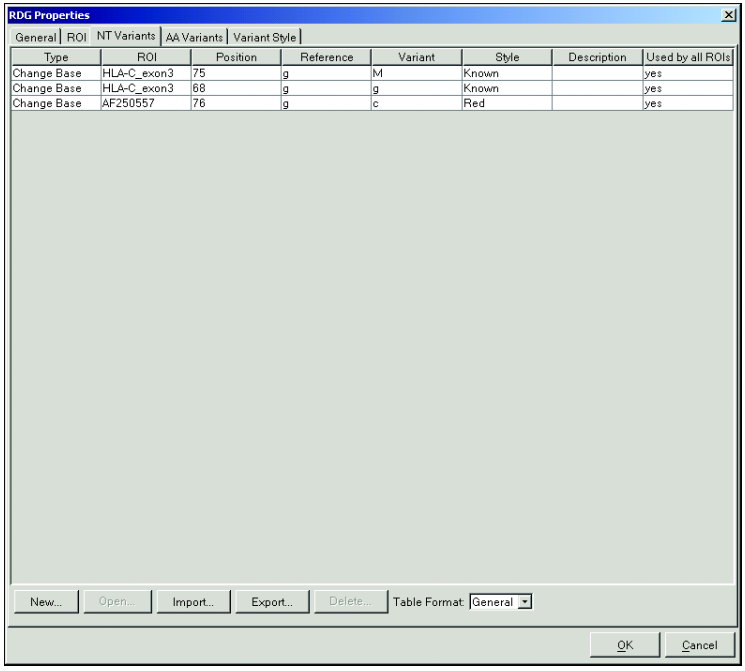
5. Enter the Variant base.

6. If desired, change the style and enter a description.

7. Select the **Used by all ROIs** box if this NT variant is to be used by all ROIs.

8. Click **Create Another**, or **OK** to save the changes.

   After you click OK, the variant additions appear in the list in the NT Variants tab.

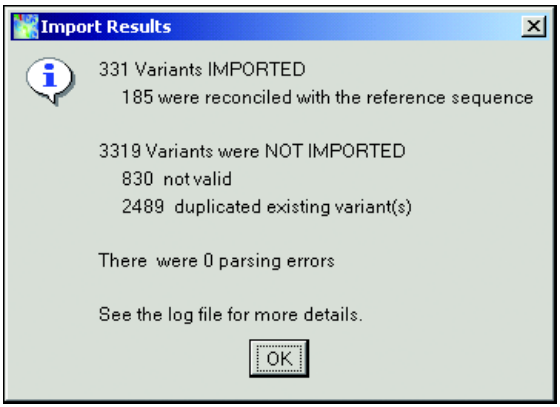


Importing NT Variants in Tab-Delimited Format

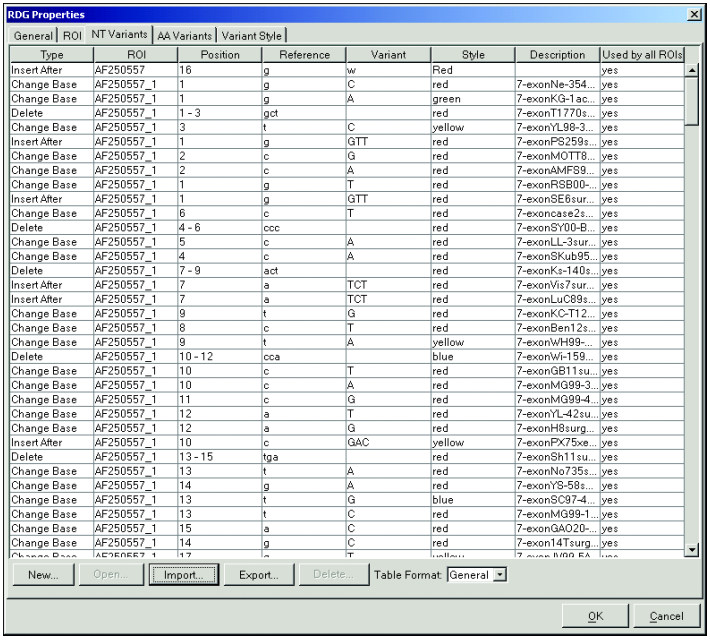
To import an NT variant from a tab-delimited NT variant file:

1. In the SeqScape® Manager, select the **Reference Data Group** tab.
2. Select the RDG in the list for which you want to import NT variants.
3. Click **Properties**, then select the **NT Variants** tab.
4. Click **Import**.
5. Navigate to the tab-delimited NT variants file, then click **OK**.

6. An Import Results dialog box opens to show the number of variants imported, as shown in the sample below. Click **OK** to close the Import Results dialog box.



7. The new variants appear in the NT Variants list. The Table Format options at the bottom of the window are General (default) and Hugo. If desired, select the format in the drop-down list.



8. Click **OK** to save the imported variants and close the RDG Properties window.

Creating an RDG from Aligned Consensus Sequences

About Creating an RDG

SeqScape® software will create a new reference sequence and variants from a set of aligned sequences imported into a blank RDG that contains no reference sequence. The file format of the imported aligned sequences must be in FASTA text. For more information on FASTA format, see Appendix F, "Aligned Variant and FASTA File Format."

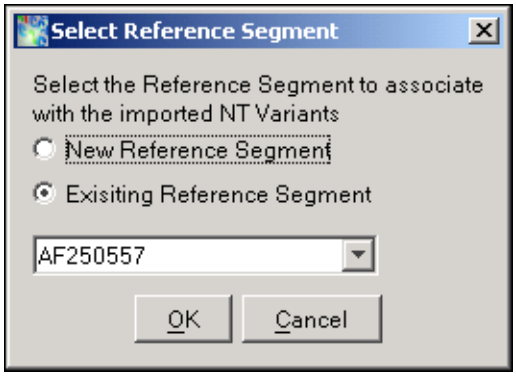
SeqScape® software uses the first sequence in the set of aligned sequences in the FASTA file as the reference. The rest of the sequences are evaluated relative to that first sequence to derive variants. Any positions that differ from the first sequence are used to populate the variants table.

Importing NT Variants from an Aligned FASTA File

To import NT variants using an aligned FASTA file:

1. Select **Tools > SeqScape Manager**.
2. Select the **Reference Data Group** tab, then select the RDG for which the variant will be added.
3. Click **Properties**.
4. In the RDG Properties window, select the **NT Variants** tab.
5. Click **Import**.
6. In the Import NT Variants dialog box, navigate to then select an aligned sequence FASTA file (.fsta extension).
7. Click **Import**.

8. In the Select Reference Segment dialog box, select the reference segment for which the variants are to be added.

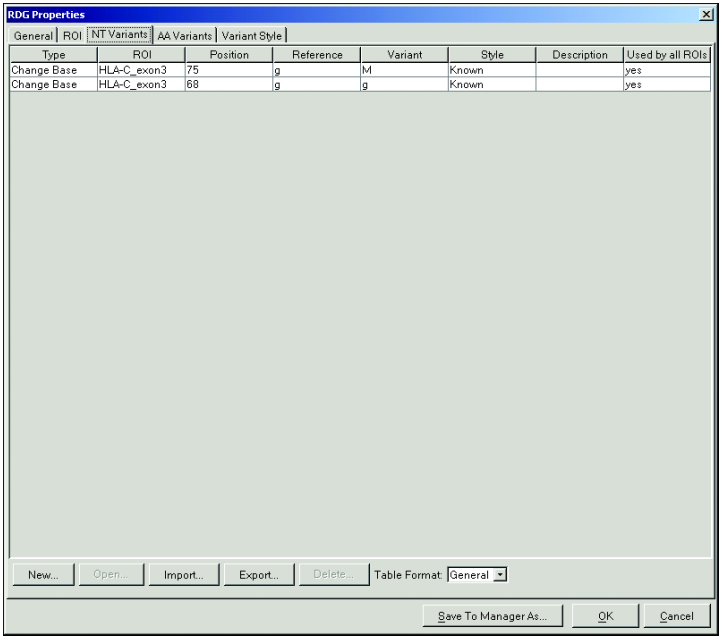


9. Click **OK**.

   After the data are imported, the Import Results dialog box opens, displaying information about the import.

   The first sequence in the imported file populates the reference. The subsequent sequences are used to derive variants by comparison to the first sequence. These variants appear in the Variants table.

10. Click **OK** to close the Import Results dialog box. The list of variants is displayed in the NT Variants tab.



11. Click **OK** to close the RDG Properties dialog box.

Entering New AA Variants

The AA Variants tab lists the known amino acid variants associated with a reference sequence. The entries you define in this tab are used to identify known and unknown amino acid variants in your projects.

You can enter AA variants in two ways:

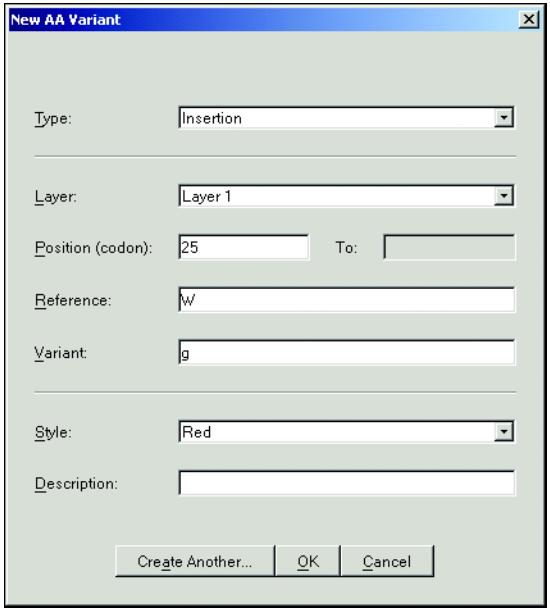
- Click Add Variant in the ROI tab, then enter the variant attributes in the New AA Variant dialog box.
- Create a table of variants using Microsoft® Excel, then import the table. The columns in the Excel table must map to the columns in the AA Variants tab. Refer to Figure 4-2 on page 4-30 for the column names.

To enter a new AA variant:

1. In SeqScape® Manager, click the **Reference Data Group** tab.
2. Select a listed RDG, then click **Properties**.

3. Select the **AA Variants** tab, then click **New**.

4. Select the type of variant (**Insertion**, **Deletion**, or **Residue Change**).



5. Enter the Position (codon) in the reference sequence that you want changed. The Reference appears after you enter the position in the sequence.

6. Enter the Variant.

7. Select a color style and enter a description, if desired.

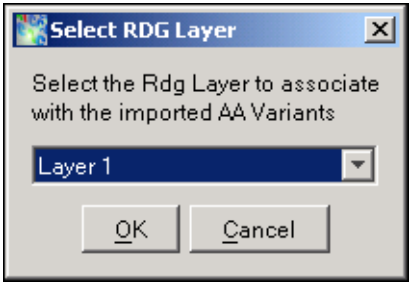
8. Click **OK**. The new variant appears in the AA Variants list.

9. Click **OK** to save the new variant.

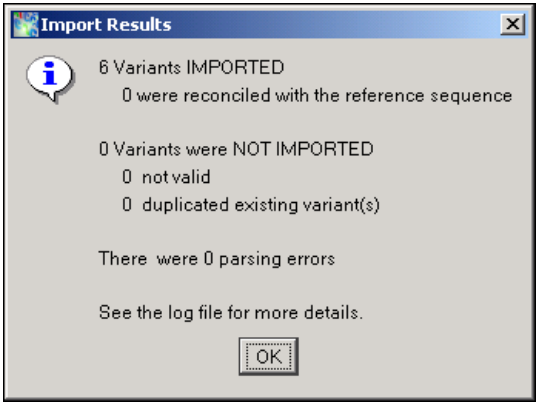
Importing AA Variants

To import an AA variant from a tab-delimited file:

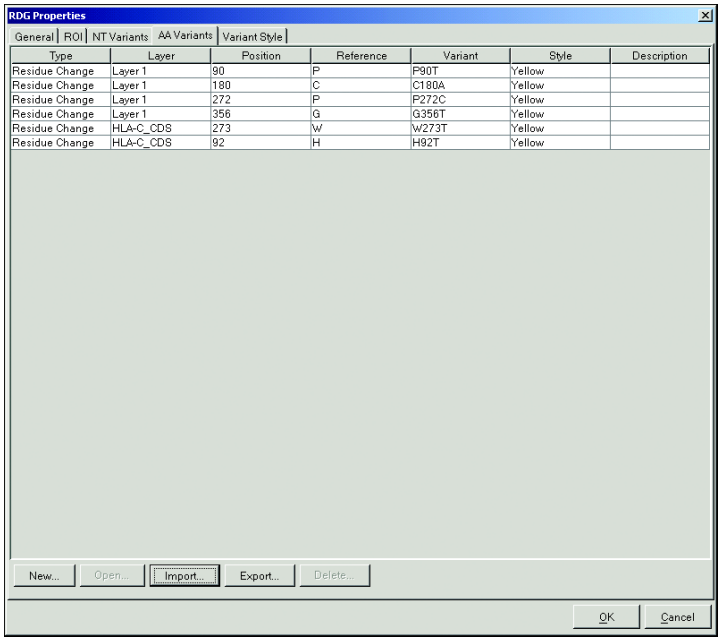
1. In SeqScape® Manager, select the **Reference Data Group** tab, then select the RDG you created.
2. Click **Properties**, then select the **AA Variants** tab.
3. Click **Import**, then navigate to the variant data file. It can be a tab-delimited text file (.txt file).
4. Click **Import**.



5. Select any layer from the drop-down list, then click **OK**.
6. Click **OK** in the Import Results dialog box.



The amino acid variants are imported and appear in the list in the AA Variants tab. A sample of AA variants is shown below.



Assigning Styles to Variants

Use the Variant Style tab to assign styles to the variants as desired. The Variant Style tab allows you to define text coloring styles that identify different types of variants and change the display characteristics of variants in the Project view.

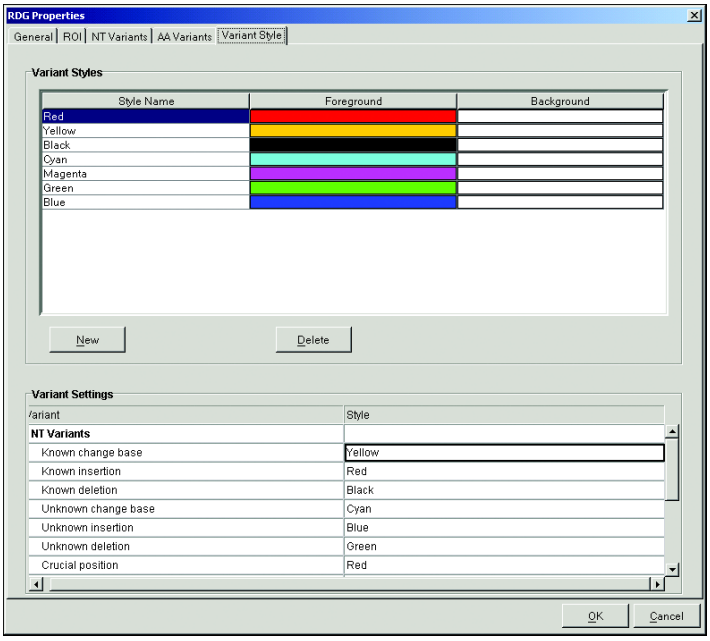
The table at the top of the dialog box displays the generic styles. The table at the bottom of the dialog box lists the different types of variant conditions and their associated styles. The styles you set appear in the Project view to identify the different types of variants.

To assign styles to the variants:

IMPORTANT! When assigning color to text, select light background colors so the text is easy to read.

1. In the RDG Properties dialog box, select the **Variant Style** tab.

   The Variant Styles pane shows the available default colors of the variants.



2. Select the colors in which you want the base changes, insertions, and deletions for known variants to display.

   a. To add a new color and style, click **New**.

   b. To name the variant style, click the **Foreground Color** box, select a new color in the color palette, then click **OK**.

   c. Select a color from the color palette for the **Background Color**, then click **OK** in both dialog boxes to set the new variant style.

   The variant styles you set appear in the Project view to identify the different types of variants.

3. To delete a color, select the color, then click **Delete**.

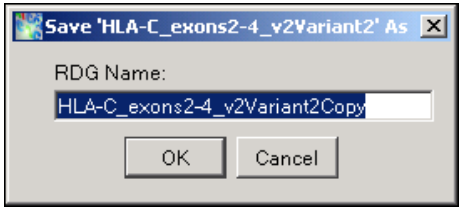
   Note: The first seven Foreground colors cannot be changed or deleted.

4. In the Variant Settings pane, select the colors in which you want to display the base changes, insertions, and deletions for unknown variants. The Variant Styles area shows a list of the available default colors.

Saving a Copy of the RDG

To save a copy of the RDG:

1. In SeqScape® Manager, select the RDG you want to save.
2. Click **Save As**.
3. When the confirmation window opens, rename the RDG or click **OK**.



Saving the RDG for Other Projects

If you are working with an RDG that is embedded in a project or project template, you can save a copy of the RDG into SeqScape® Manager. This is useful if you make edits to an RDG and want to reuse the RDG for other projects.

To save the RDG:

1. With the project open, in the RDG Properties dialog box, click **Save To Manager As**.
2. Enter a name for the RDG, then click **OK** to save a copy of the RDG under a new name.

If you accept the default name, a copy of the original RDG is saved with the default name and is available to use with another project.

Save To Manager As Button

Use the Save To Manager As button to save copies of project elements so that you can import the elements into other projects to change a project template.

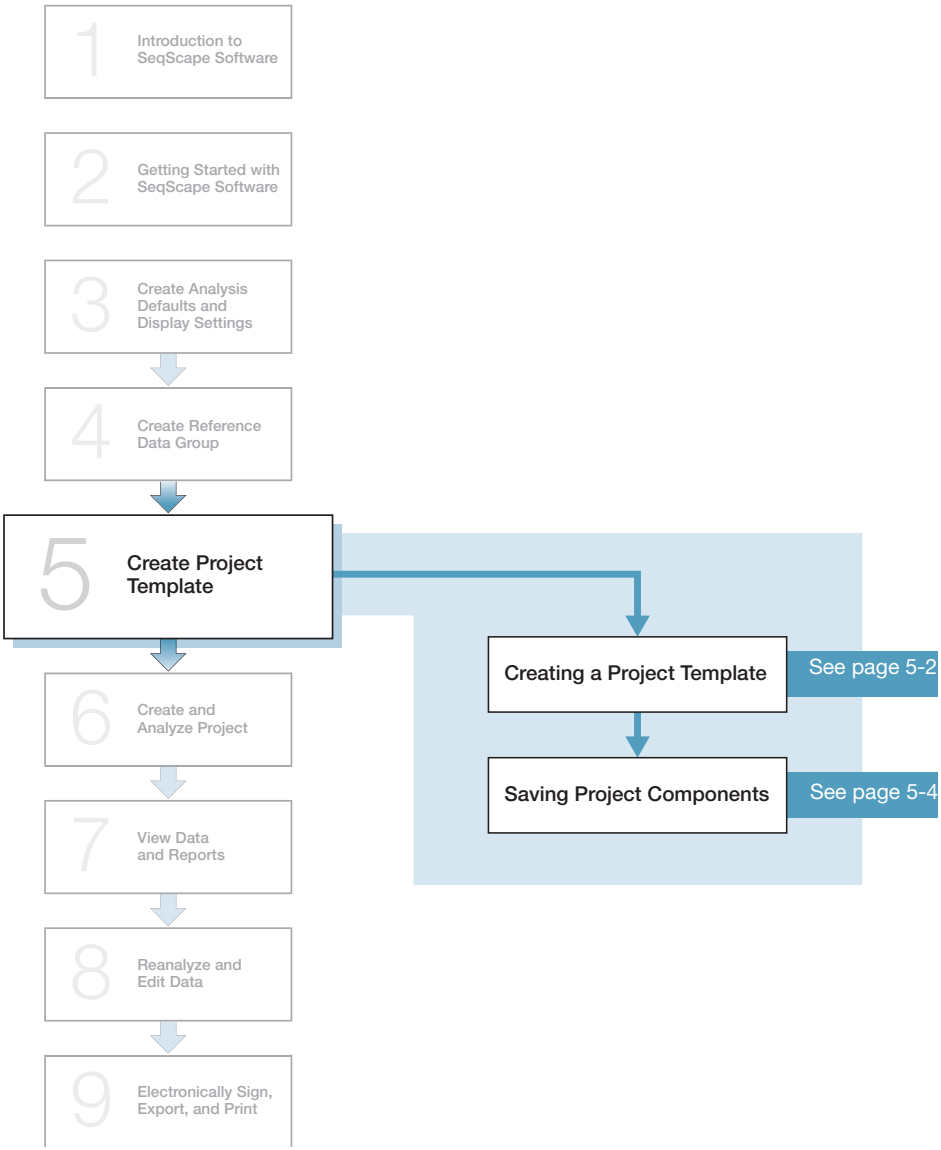
For all tabs in each dialog box in the Analysis menu, RDG Properties, Analysis Defaults, and Display Settings have a Save To Manager As button.

To use the Save To Manager As button:

1. With the project open, select the **Analysis** menu then select **RDG Properties**, **Analysis Defaults,** or **Display Settings**.
2. Select any tab in any of these three dialog boxes, then click **Save to Manager As**.
3. In the Name field, enter a new name, or accept the default nameCopy, then click **OK**.

   The saved copy is available to import into another project.

5 Creating a Project Template



Creating a Project Template

Before you can effectively use SeqScape® software, you must create and configure a project template. A project template contains all the reference data and settings needed to analyze your data automatically. It defines how the software analyzes and displays your samples. When project templates are created in the SeqScape® Manager, they can be imported, exported, and edited.

About Creating a New Project Template

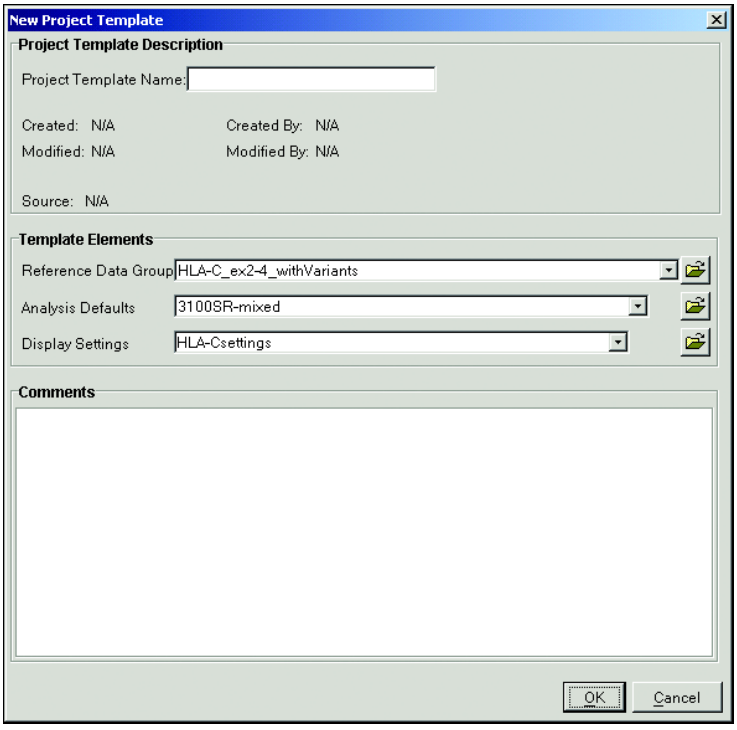
When you create a new project template from the SeqScape® Manager, you select:

- **Reference Data Group** – Reference sequence and associated data to which all the specimens in a project are compared. See "Creating a New RDG Using SeqScape® Manager" on page 4-13 for more information.
- **Analysis defaults** – Settings that are used to analyze the data. See "Specifying the Analysis Settings" on page 3-11 for more information.
- **Display settings** – Settings that are used to display the data. See "Specifying Display Settings" on page 3-16 for more information.

Creating a New Project Template

To create a new project template:

1. In the SeqScape® window, select **Tools > SeqScape Manager**.
2. Select the **Project Templates** tab, then click **New**.



3. Enter a name for the project in the Project Template Name field.

   Note: The project template name must contain only characters that conform to the Windows file system. Refer to "File-Naming Convention" on page 2-10 for a list of all invalid characters.

4. Select the desired Template Elements from the drop-down lists, then click **OK**.

Saving Project Components

About Saving Template Components

If you modify RDG, analysis, or display settings within a project, the changes are valid only in that one project. However, if you want to save those settings so they can be applied to other projects, you can create new SeqScape® Manager template components based on existing template components.

Saving Template Components from Within a Project

To save project template components:

1. Within a project, select the **Analysis** menu, then select one of the template components that you want to modify:
   - **RDG Properties**
   - **Analysis Defaults**
   - **Display Settings**
2. Make the desired modifications to the component.
3. Click **Save To Manager As**.

   An appended name of the current template component appears in the Save.xx to the SeqScape® Manager As dialog box.
4. Leave the name unchanged or change it.

   IMPORTANT! You cannot save over an existing template component. You must delete the existing master component from the SeqScape® Manager before you can save a new template component.
5. Click **OK**.
6. To use the modified component for other projects, make a new project template that uses the new components.

Examples of Changing the Settings Within a Project

Example 1

1. Create a project template and apply it to a project.
2. Select **Analysis** > **Analysis Settings**, then change a sample analysis setting.

   The underlying Analysis Defaults are unchanged in the SeqScape® Manager.

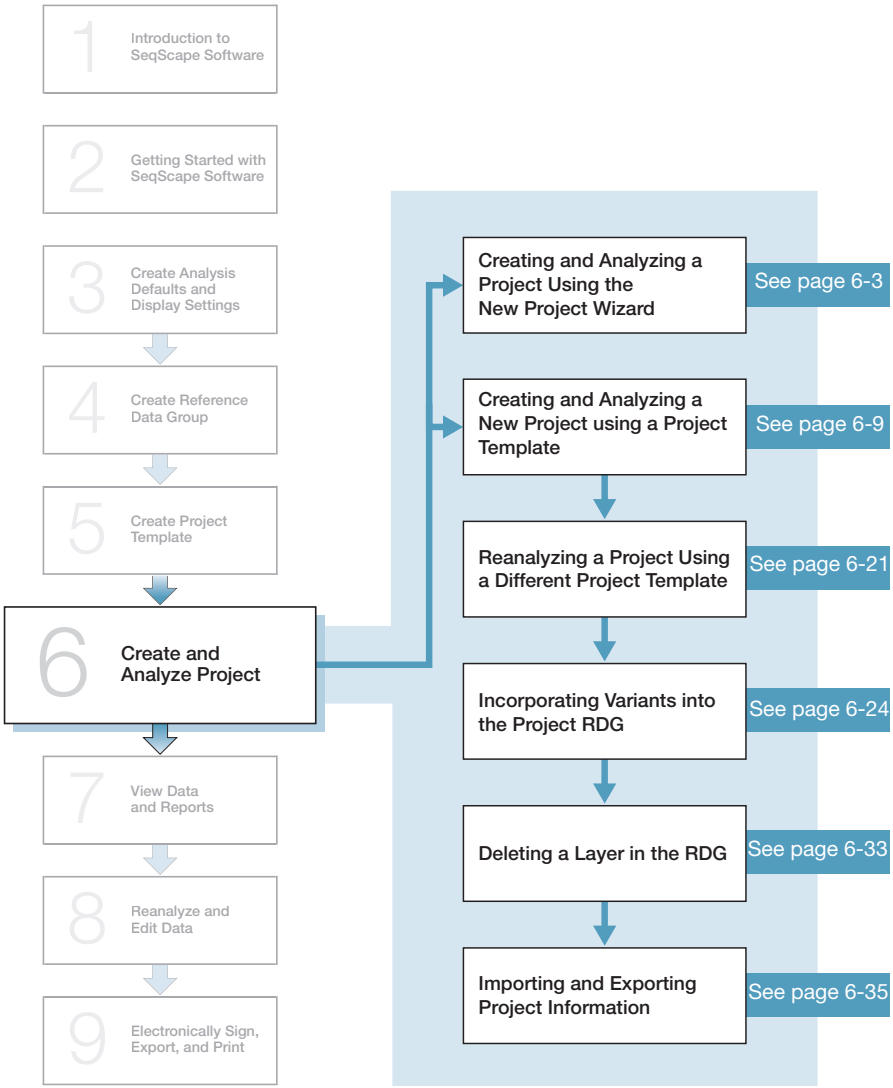
Example 2

1. Create a project template and apply it to a project.
2. Modify each component of the template.
3. Change a variant style in the RDG, then select **Save To Manager As**.

   A new RDG in the SeqScape® Manager reflects this change, but the old RDG in the SeqScape® Manager remains unchanged. Therefore, the project template using the old RDG is also unchanged.

Note: In both examples, the open project displays the changes.

6 Creating and Analyzing a Project



About Projects

Before Creating a Project

Before you can create a project in the SeqScape® Software Version 3, you must have created a project template that contains:

- A Reference Data Group
- Analysis Defaults
- Display Settings

Components of a SeqScape® Project

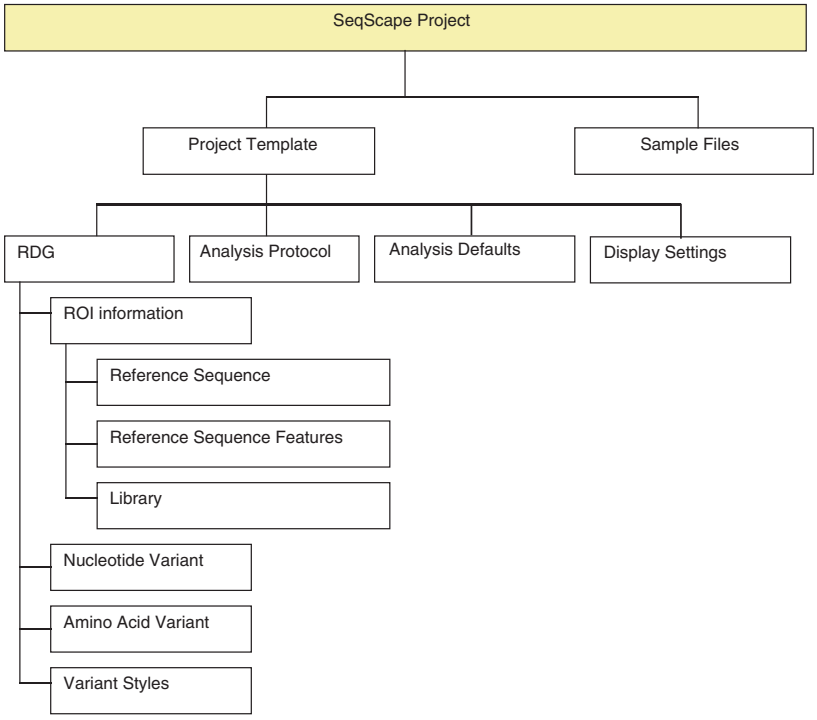


Figure 6-1 Project components

Ways to Create and Analyze a New Project

After the analysis defaults are set up, you can create a new project for data analysis by:

- Using the New Project wizard (see page 6-3)
- Using an existing project template (see page 6-9)

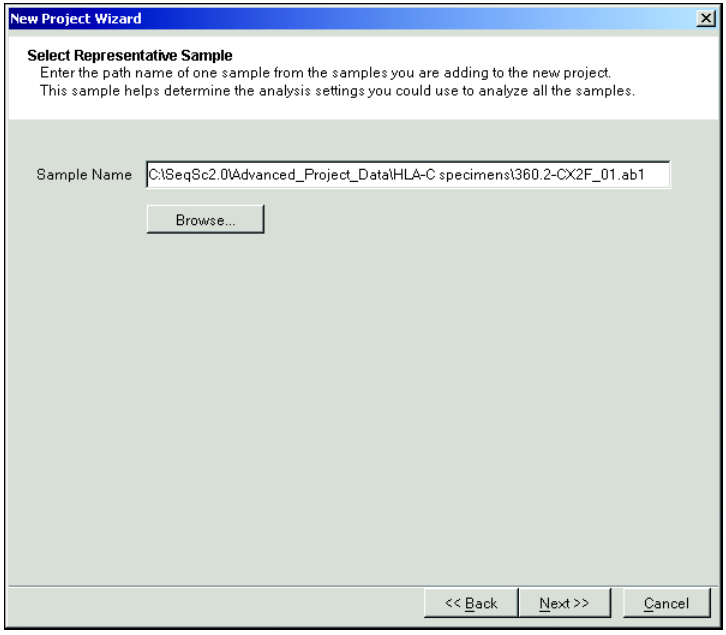
Section 6.1 Creating and Analyzing a Project Using the New Project Wizard

The New Project Wizard

The New Project Wizard takes you through the process of setting up and analyzing a new project.

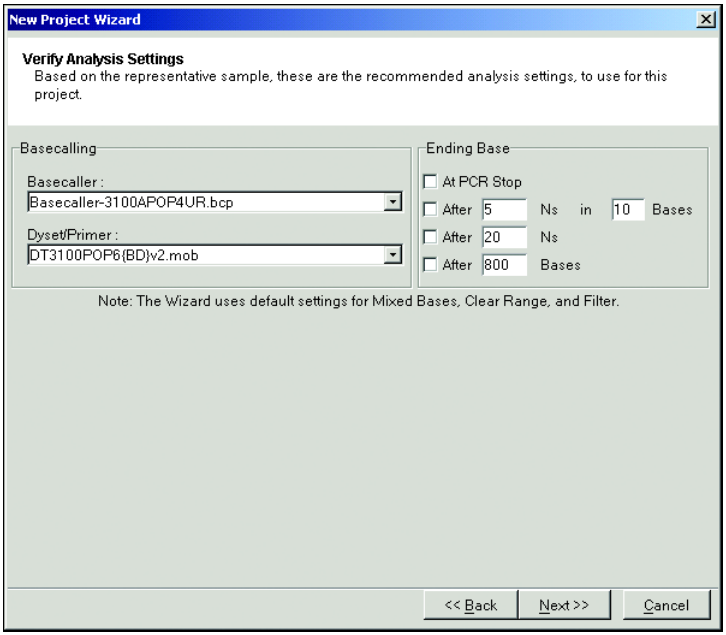
To create a new project using the New Project wizard:

1. Launch SeqScape® software.
2. Select **File** > **New Project Wizard**.
3. Enter a name for the new project in the Project Name field, then click **Next**.
4. Enter a sample name, or click **Browse,** then navigate to the sample you want.



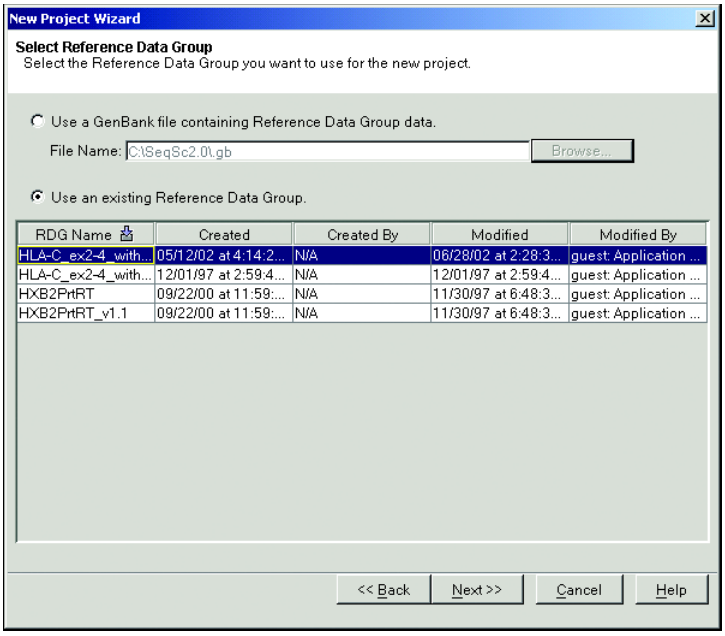
5. Select a sample that has the .ab1 extension, then click **Open**.

6. Click **Next**. The wizard uses analysis settings based on your sample choice.



7. Verify the analysis settings (Basecaller, DyeSet/Primer files, and Ending Bases), then click **Next**.

8. In the Select Reference Data Group page:



a. Do one of the following:

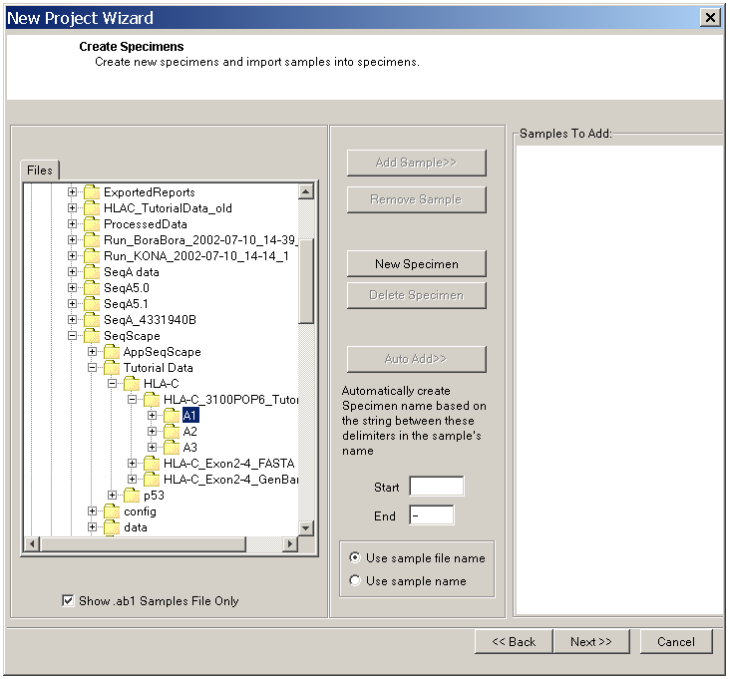
- Select **Use a GenBank file containing Reference Data Group data**, then specify a GenBank file.

  Or,

- Select **Use an existing Reference Data Group**, then select a Reference Data Group file in the list.

b. Click **Next**.

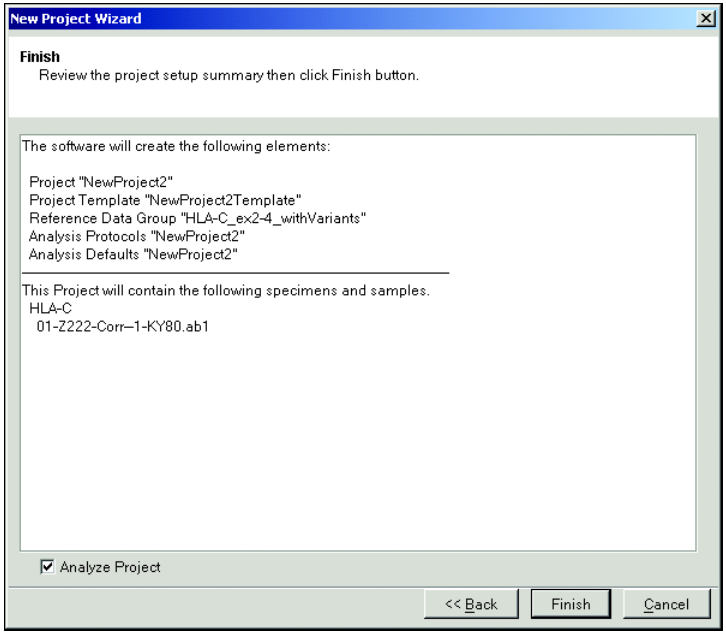
9. Add specimens and import samples in the Create Specimens page:



a. In the Files section, select a sample, multiple samples, or a folder, then click **Auto Add**.

b. Click **Next**.

Note: For information on adding specimens, see "Adding Specimens and Importing Data into a Project Overview" on page 6-10.



When you use the Project Wizard for the first time, master display settings are created. These same settings are used if the wizard is used again.

10. Review the setup. Click **Back** to change the setup, if necessary.

11. Do one of the following:

To analyze …	Then …
Now	Click **Finish**.
Later	1. Deselect **Analyze Project** at the bottom left corner of the page. 2. Click **Finish**.

12. When you close the new project, click **Yes** to save it.

This project is now available in the list of available projects in the SeqScape® Manager.

Section 6.2 Creating and Analyzing a New Project Using a Project Template

In This Section

You can use an existing project template to create a new project. For convenience, one example project template is included in the software. To create your own project template, see "Creating a Project Template" on page 5-2.

Project Template Included

Table 6-1 Components of the project template included in SeqScape® software:

Template Component	File Name
Project Template Name	• HLA-3100_POP6mixed_v2 • p53_exon7-v2
Reference Data Group	• HLA-C_exons2-4_noNT_v2 • p53_Exon7_v2
Analysis Defaults	• 3100-SR_POP6_BDTv1_mixed_v2 • 3700LR_POP5_BDTv1_v2
Display Settings	• DefaultDisplaySettings_v2

Creating a New Project Using a Template

To create a new project using a project template:

1. Select **File** > **New Project**.
2. When the New Project window opens, select a template from the list and enter a project name.

   Note: To see the whole name in the list, click-drag the Project Template heading to the right when the double-headed arrow cursor appears on the column bar.

3. Click **New**.

   The new project using the selected template opens.

Adding Specimens and Importing Data into a Project Overview

All sample data from a single biological source should be placed inside a specimen within a project. All sample data inside a specimen are assembled, and a consensus sequence is produced. You can think of each specimen as holding the assembled samples from one PCR product, for example. The consensus that is generated is compared to the references and aligned to the other consensus sequences from other specimens.

If you have new, unanalyzed data, you need to create specimens in the project to hold the data. You can add specimens to a project automatically or manually.

For more information on what types of data can be imported into a project, see "Adding Specimens and Importing Samples Manually" on page 6-14.

IMPORTANT! Unanalyzed specimen and sample data show a red slash line through their icons, indicating that analysis is needed.

IMPORTANT! Specimen names can be edited only after they are imported. Sample names cannot be edited from within SeqScape® software at any time.

Adding Specimens and Importing Samples Automatically

Using a text delimiter, SeqScape® software simultaneously and automatically creates specimens and imports unanalyzed or analyzed samples into a project.

Sample IDs and Sample Filenames

To take advantage of this feature, your sample ID (which you assign to the sample in the data collection software and which is stored within each sample file) needs to have the same prefix for all samples in each specimen.

IMPORTANT! You *cannot* modify the sample ID (name).

The sample filename is longer than the sample ID, and often is derived from the sample ID. The sample filename is what you see when looking for the sample. You *can* modify the sample filename.

The text delimiter is chosen from the sample ID. Using this function, a set of sample files that are grouped into the same folder and that share a similar delimiter can be imported into their corresponding specimens in a single step.

In the example shown in Figure 6-2, the delimiter is a dash. Everything to the left of the delimiter determines the specimen name. When you select Add Automatically, the sample files are automatically transferred into specimens that are also created and named automatically. In this example, the first specimen includes all files that start with A1.

The sample ID also appears in the Annotation view of the sample, as indicated in Figure 6-2.

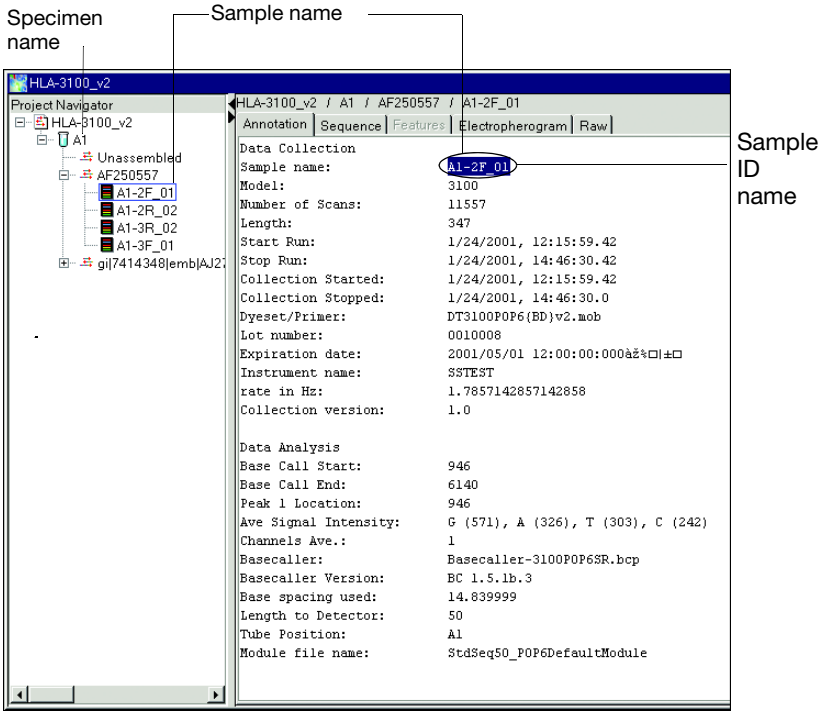


Figure 6-2 Annotation tab showing sample name example

Creating a Specimen

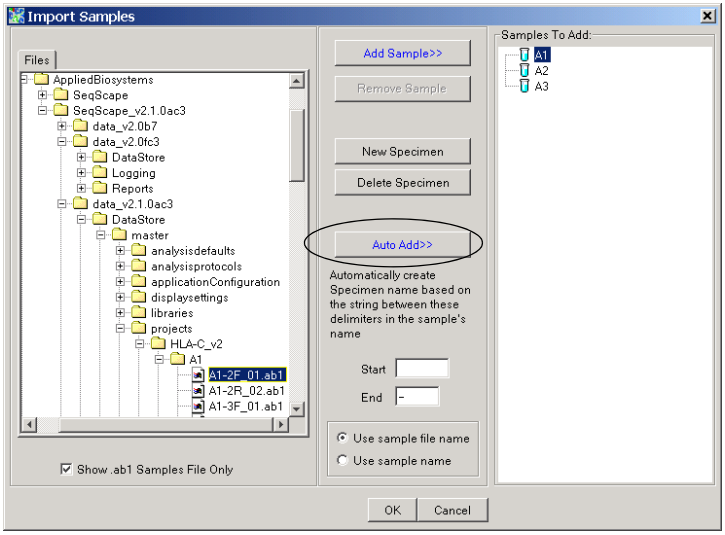
To create a specimen and import samples automatically:

1. With the Project window open, select **File > Import Samples To Project** or click  .

2. In the Specimen name delimiter field, enter the delimiter text.

   Note: The delimiter text is derived from the sample ID name in the data collection software sample sheet or plate record. In the figure in step 3, the delimiter is a dash. The sample ID name from the data collection software appears in the Sample name section of the Annotation view of the sample.

3. Select the folder containing the samples to be imported, then click **Auto Add**.



Based on the text delimiter, the samples are automatically imported into the appropriate specimens (in this example, the specimens are shown under HLA-C specimens).

4. Click **OK** to import the specimens and samples into the project.

Adding Specimens and Importing Samples Manually

You can import the following types of sample data into specimens within a project:

- Sample data files from ABI PRISM instruments
- Database files
- Specimen text-only files

To import …	See …
Sample data files	"Adding Specimens and Importing Data Files" on page 6-14
Specimen text-only files	"Importing Text-Only Files" on page 6-19

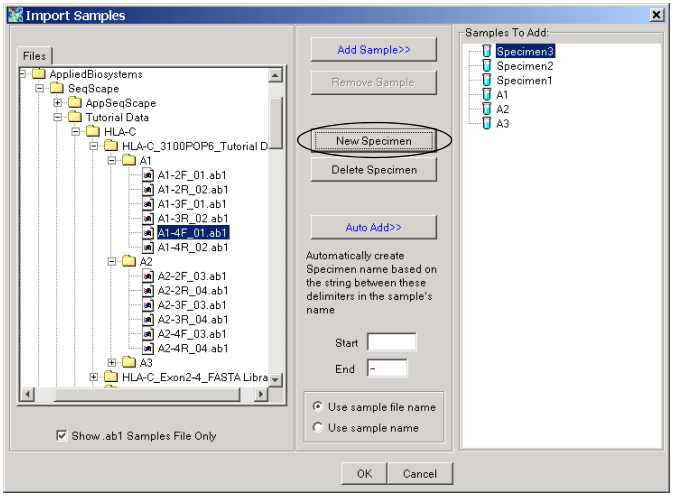
Adding Specimens and Importing Data Files

To import unanalyzed or analyzed sample data, the files must be in ABI format. Sample data is imported into specimens in the project. New specimens are created in the Import Samples dialog box.

To add specimens and import sample data files:

1. With the Project window open, select **File > Import Samples To Project**, or click  to open the Import Samples dialog box.

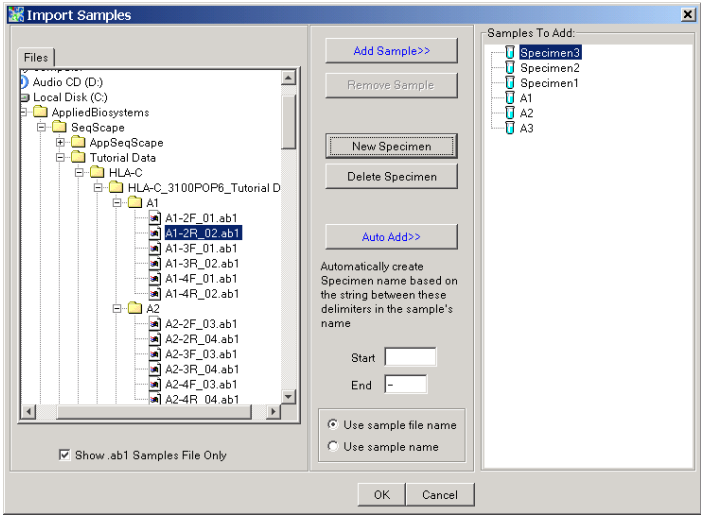
2. Create a new specimen:
   a. Click **New Specimen**.
   b. Add two more specimens.



3. In the Samples To Add section on the right, select the specimen into which to import the data.

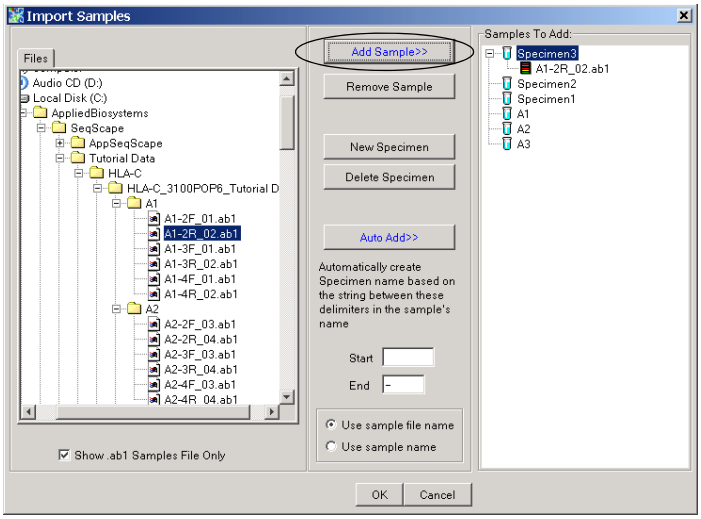
4. In the Files pane, navigate to the samples you want to add.

5. Select the first specimen in the Samples to Add pane.



6. In the Files section, select the sample data files.

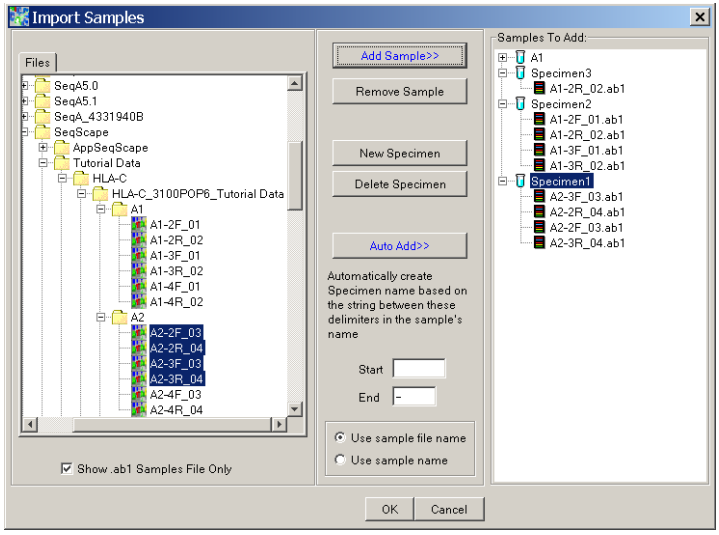
To import …	Then …
A single sample	Select the single sample.
Multiple samples	**Ctrl+Click** to select contiguous or noncontiguous samples.
All samples in a folder	Select the folder.



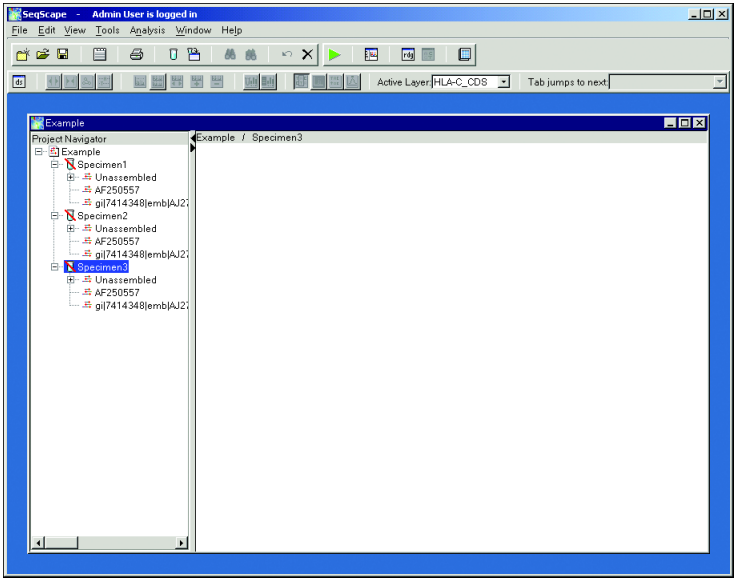
7. Click **Add Sample**.

   The sample data appear in the selected specimen, showing where the data will be imported. No data are imported into the project until you click OK.

8. Select the second specimen, select the samples, then click **Add Sample**. Repeat this for the third specimen.

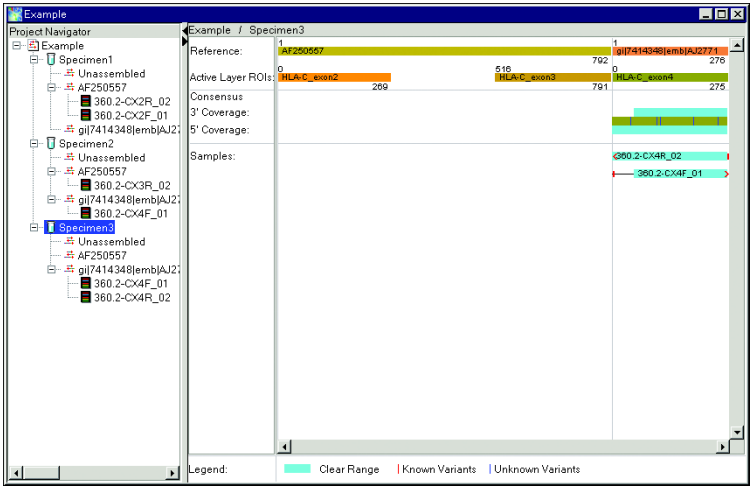


9. Click **OK** to perform the imports and return to the Project window. The project reflects the new specimens and samples. Specimens shown with a red line through are unanalyzed and unassembled.



10. If desired, select each specimen, type a new name for the specimen, then press **Enter**.

11. Click the green arrow button (Analyze) on the toolbar, which indicates that the samples need to be analyzed. After the samples are analyzed, the red line through the specimen disappears and the samples are assembled, as shown in the figure below.



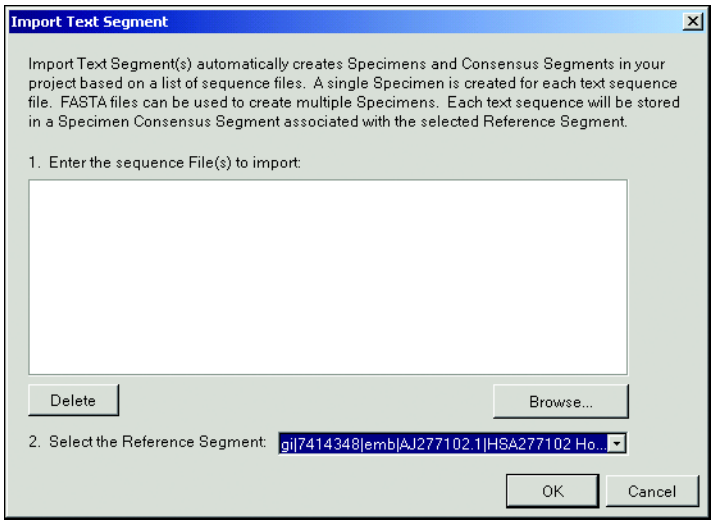
12. Save the project then close it.

Importing Text-Only Files

You can import into a project a consensus sequence in text format as a text-only specimen.

To import text or previously assembled sequences:

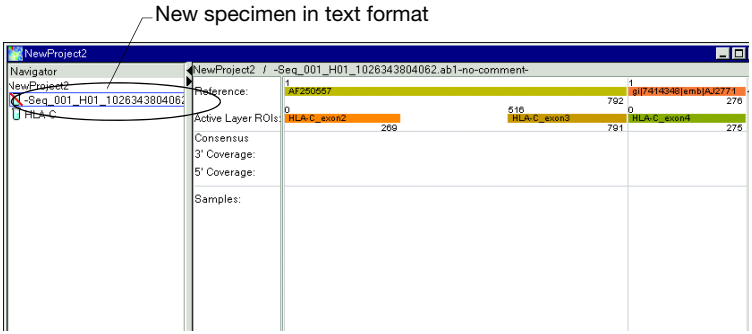
1. In the Project Navigator, select the project name, then select **File > Import Text Segment**.



2. Click **Browse**, then navigate to and select the target segment.
3. In the Import Text-Only Segment dialog box, select the text file (.fsta format), then click **Import**. The segment appears in the previously blank section in the Import Text Segment dialog box.
4. Repeat steps 2 and 3 to add additional text segments.

5. Click **OK**.

   A new specimen is created with the name specified in the first line of the file.



Removing Samples or Specimens

To remove samples or a specimen from a project:

1. In the Project Navigator of the project, select the samples or specimen you want to remove from the project.

2. Press the **Delete** key.

   IMPORTANT! This deletes the results and cannot be undone. If you press Delete in error, close the project without saving to restore the results.

3. In the Confirm Delete dialog box, click **Yes**.

Analyzing the Data

After you import all your data, you can run the analysis. After new data are imported or analysis settings are changed for a sample, the Analyze icon in the toolbar appears green, indicating that there are unanalyzed data.

To run an analysis in the project, click (Analyze), or select **Analysis > Analyze**.

Section 6.3 Reanalyzing a Project Using a Different Project Template

In This Appendix

When You Would Want to Do This

After you analyze an entire project that contains many samples, you may want to reanalyze all the data using a project template that contains different settings or reference data.

Saving a Project Before Reanalyzing

IMPORTANT! Applying a new project template to an existing project overwrites all analyzed project data, including basecalls, features, alignments, and manual edits. To avoid overwriting the data, you can rename the project to keep your original analysis.

To save a project that you want to reanalyze:

1. Select **Tools** > **SeqScape Manager**.
2. In the Project list, select the project that you want to save before reanalyzing.
3. Click **Save As** and rename the project.
4. Click **OK**.

   The project is saved under a new name and your original project remains in the list.

Applying a New Template to an Existing Project

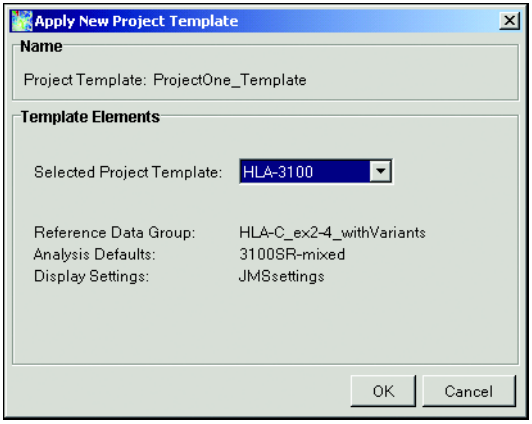
To reanalyze a project with a different project template:

1. Create a template containing the desired changed settings and/or reference sequence (see "Creating a New Project Using a Template" on page 6-9).

2. Open the existing project that has the data that was analyzed using the old settings.

   IMPORTANT! Save the project under a new name if you want to keep the current project data to compare to the new project data. If you do not save the project, all the data are overwritten when you apply a new project template.

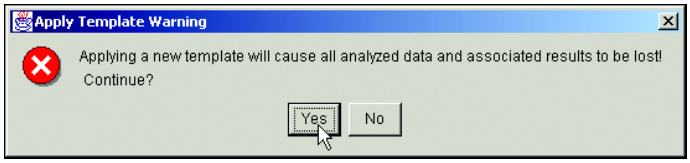
3. Select **Analysis > Apply Project Template** to open the Apply New Project Template dialog box.



4. In the Selected Project Template drop-down list, select the project template that you want to apply to the project.

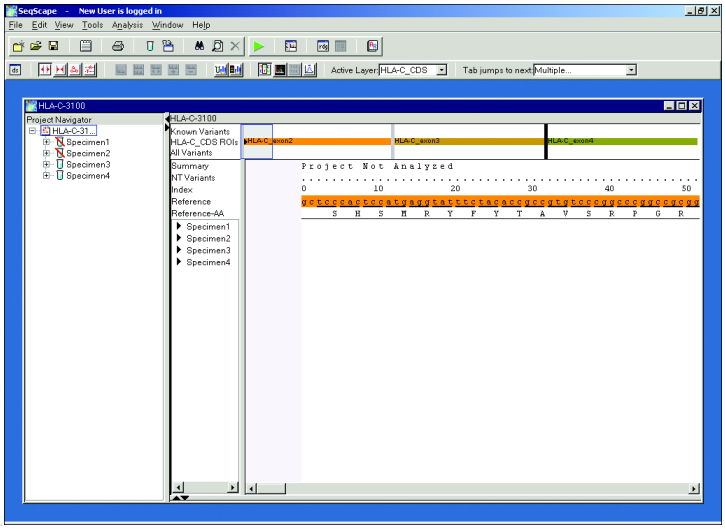
5. Click **OK**.

6. A dialog box opens warning you that all analyzed data and results will be lost. To continue, click **Yes**.



The project opens, containing all the specimens and samples, but the data are unanalyzed.

7. To analyze the data with the new template, click  (Analyze) or select **Analysis > Analyze**.



Incorporating Variants into the Project RDG

About Incorporating Variant Sequences

You can incorporate variants into an active project RDG by one of the following:

- Changing an unknown variant in a specimen to a known variant
- Adding a variant
- Importing a file containing variant sequences
- Importing a set of variants from a TXT file

Note: If you have a master RDG and want to include additional variants in the RDG, you must incorporate them using the SeqScape® Manager.

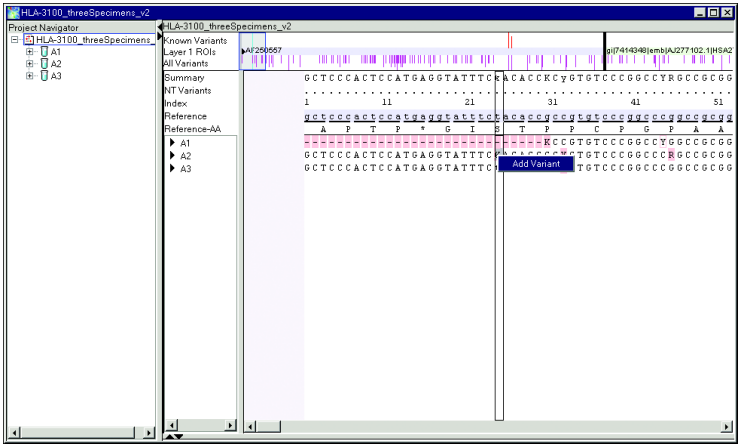
You can incorporate variants automatically by importing a file of a tab-delimited text file of variant positions and descriptions. By this method, variants are created and styles are applied to all the variants in the file.

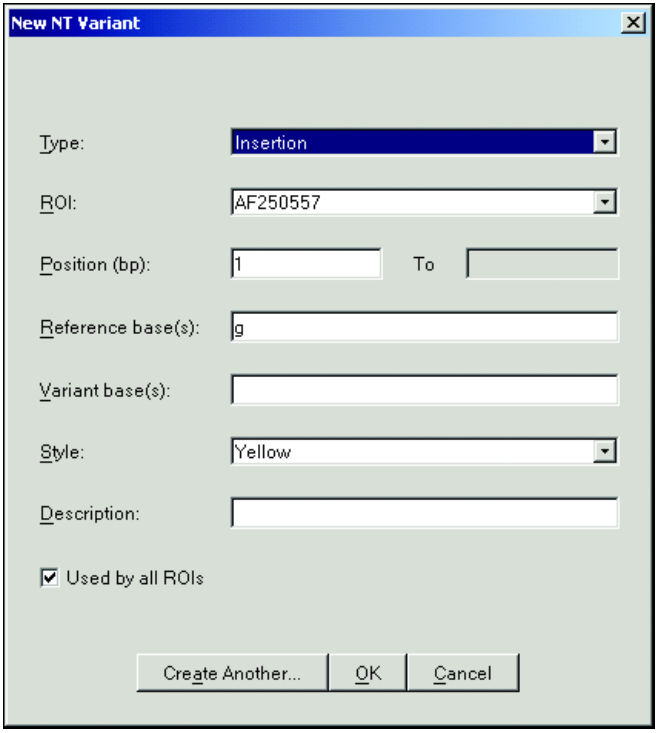
Alternatively, you can change an unknown variant in a specimen to a known variant, or you can create variants by adding them to the Reference Data Group.

Changing a Single Unknown Variant to a Known Variant

To change an unknown variant to a known variant:

1. Select a variant base in a specimen, right-click the base, then select **Add Variant**. The New NT Variant dialog box opens, displaying the type and position of the variant.





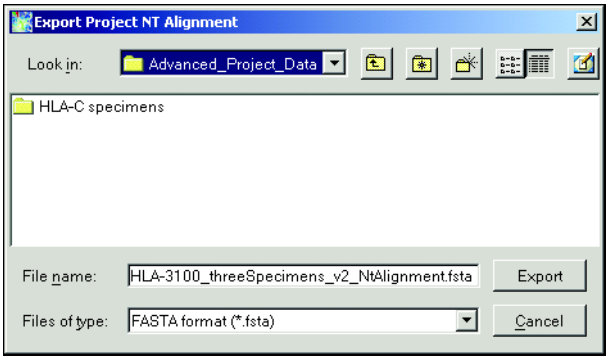
2. Select a variant style from the Style drop-down list.
3. In the Description field, enter text, if desired, then click **OK**.
4. Repeat steps 1 through 3 for another variant.

Changing Multiple Unknown Variants

Changing multiple unknown variants to known variants, requires that you export unknown variants in a project alignment file and then import them into the project.

To change multiple unknown variants to known variants:

1. Open the project, then select the specimen containing the unknown variants.
2. Select **File > Export > Project Alignment-Nucleotides**.

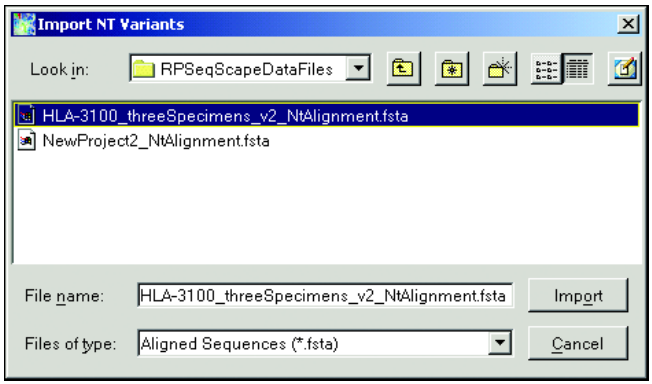


3. In the Export Project NT Alignment dialog box, select a destination for the exported data, then click **Export**.

Importing Variants

To import variants:

1. Select **Analysis > RDG Properties**.
2. Select the **NT Variants** tab.
3. Click **Import**.
4. In the **Import NT Variants** dialog box, navigate to then select the project alignment file. Make sure Files of type is set to **All Files** or **Aligned Sequences**.



5. Navigate to or select the file to import.
6. Click **Import**.
7. In the Select Reference Segment dialog box, select the reference segment in the drop-down list then click **OK**.

   The variants appear in the NT Variants table as Known variants. The descriptions are the specimens in which the variants appear and the style is the default style for the variant type.

8. Select the **Variant Style** tab in the RDG Properties dialog box to change the default style in the RDG and enter a description of the imported variants.
9. Select the **NT Variants** tab to be sure the variants are Known.
10. Click **OK** to save the variants.

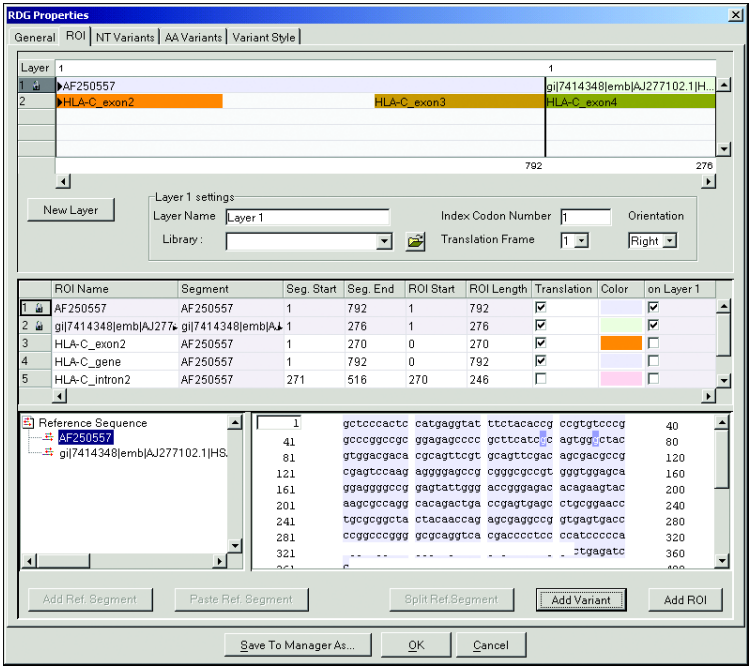
Creating a New Variant in a Project

You can add a variant to a project by:

- Entering the type and position of the variant in the Variants tab of the RDG Properties dialog box.
- Selecting the location on the reference sequence in the Sequence tab in the RDG Properties dialog box. The appropriate information regarding the variant is automatically entered in the variant dialog box.

To create a new variant in the project:

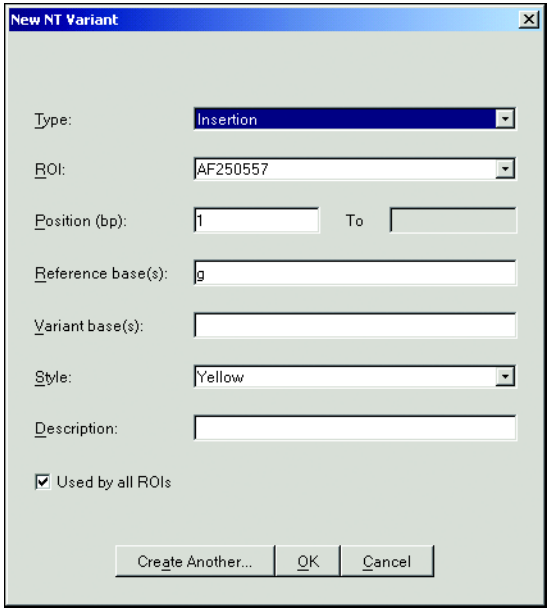
1. In the Project window, select **Analysis > RDG Properties**.
2. Select the **ROI** tab, then select **Add Variant**.



3. In the New NT Variant dialog box, select the type of variant (**Base Change**, **Insertion**, or **Deletion**).

4. Select the Position and either the **To** (position) or **Variant base**.

   Note: The Reference base is entered by the software based on the position.



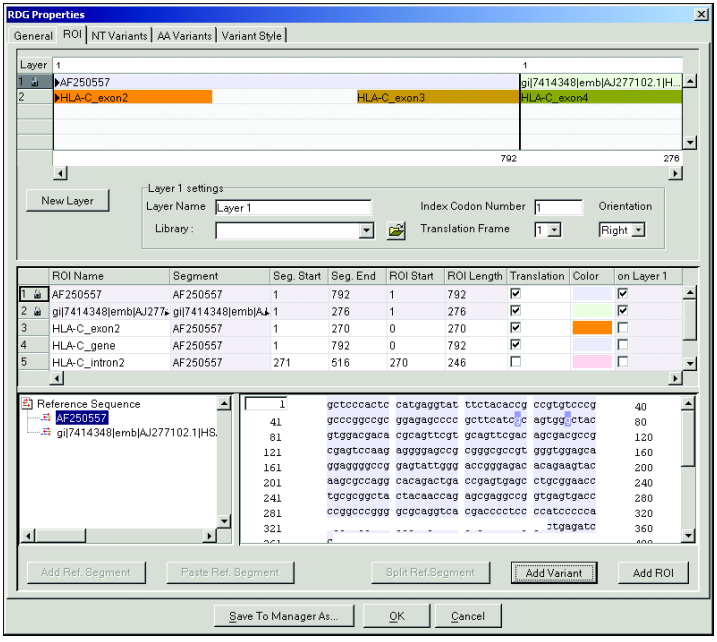
5. Select a style (color) in which you want the variant to be displayed, then enter a description of the variant, if desired.

6. Click **Create Another** to add more variants, or click **OK** to save the variant to the RDG.

Adding a Variant in the Project

To add a variant in the project:

1. In the Project window, select **Analysis > RDG Properties**.
2. Select the **ROI** tab.
3. Indicate your variant by doing one of the following:
   - Select the base that corresponds to the substitution variant or range of bases for a deletion variant.
   - Click the position at which you want an insertion variant.
4. Select **Add Variant**.



5. In the New NT Variant dialog box, note that the Position and the Reference base are already entered.
6. Select the type of variant by clicking **Base Change**, **Insertion**, or **Deletion,** then enter the Variant base.
7. Select a style for the variant, then enter a description of the variant, if desired.

8. Click **OK** to save the variant to the project.

Importing Variants to the Project

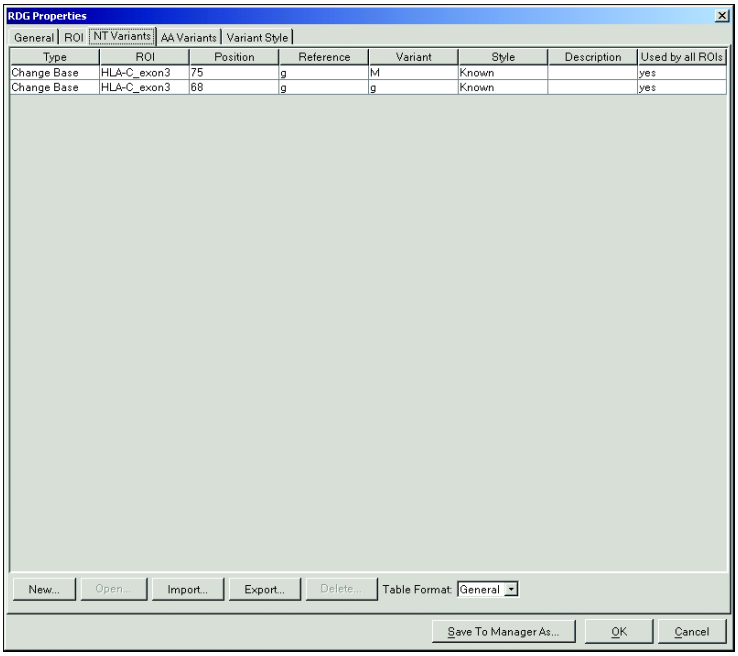
When you import variants into a project, they must be in one of the following configurations:

- Tab-delimited text file format
- Text file format containing aligned sequences

To import variants into a project:

1. In the Project window, select **Analysis** > **RDG Properties**, then select the **NT Variants** tab.

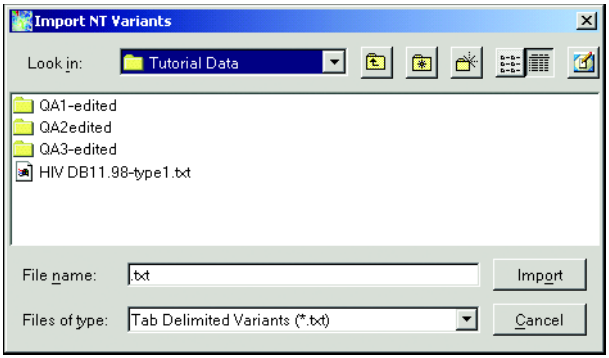
2. Click **Import**.



3. Browse to the appropriate file, then select it.

   Note: The files must be tab-delimited text files as indicated in the Files of types field.

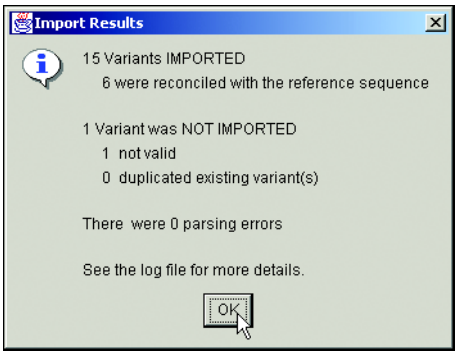
4. Click **Import**.



5. Select the reference segment, then click **OK**.

6. After the data are imported, the Imports Results dialog box opens with information regarding the import.

7. Note the information, then click **OK**.



The variants now appear in the NT Variants tab of the RDG Properties dialog box.

Deleting a Layer in the RDG

SeqScape® Software 3 allows you to selectively delete layers from the Reference Data Group (RDG).

To delete a layer:

1. Select **Analysis** > **RDG Properties**.
2. Select the **ROI** tab.
3. Click to select the layer you want to delete.
4. Click **Delete Layer**.

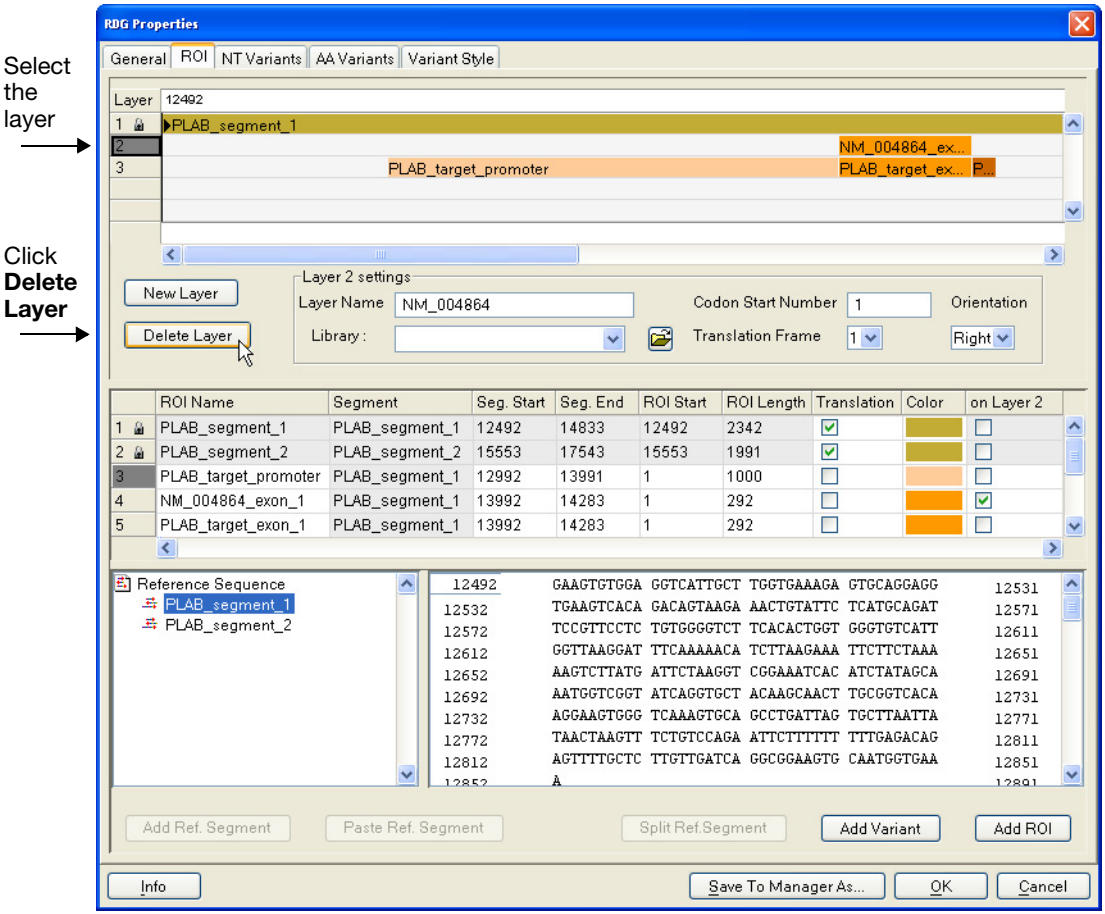
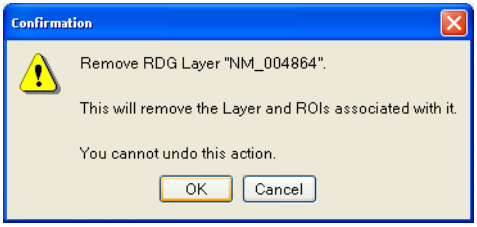


Figure 6-3 ROI tab of the RDG Properties dialog box

The Remove RDG Layer Confirmation box appears.



5. Click **OK** to permanently remove this layer and all associated Regions of Interest (ROI).

Section 6.4 Importing and Exporting Project Information

About Importing and Exporting

The purpose of importing and exporting project information is to transfer the project information to another computer. You can export or import projects, project templates, reference data groups, nucleotide and amino acid variant tables, libraries, and analysis defaults from the SeqScape® Manager. This allows you to examine and compare results from different data stores.

Note: The export and import functions of SeqScape® Manager use the file extension CTF.

Importing from SeqScape® Manager

To import from SeqScape® Manager:

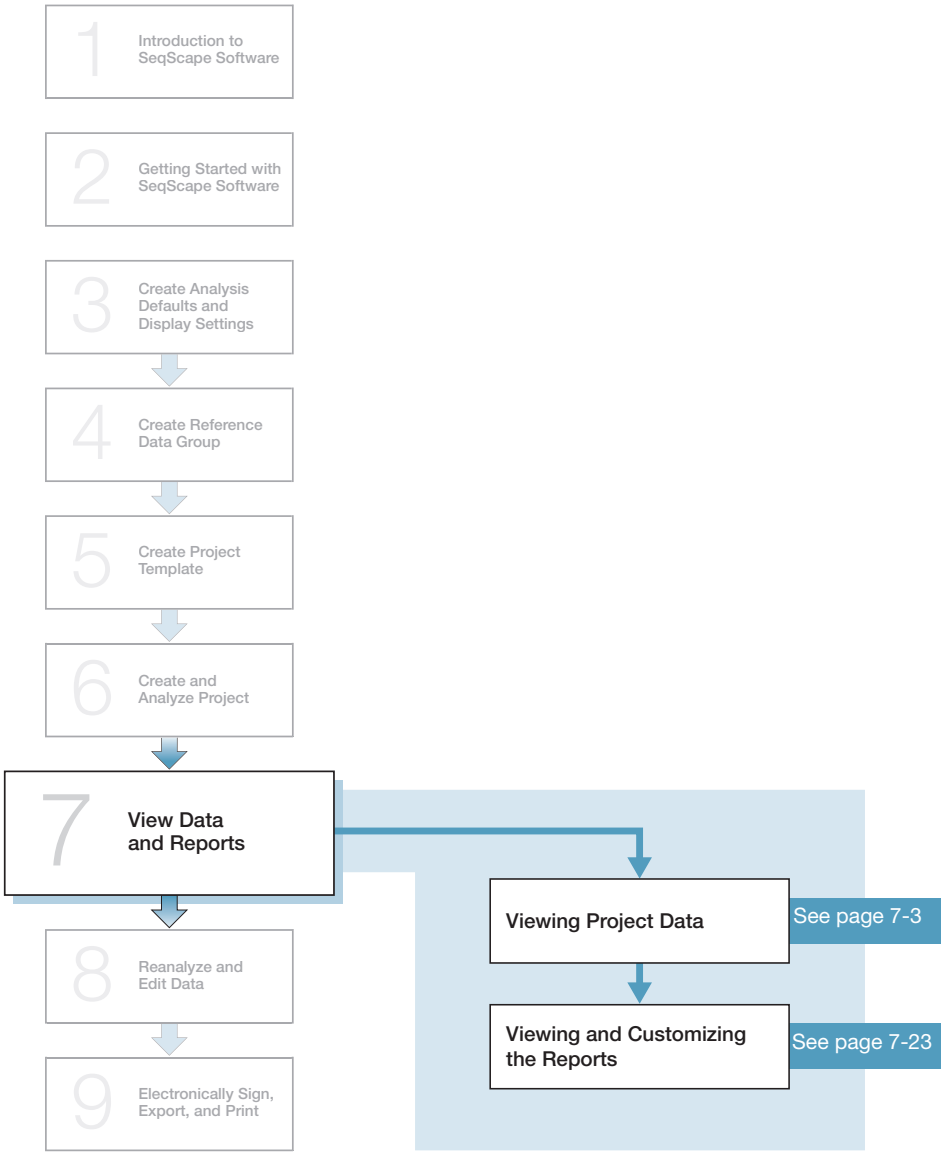
1. Select **Tools** > **SeqScape Manager**.
2. Select any tab into which you want to import.
3. Click **Import**.
4. Navigate to the file that you want to import.
5. Click **Import**.The imported file appears in the list under the appropriate tab.

Exporting from SeqScape® Manager

To export from SeqScape® Manager:

1. Select **Tools** > **SeqScape Manager**.
2. Select any one of the tabs from which you want to export.
3. From the list, select the file that you want to export, then click **Export**.
4. Navigate to the location to which you want to export.
5. Rename the file, if necessary, using the .ctf extension.
6. Click **Export**. The exported file is available to import into another project.

7 Viewing the Results



Section 7.1 Viewing Data

In This Section

View Formats and Displays

View Formats

You can view the results in multiple formats:

Project view – To view all data

Specimen view – To view samples within a specimen

Segment view – To view a summary of all the samples assembled within each reference segment

Sample view – A summary of the data for each sample

Data Display Conventions

The sequence data are displayed using the following conventions:

- Every mixed base (or choice of mixed bases) is represented as a single IUB code. For more information, see Appendix D, "Translation Tables."
- Spaces in aligned sequences are displayed as dashes and are not part of the original sequence.
- In the Dots view and in the collapsed NT view, characters that are identical to the reference are displayed as dots.
- The aligned reference sequence appears at the top of the table and the aligned sequences appear in the rows below in the Project view.

Quality Value Display

The QV (quality value) is displayed as a bar above each called base for the sample sequence and consensus sequence. The height of a bar corresponds to a 1 to 99 value that is determined by the analysis.

Note: For QVs from 50 to 99, all QV bars are identical in height and color.

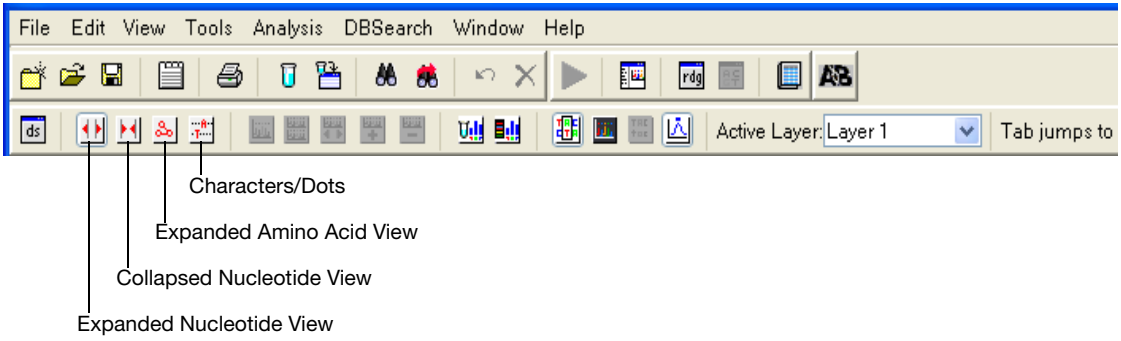
Note: For more information on quality values, see Chapter A, "Sample and Consensus Quality Values."

Exporting and Printing Project Data

To export the project data, see "Exporting" on page 9-11, to print data, see "Printing Data and Reports" on page 9-21.

Project Views

There are four main project views, only one of which can be displayed at a time:



Displaying the Project Views

To display project views:

1. Open the project of interest.
2. Select a layer in the Active Layer drop-down list.
3. At the top of the navigation pane, select the project icon.
4. Use the instructions in Table 7-1 to display the project views of interest.

Table 7-1 Project Views

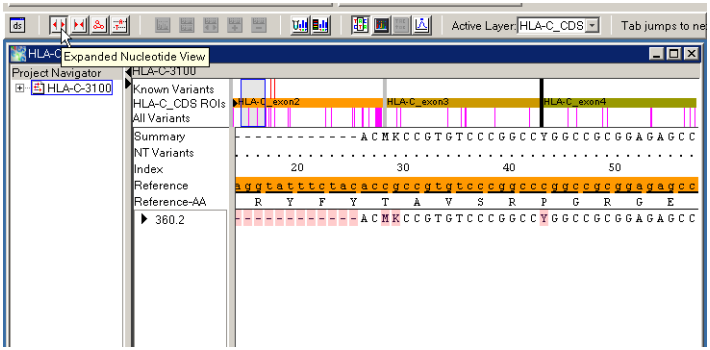
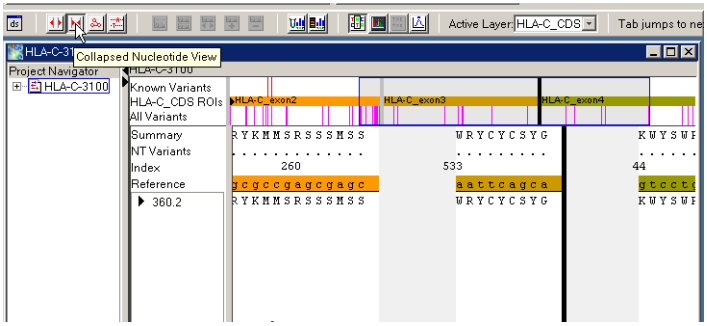
View	Procedure	Display
Expanded Nucleotide	Click  .	
Collapsed Nucleotide	Click  .	 This view shows only those nucleotides that differ from the aligned reference. Bases that match the reference sequence are displayed as dots, regardless of the state of the Dots setting. **Note:** Click  (Expanded Nucleotide View) to return to the expanded view.

Table 7-1 Project Views *(continued)*

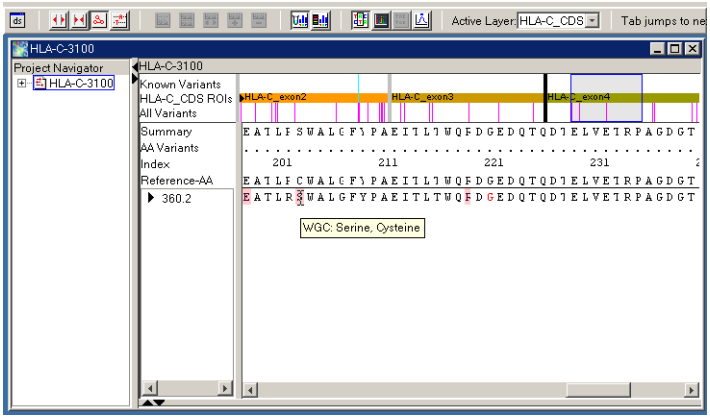
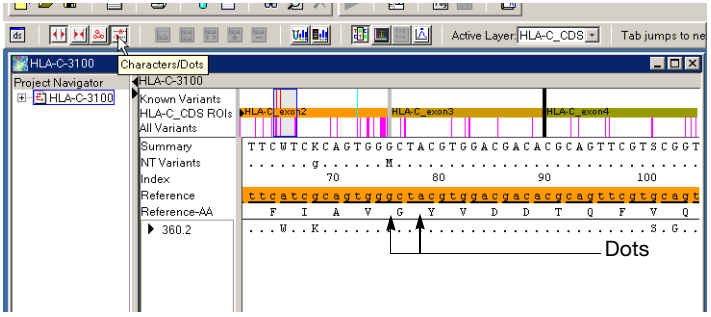
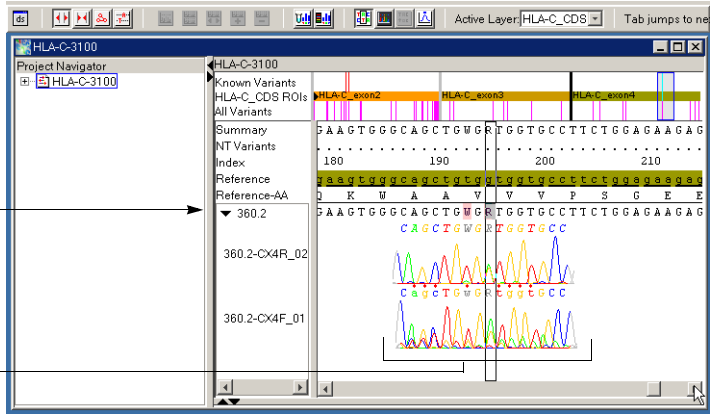
View	Procedure	Display
Expanded Amino Acids (translation of the nucleotide sequence)	Click  .	 **Note:** Holding the pointer over an amino acid displays the possible translations and the codon at that position. **Note:** Bold red characters (default) indicate the location of a degenerate codon.
Characters/ Dots	Click  .	 **Note:** The characters that are identical to the reference are displayed as dots.

Table 7-1 Project Views *(continued)*

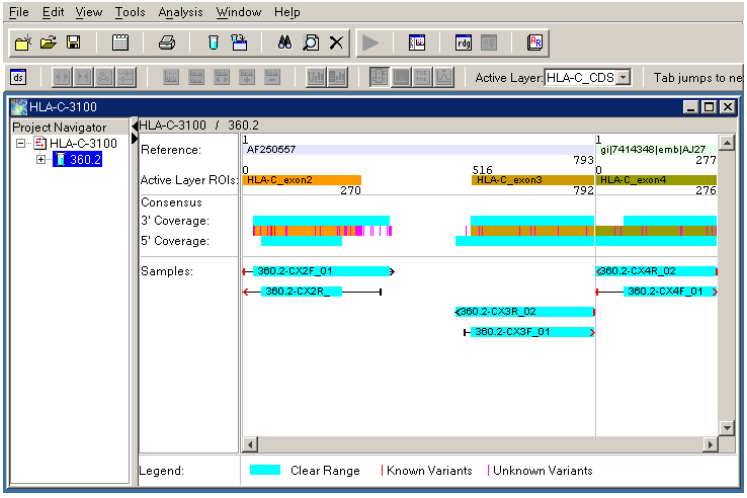
View	Procedure	Display
Electro-pherogram Snippet	1. In the Expanded Nucleotide or Characters/Dots view, select a base in the summary or specimen sequence. 2. Click the triangle next to the specimen name.  Triangle Electropherogram snippets **Note:** Pressing Ctrl+Z centers any electropherogram snippets in the middle of the view.	

Specimen Views

The specimen result is displayed as a schematic of the location and orientation of all samples within a specimen with respect to the reference, ROIs in the current layers, and consensus sequence.

To display the specimen view:

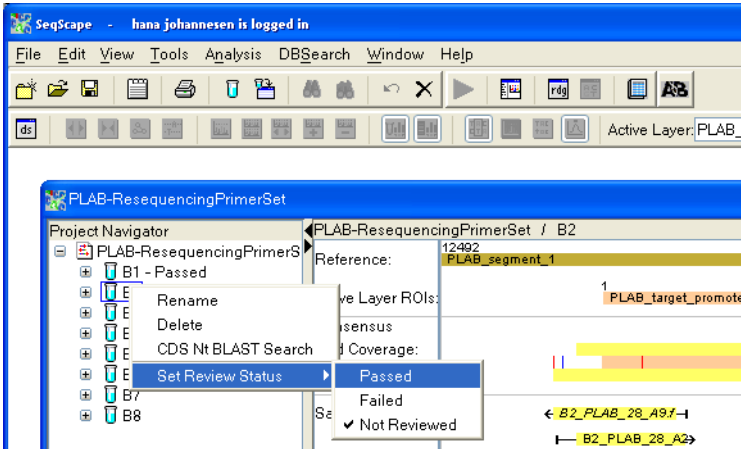
1. Open the project of interest.
2. Select a layer in the Active Layer drop-down list.
3. In the navigation pane, open the project (if necessary), then select a specimen icon.



Labeling the Review Status of a Specimen

In the Specimen view, you can now label the review status of a specimen.

1. Select a project in the Project Navigator.



2. Right-click a specimen, point to **Set Review Status**, then click the appropriate label.

   The review status will display next to the specimen in the Project Navigator.

Segment Views

There are two segment views:

- **Layout view** – Displays a schematic of the location and orientation of the samples with respect to the reference segment.
- **Assembly view** – Displays the nucleotide sequence of the consensus and samples and sample electropherogram data for the reference segment retained from the project view.

Note: The view position in the Assembly view (blue box) is represented by red lines in the Layout view. Click the Layout view to navigate to a desired position in the Assembly view.

Table 7-2 describes the multiple Assembly view types.

Displaying the Segment Views

To display segment views:

1. Open the project of interest.
2. Select a layer in the Active Layer drop-down list.

3. In the navigation pane, open a specimen, then select a segment.

4. Use the procedures in Table 7-2 to display the segment views of interest.

Table 7-2 Segment Views

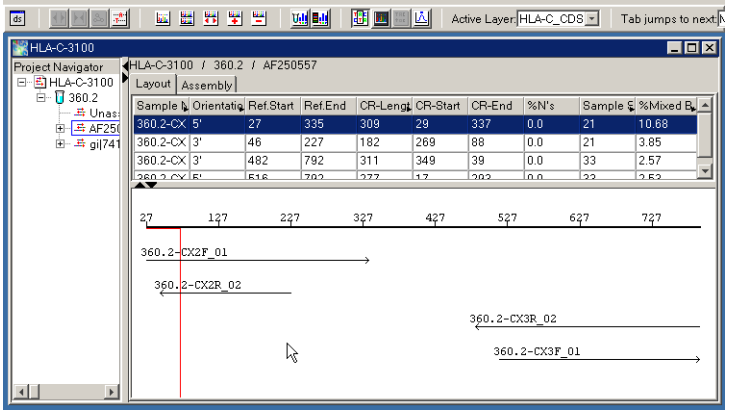
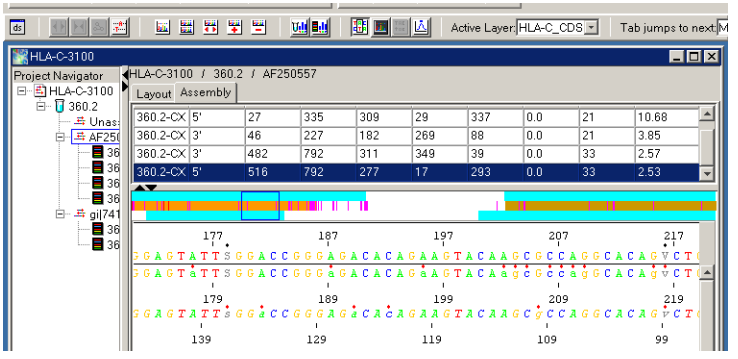
View	Procedure	Display
Layout	Select the **Layout** tab.	
Assembly	Select the **Assembly** tab.	

Table 7-2 Segment Views *(continued)*

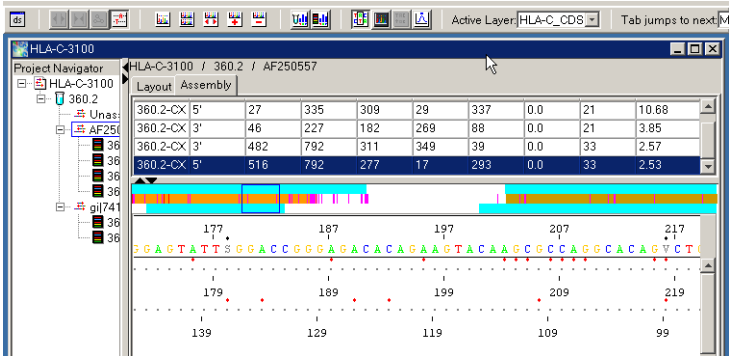
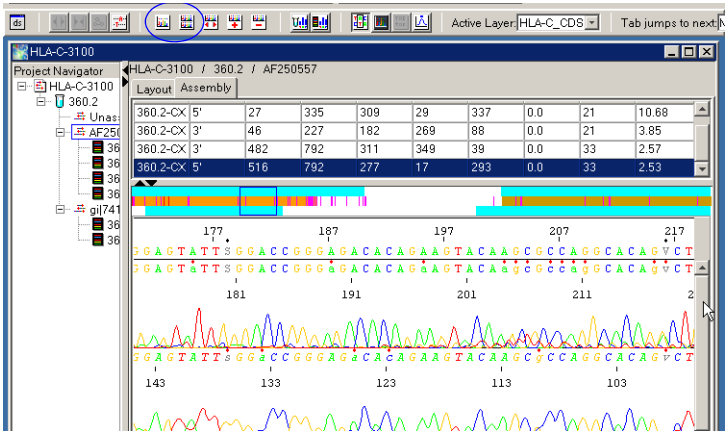
View	Procedure	Display
Dot Assembly	1. Select the **Assembly** tab. 2. Click 	
Electro-pherogram Assembly	1. Select the **Assembly** tab. 2. Select a sample in the sample table. 3. Click  for multiple EPS, or click  for one EP.	

Table 7-2 Segment Views *(continued)*

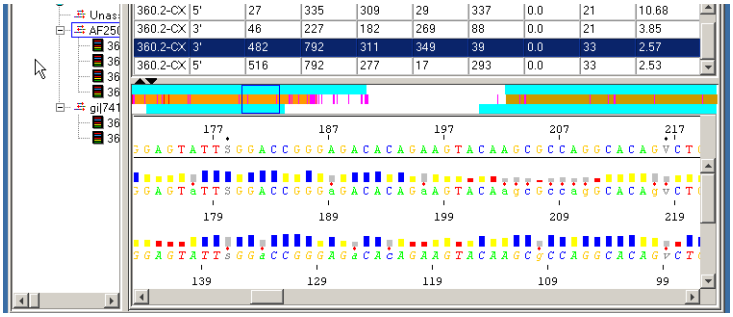
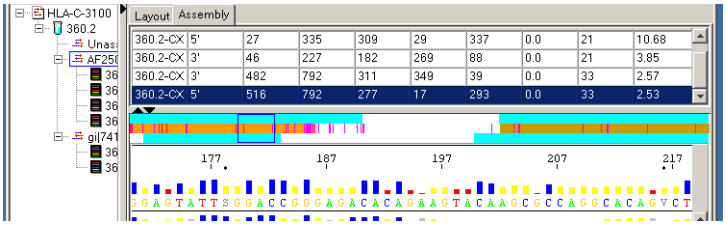
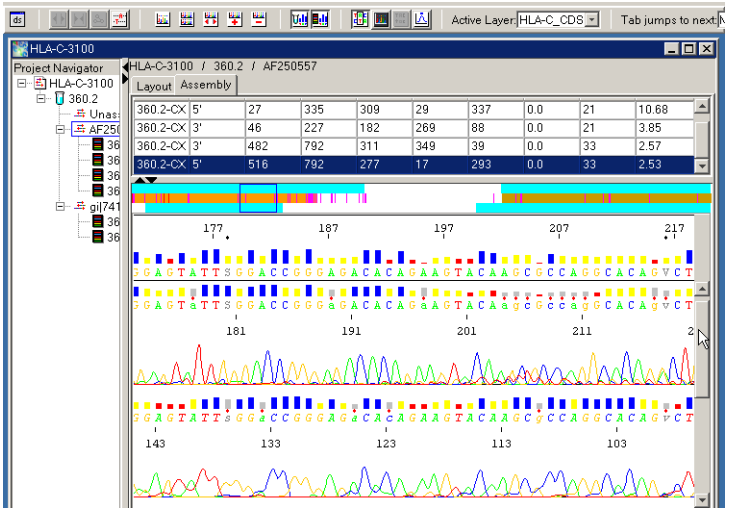
View	Procedure	Display
QV Assembly	1. Select the **Assembly** tab. 2. Click  for sample QVs, and/or click  for consensus QVs.	 

Table 7-2 Segment Views *(continued)*

View	Procedure	Display
Assembly (The EP and QV views can be used with the Sequence or Dot view.)	1. Select the **Assembly** tab. 2. Click  . 3. Click  . 4. Click  .	 **Note:** A black dot above a base in the consensus sequence indicates a discrepancy between the consensus and the reference sequences, and a red dot above a sample base indicates the base was edited by the consensus caller. Options: • Use  (zoom tools) or Ctrl+/Ctrl- to zoom in and out on the horizontal view. • Use Find and Find Again to search for text.

Sample Views

The sample result includes all the data characteristics of a sample. Sample data characteristics are displayed in the following tabs:

Tab	Displayed Information
Annotation	Information about the data and its analysis.
Sequence	Sequence of the sample in NT codes. For readability, the display clusters the sequences into substrings of 10 characters each, separated by blanks.
Features	Calculated clear range and multiple base positions.
Electropherogram	Electropherogram and basecall data for the sample. The data excluded from the clear range are shown in gray.
Raw	Raw data collected by the genetic analyzer.

Displaying the Sample Views

To display sample views:

1. Open the project of interest.
2. In the navigation pane, open a specimen, then open a segment.
3. Select a sample, then select a tab (see Table 7-3 on page 7-15).
4. Select a new tab to change the view.
5. To view a different sample, select a new sample from an open segment, then select a tab.

Examples of the five tab views are displayed in Table 7-3, “Sample Views,” on page 7-15.

Table 7-3 Sample Views

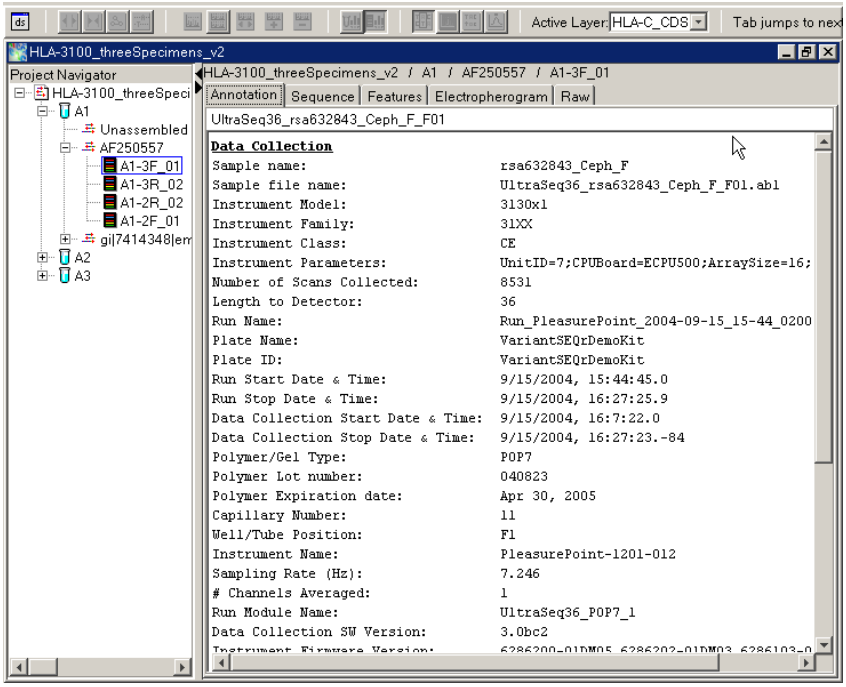
Tab	Display
Annotation	

Table 7-3 Sample Views *(continued)*

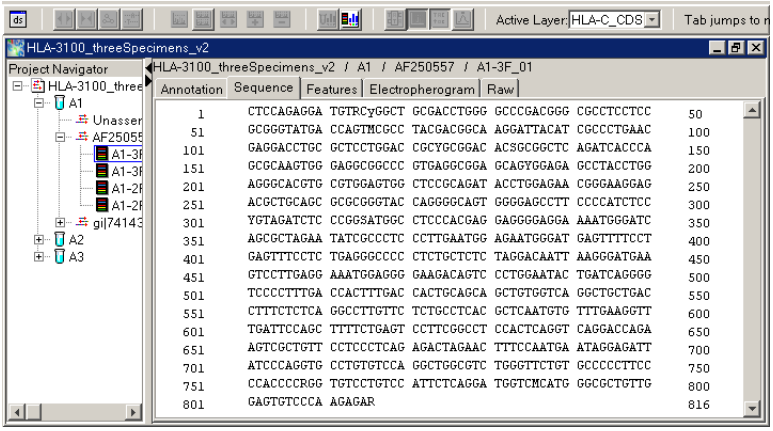
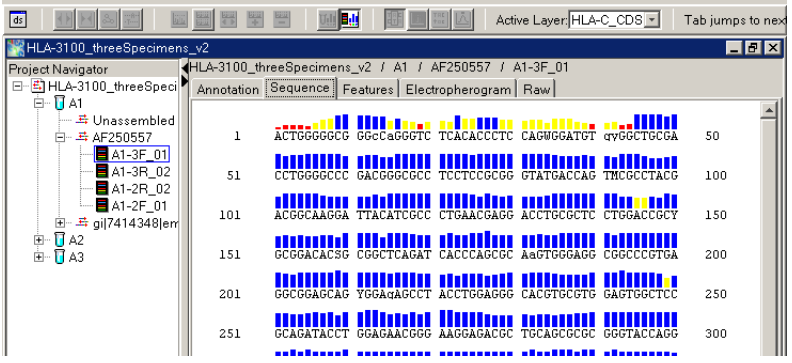
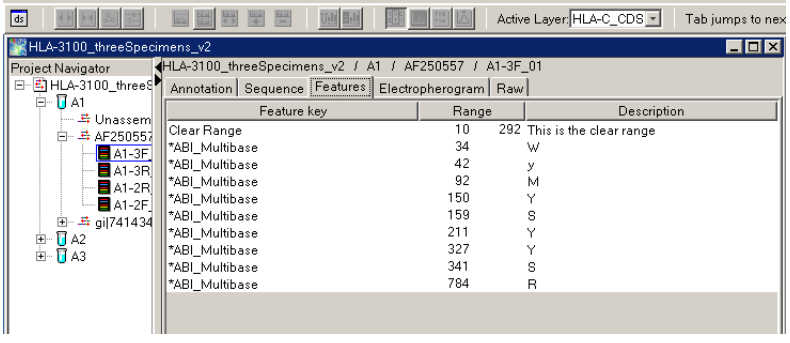
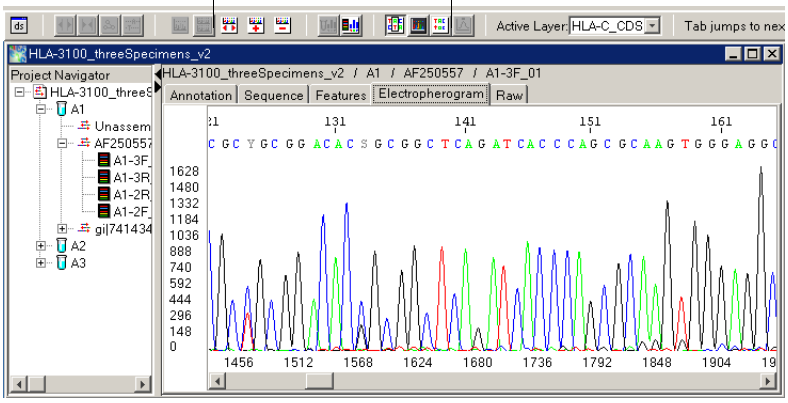
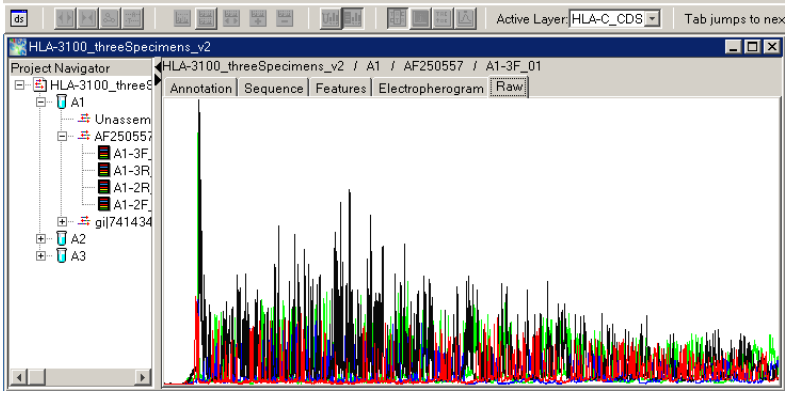
Tab	Display
Sequence	 Use  (Show/Hide Sample QV) to display/hide sample QVs. 
Features	

Table 7-3 Sample Views *(continued)*

Tab	Display
Electropherogram	Active tools  Options: • Use (zoom tools) to zoom in/out on the horizontal view. • Use (Show/Hide Sample QV) to display/hide sample QVs. • Use (Inverse View) to switch between white and black background displays. • Use (View Original Sequence) to display/hide the original sequence called by the software. • Use Find and Find Again to search for text.
Raw	

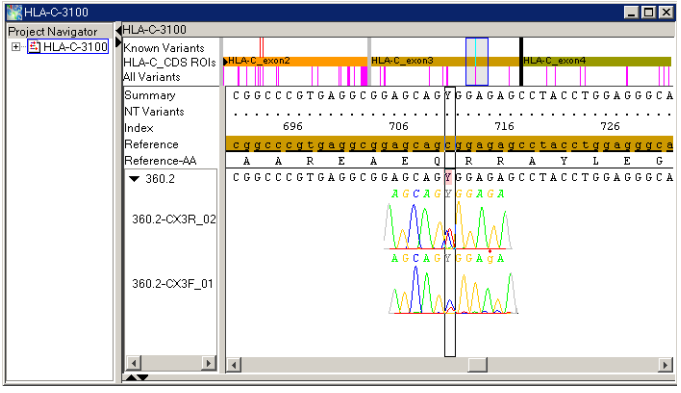
Viewing Variant Data

Note: To edit variant data, see "Editing Variants" on page 8-19.

Method 1 To view variant data:

1. Open the project of interest.
2. Click a consensus base.
3. View the electropherogram snippets by clicking the triangle next to the specimen name.
4. In the Tab Jump to Next drop-down list, select **Multiple**, then select **Known Variant** and **Unknown Variant**.
5. Press **Tab** to move to the next variant or press **Shift-Tab** to move to the previous variant.

   Note: Pressing Ctrl+Z moves any electropherogram snippets of the selected variants to the middle of the view.



Method 2 To view variant data:

1. Open the project of interest.
2. Select **Analysis > Report Manager** or click (Report Manager).

3. In the navigation pane, select the report you want to view.
4. Select **Window > Tile**.
5. Review the positions by selecting a base change in the Mutation table. This action brings the alignment view to the correct position in the alignment.

VariantSEQr® System Data

Amplicon View for VariantSEQr® Resequencing System

SeqScape® Software 3 has been integrated with VariantSEQr® Resequencing System to provide an easy and accurate data analysis solution.

This system includes:

- Validated PCR primer sets (resequencing amplicons)
- Universal protocols for PCR and PCR clean-up, sequencing and sequencing clean-up, and data analysis with SeqScape® software
- A pre-configured project template for analyzing data with SeqScape® software, including:
  - A reference sequence with transcript information
  - Recommended analysis and display settings

Using the System for Data Analysis

To use the system, you import the project template into SeqScape® software, add sample data, then run data analysis.

Reviewing the VariantSEQr® Data

After the data analysis is complete, you can review the data at various levels. Reviewing data at the project level allows you to review and compare all samples in the project.

When you select the project name in the project navigator window, two tabs appear on the right:

- **Project View**
- **Amplicon View**– Enables effortless review of data with respect to the VariantSEQr® system resequencing amplicons (RSA).

Figure 7-1 shows how you can compare the myScience environment graphical view (available when searching for a gene for VariantSEQr® system) with the Amplicon View in SeqScape® software to review an RSA assembly.

In Figure 7-1, the green bar above two of the amplicons represents 2x coverage. For RSA000575278, there is coverage of only one sample in the reverse orientation for the given data set.

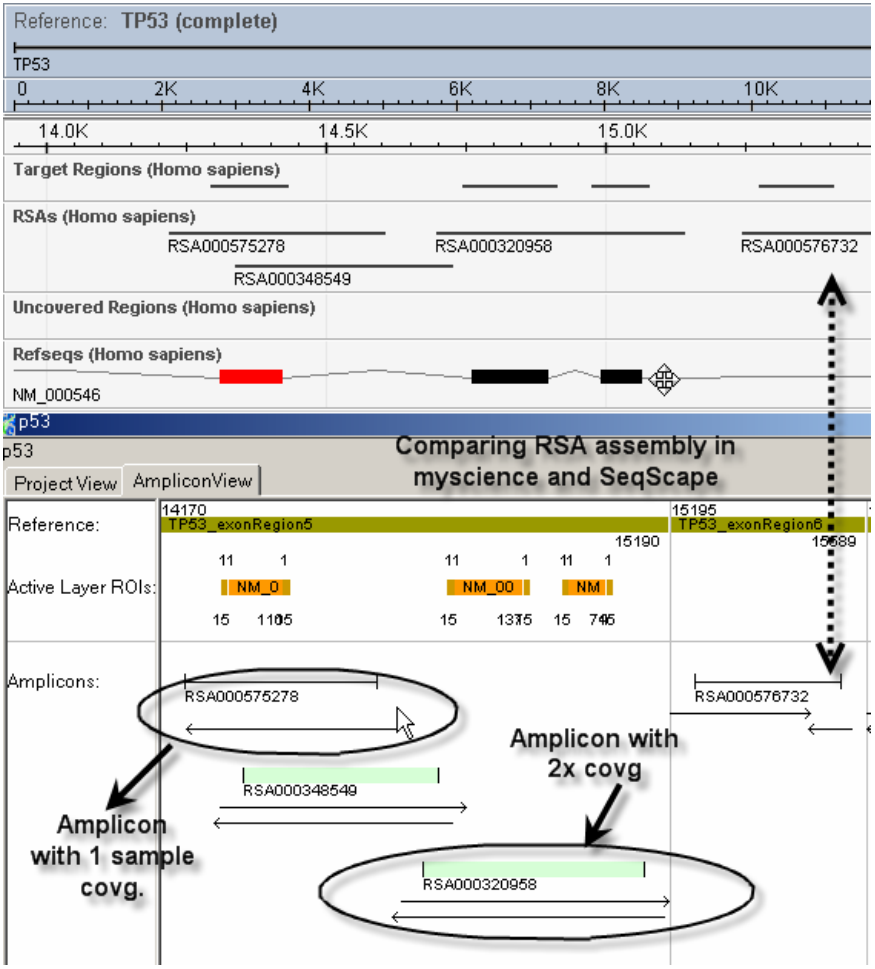


Figure 7-1 Comparing an RSA assembly in myScience environment and SeqScape® software

Troubleshooting the VariantSEQr® Data

To troubleshoot the data, evaluate the sample quality in the unassembled mode in the project navigator window.

The Amplicon view (Figure 7-2) provides a complete status view of the resequencing project, indicating the amplicons that passed and failed. A legend at the left corner allows you to look at variants, both known and unknown, in reference to specific amplicons. A zoom-in feature is available for focusing on specific exon regions.

Another VariantSEQr® system project example is shown in Figure 7-2. The lower pane of the Analysis QC Report displays additional information for RSA Coverage. As shown in Figure 7-2, a table in the Analysis QC Report provides a detailed percentage coverage of the Resequencing Amplicons. You can export this information by exporting the Analysis QC Report.

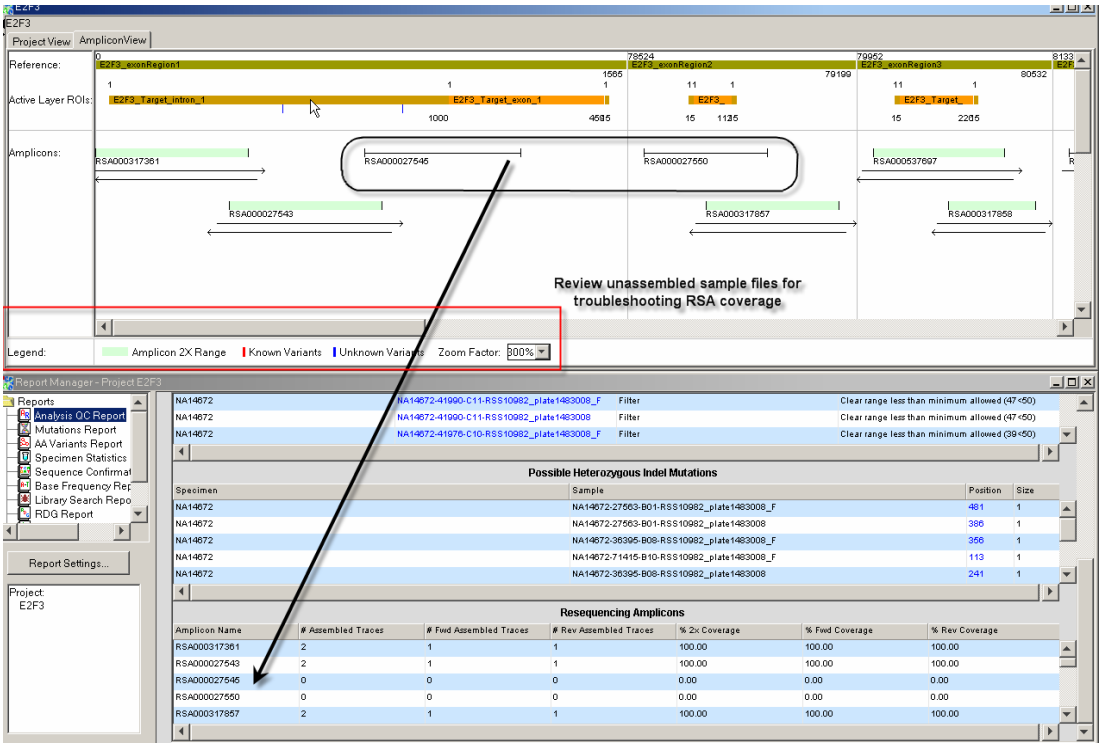


Figure 7-2 Resequencing Amplicon Table in the Analysis QC Report

Section 7.2 Viewing Reports

In This Section

About the Reports

After the data are analyzed, you can view, export, and print reports. You can use reports with project results to evaluate your samples, modify the analysis settings, and edit the basecalling. Also, you can use reports to troubleshoot your results because reports contain hyperlinks to the primary sequence data.

Types of Reports

Eleven reports are generated with every project analysis. Each project has its own Report Manager window containing all the following reports:

- Analysis QC
- Mutations
- AA Variants
- Specimen Statistics
- Sequence Confirmation

- Base Frequency
- Library Search
- RDG
- Audit Trail
- Electronic Signature History
- Genotyping

Note: Only one report can be viewed at a time.

All reports have a Summary table that includes project information and the specimens in the report.

Exporting and Printing Reports

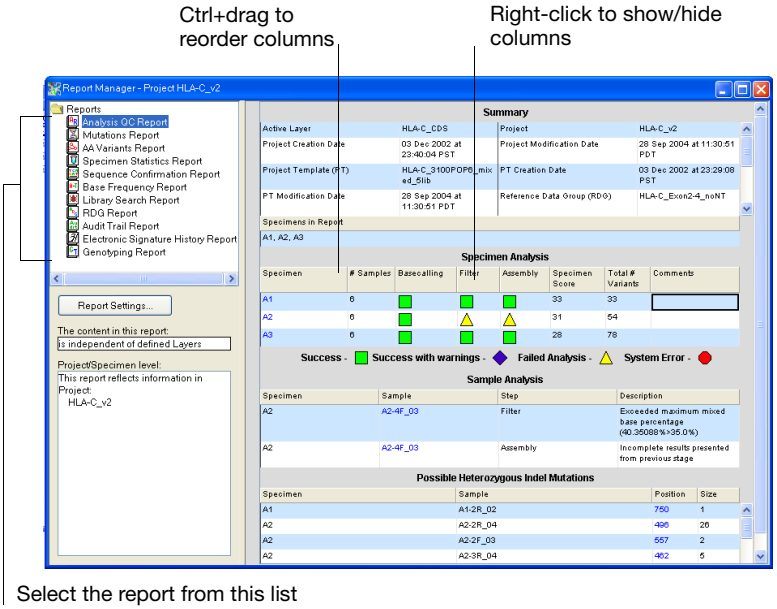
To export a report, see "Exporting Reports" on page 9-18; to print a report, see "Printing Data and Reports" on page 9-21.

Viewing the Reports

The data in the reports are filtered, based on the view (project, specimen, segment, or sample) selected in the navigation pane of the project and the layer selected from the Active Layer drop-down menu.

To view a report:

1. Open the project of interest, then select the active layer.
2. In the navigation pane, select the project, specimen, segment, or sample view.
3. Select **Tools** > **Report Manager** or click  (Report Manager).



4. In the navigation pane, select the report you want to view.

5. To view other reports, select a different report in the navigation pane.

   Note: Only one report can be viewed at a time.

6. To update the reports with additional data:

   a. In the project, select additional or different specimen/samples.

   b. Click (Report Manager) to update the data in the open report.

Viewing the Reports and Project Results Together

When you view reports, you can tile the report with the Project window to easily view the data when you click a hyperlink in the report.

To view the project results and reports together:

1. Open the project of interest.
2. Select a layer in the Active Layer drop-down list.
3. Select **Analysis > Report Manager** or click .
4. In the navigation pane, select the report you want to view.
5. Select **Window > Tile**.
6. Click a hyperlink (blue text) in the report, then view the data in the Project view.

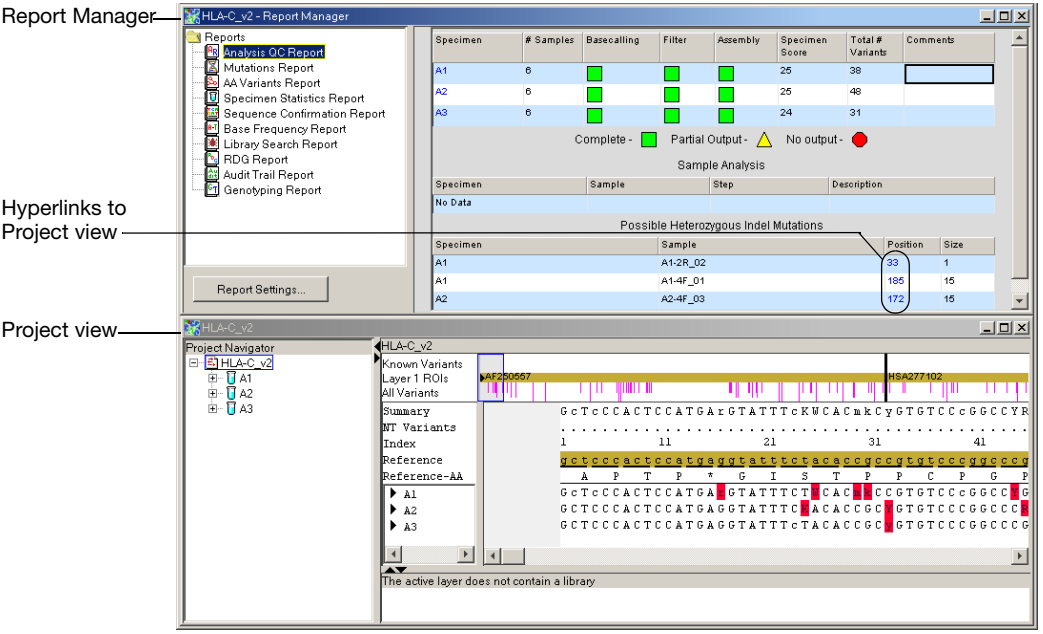


Figure 7-3 Viewing results and reports together

Analysis QC Report

Life Technologies recommends that you click (Analyze) before viewing the Analysis QC report.

Note: All blue text is hyperlinked to the project navigator.

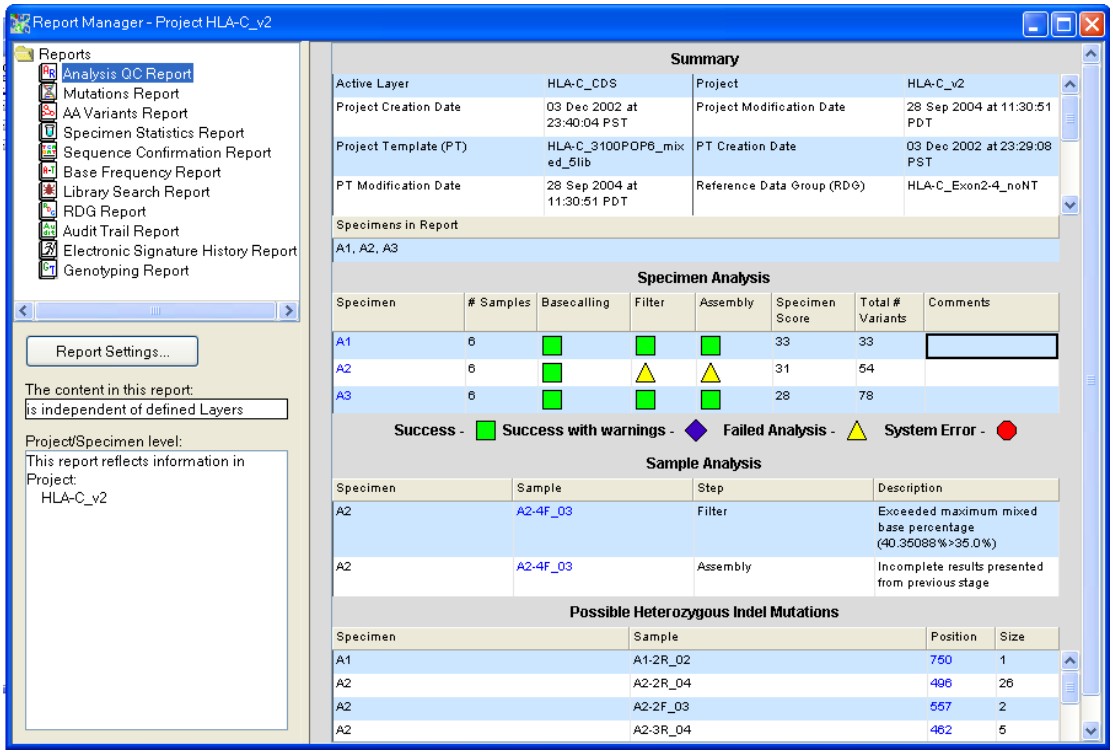


Figure 7-4 Analysis QC report

Table 7-4 Parts of the Analysis QC Report

Table	Description
Summary	Displays project information and the specimens in the report.
Specimen Analysis	Displays specimen analysis results, specimen score (average consensus QV) and total number of variants.

Table 7-4 Parts of the Analysis QC Report

Table	Description
Sample Analysis	Displays sample analysis errors and details.
Possible Heterozygous Indel Mutations	Displays possible mutations and their positions and size for each specimen.

New Blue Diamond Status Indicator

A fourth indicator of basecalling status was added to the Analysis QC Report in SeqScape® software v2.5. This indicator, a blue diamond, designates successful basecalling with some anomalies. In earlier versions of the software, there were only three indicators of basecalling status.

Also, the BC check box in the Sample Manager window now displays blue, when appropriate, in addition to green, yellow, and red.

Understanding the Basecalling Status Indicators

Depending on the basecaller you are using and the quality of your sample data, either two or four status indicators may be shown.

- KB™ Basecaller: all four indicators
- ABI Basecaller: only green and red indicators

Typical samples that are run according to recommended protocols should produce the green status from the KB™ Basecaller. However, be aware that a sequence called by the ABI Basecaller with the green status is not necessarily superior to the same sample called by the KB™ Basecaller with blue or yellow status. The basecalling status from the KB™ Basecaller is more precise because the KB™ Basecaller has more signal processing quality control than the ABI Basecaller. Thus, when the ABI Basecaller returns the green status, it includes the green, blue, and yellow statuses for the KB™ Basecaller.

In addition, the KB™ Basecaller produces per-base quality values, but the ABI Basecaller does not. (The ABI Basecaller works with ™ software to generate quality values.)

Therefore, if you are using the ABI Basecaller, carefully review all sequences against the processed electropherograms.

Basecalling Status Indicators

Table 7-5 Basecalling Status Indicators

Symbol	Explanation of Basecalling Status
	The basecaller analyzed the sample file correctly.
	The basecaller analyzed the sample file successfully; however, some anomalies that may or may not be serious were detected. Review the error message, the sample score, and the data. The KB™ Basecaller returns the blue status when the sample data are sub-optimal or contain anomalous characteristics, but otherwise may still be called. Consult the analysis report for detailed diagnostic messages, and examine the processed electropherogram and basecalls to determine the severity of the problem. Note that samples in this category may still contain long regions of high-quality calls, as indicated by the per-base quality values.
	The basecaller cannot analyze the sample file due to its low data quality. Troubleshoot at the sample preparation and/or the electrophoresis steps. The KB™ Basecaller returns the yellow status when the sequencing experiment fails or is severely compromised in some way that prevents a successful analysis. In this case, the software saves the results to the sample file, but the called sequence is replaced with a placeholder sequence of five Ns instead of basecalled data. A common reason for this status is a failed sample, where few or no DNA peaks are evident in the raw data. Note that if such a failed sample is basecalled using the ABI Basecaller, it is likely the green status will be returned, but the data are usually unreliable.
	This is a software failure. Check the software error messages. For both basecallers, the red status typically indicates a software configuration error or invalid input to the basecalling algorithm. No analysis results are written to the sample file and the file remains in the unanalyzed state. Use the analysis report to diagnose the problem, or contact technical support.

Mutations Report

Note: All blue text is hyperlinked to the project navigator.

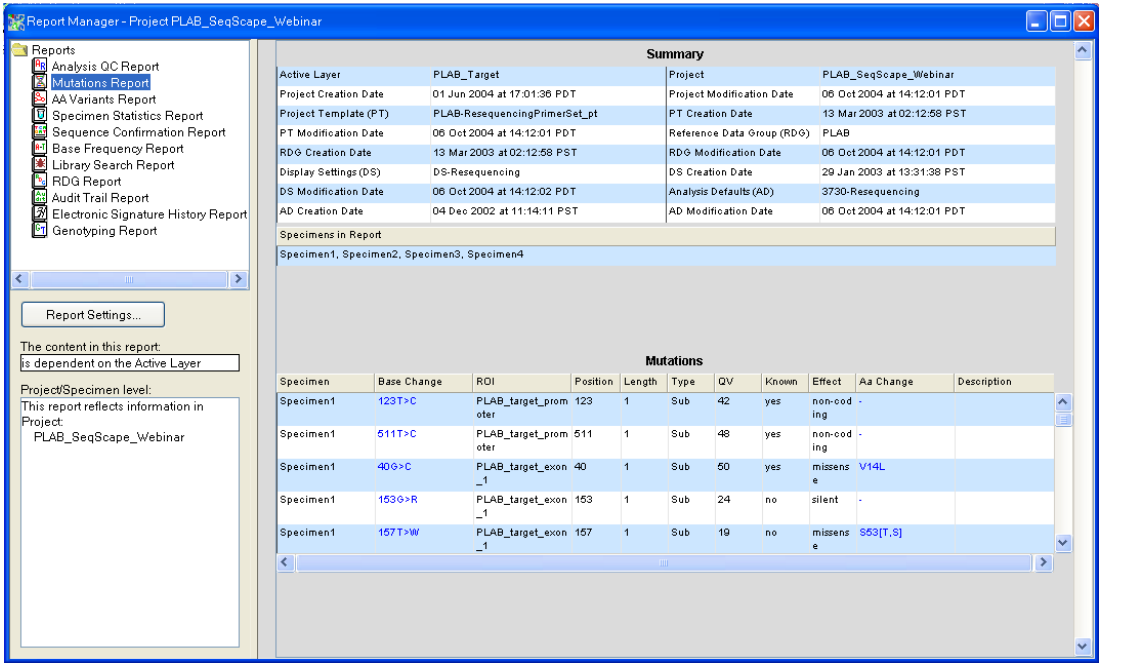


Figure 7-5 Mutations Report

Table 7-6 Parts of the Mutations Report

Table	Description
Summary	Displays project information and the specimens in the report.
Mutation	Displays the bases changed, ROI, position, length, type, QV, and effect information for each mutation detected in a specimen.

The Mutations report includes a column that provides a predicted "effect" for each nucleotide variant. Table 7-7 describes the possible values in the Effect column.

In the Mutation report, clicking an NT variant links to the corresponding base in the Project View. The corresponding AA change in the Mutation Report links to the AA Variant in the AA Variants Report. This AA variant is, in turn, linked to the AA in the Project AA view.

Table 7-7 Predicted effects of nucleotide variants

Effect	Description
Missense	The substitution variant codes for an amino acid substitution.
Nonsense	The substitution variant codes for a terminator codon. (In a mixed codon, if any codon is a terminator codon “nonsense” is displayed).
Silent	The substitution variant is in a coding region but does not code for an amino acid change.
Frameshift insertion	The insertion variant is in a coding region and codes for a frameshift in translation (the size of the insertion is not a multiple of three).
Frameshift deletion	The deletion variant is in a coding region and codes for a frameshift in translation (the size of the deletion is not a multiple of three).
In-frame insertion	The insertion variant is in a coding region and does not code for a frameshift in translation (the size of the insertion is a multiple of three).
In-frame deletion	The deletion variant is in a coding region and does not code for a frameshift in translation (the size of the deletion is a multiple of three).
Non-coding	The variant is not in a coding region.
Partial codon	The variant is in a coding region, but occurs at the beginning or end of the sequence, where you do not know the full three-base codon sequence.
No information	The variant is a result of the consensus sequence not completely covering the reference sequence. These are not real variants, so you cannot predict a real effect.
Heterozygous deletion/insertion	Specimen-level heterozygous indel mutation (HIM) identification

AA Variants Report

Note: All blue text is hyperlinked to the project navigator.

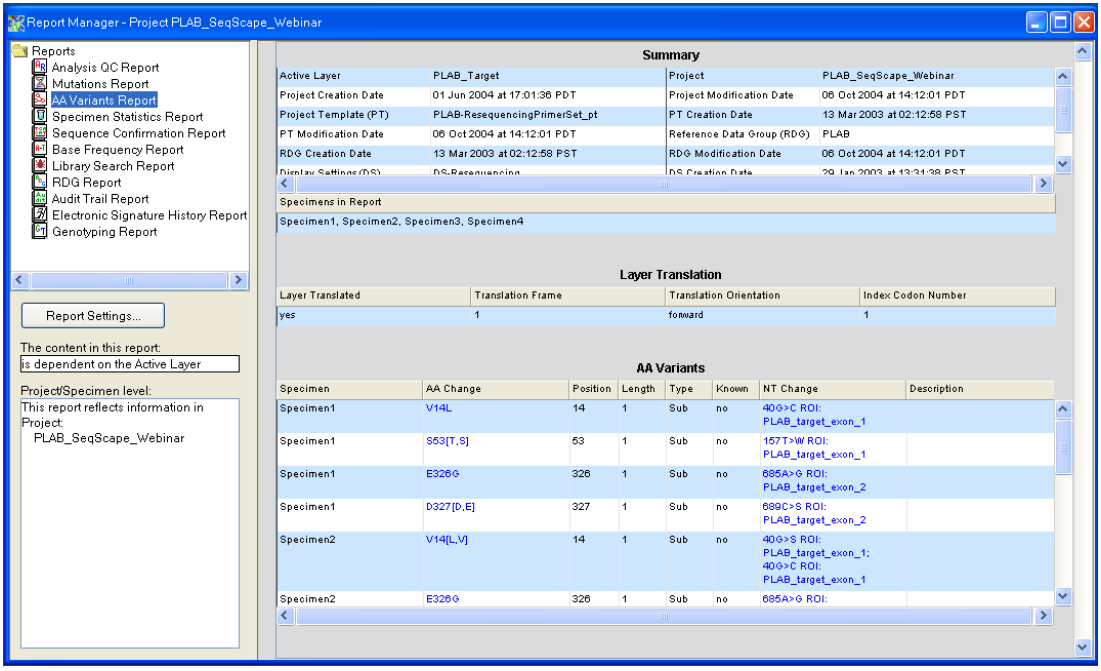


Figure 7-6 AA Variant Report

Table 7-8 Parts of the AA Variant Report

Table	Description
Summary	Displays project information and the specimens in the report.
AA Variant	Displays the AA changed, position, length, type, and description for each variant detected in a specimen.

Specimen Statistics Report

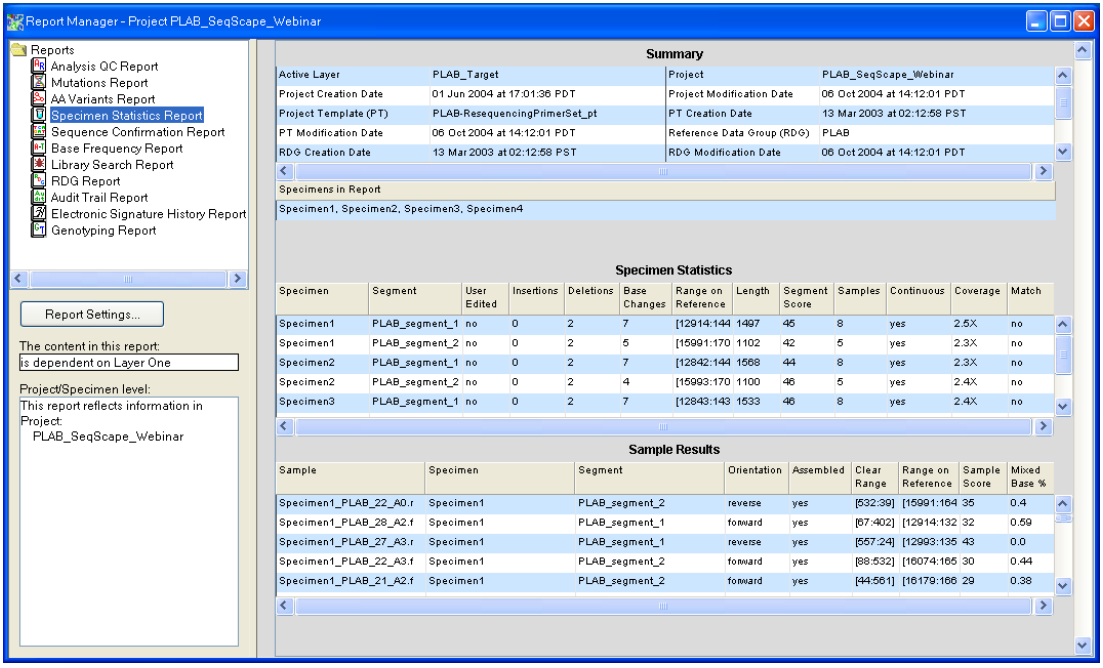


Figure 7-7 Specimen Statistics Report

Table 7-9 Parts of the Specimen Statistics Report

Table	Description
Summary	Displays project information and the specimens in the report.
Specimen Statistics	Displays the bases changed, ROI, position, length, type, QV, and effect information for each mutation detected for each specimen.
Sample Results	Displays the specimen, segment, assembly status, calculated clear range, sample score (average QV), and % mixed bases for each sample.

Sequence Confirmation Report

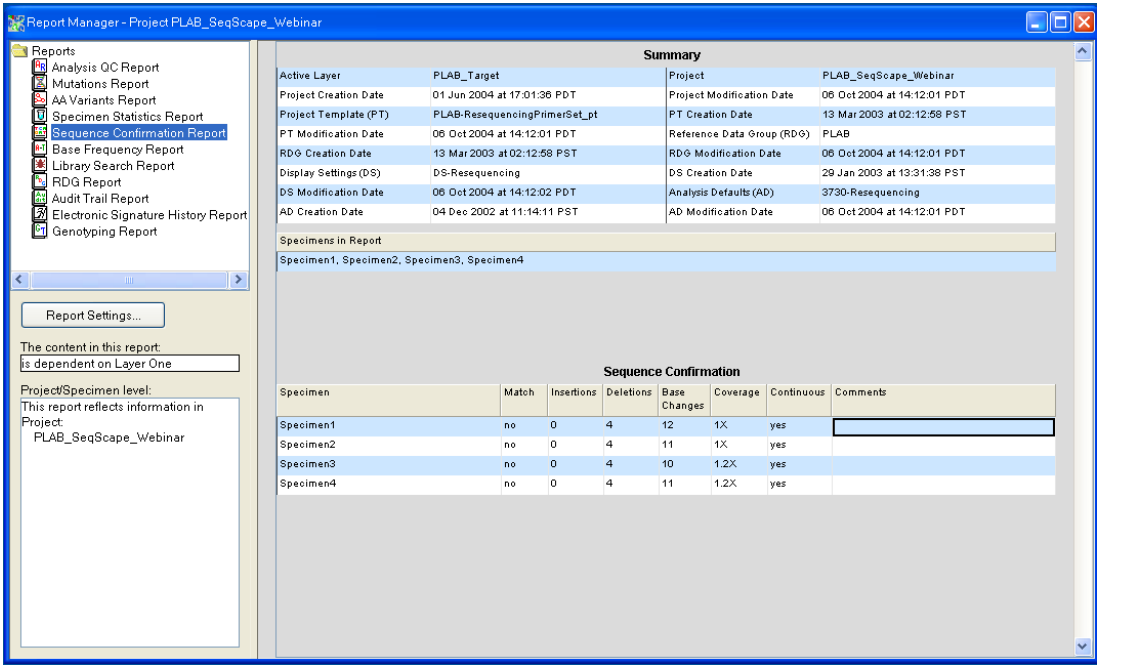


Figure 7-8 Sequence Confirmation Report

Table 7-10 Parts of the Sequence Confirmation Report

Table	Description
Summary	Displays project information and the specimens in the report.
Sequence Confirmation	Displays the match, the number of insertions, deletions and bases, and the amount of coverage and whether it is continuous for each specimen.

Base Frequency Report

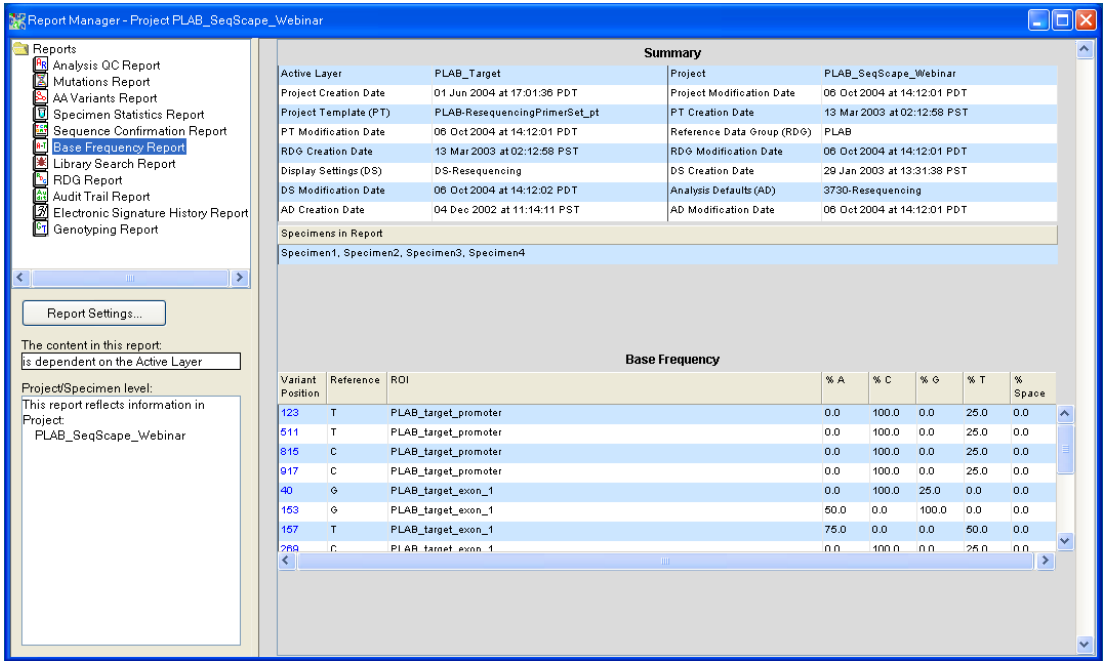


Figure 7-9 Base Frequency Report

Table 7-11 Parts of the Base Frequency Report

Table	Description
Summary	Displays project information and the specimens in the report.
Base Frequency	Displays the reference, ROI, and the % of each base and space for each variant position.

Library Search Report

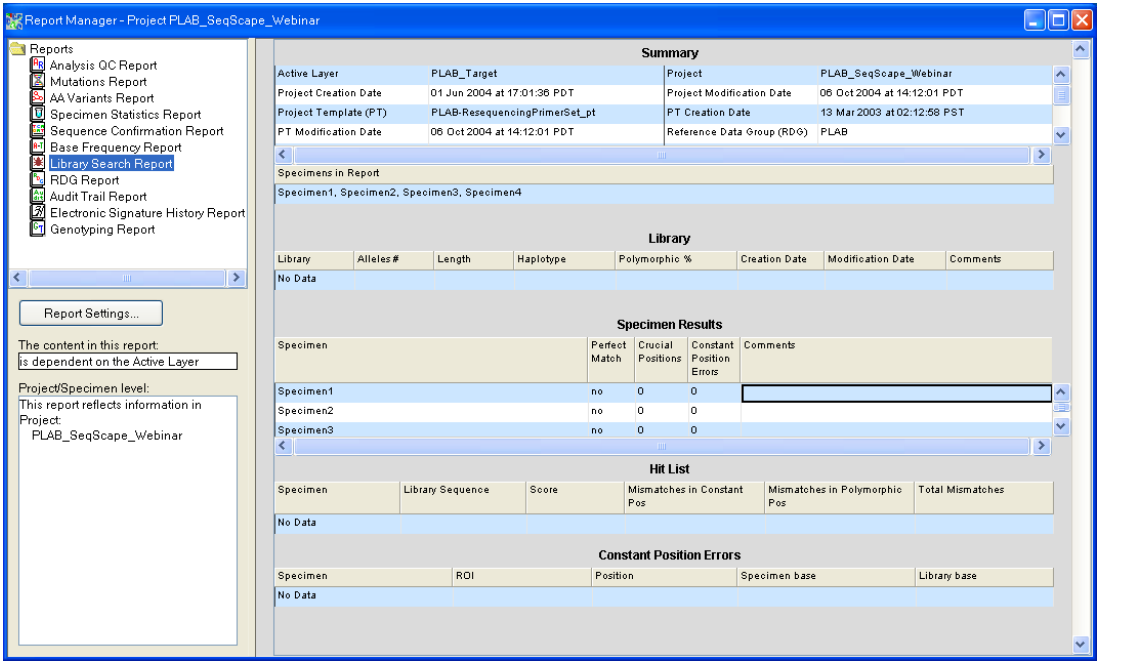


Figure 7-10 Library Search Report

Table 7-12 Parts of the Library Search Report

Table	Description
Summary	Displays project information and the specimens in the report.
Library	Displays the library and information used in the search.
Specimen Results	Displays the match status and the crucial-position and constant-position errors for each specimen.
Hit List	Displays the library matches found, their scores (closest match), and mismatch information for each specimen.
Constant Positions Errors	Displays the position, specimen base, and library base for each specimen and ROI. The values in the Position column are hyperlinked to the project navigator.

RDG Report

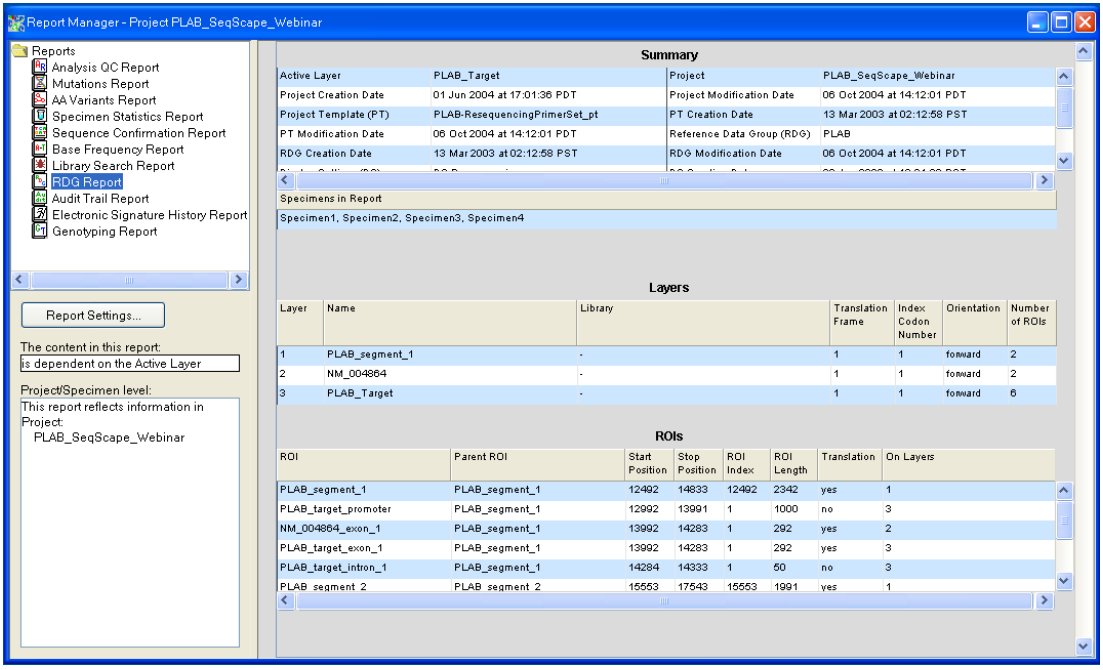


Figure 7-11 RDG Report

Table 7-13 Parts of the RDG Report

Table	Description
Summary	Displays project information and the specimens in the report.
Layers	Displays a summary of the information for each layer in the project as defined in the RDG.
ROIs	Displays a summary of the information for each ROI as defined in the RDG.

Audit Trail Report

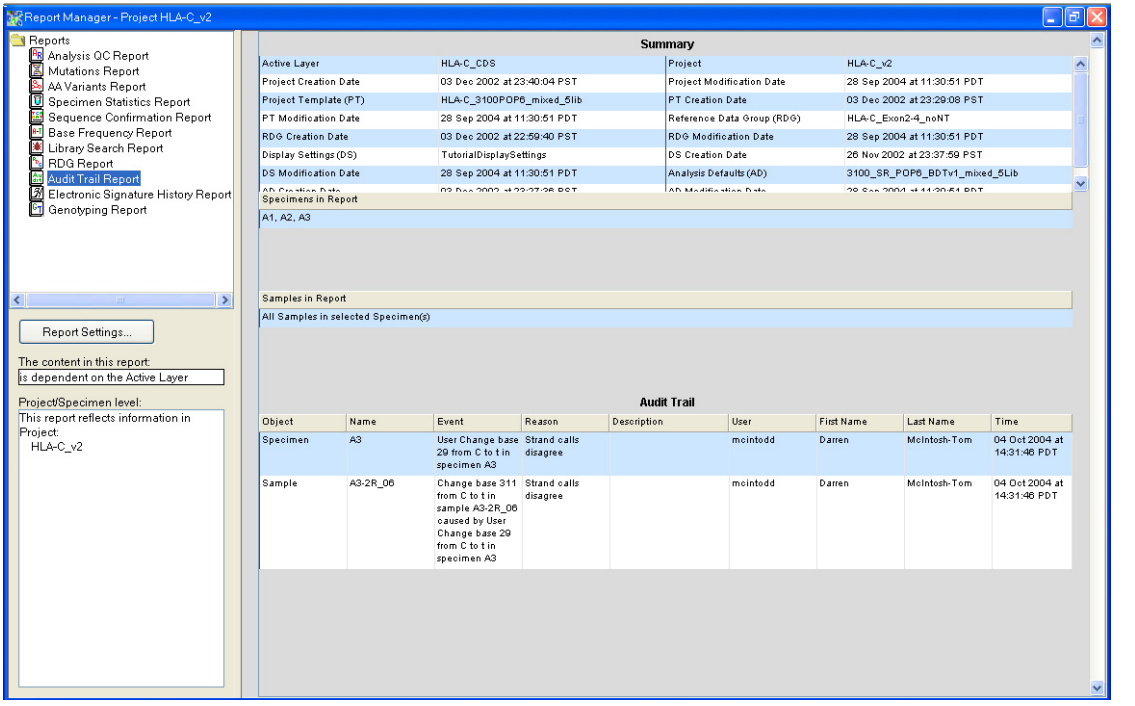


Figure 7-12 Audit Trail Report

Table 7-14 Parts of the Audit Trail Report

Table	Description
Summary	Displays project information and the specimens and samples in the report.
Audit Trail	Displays a record of the edits and changes made to data in a project, if the audit trail feature is on.

Electronic Signature History Report

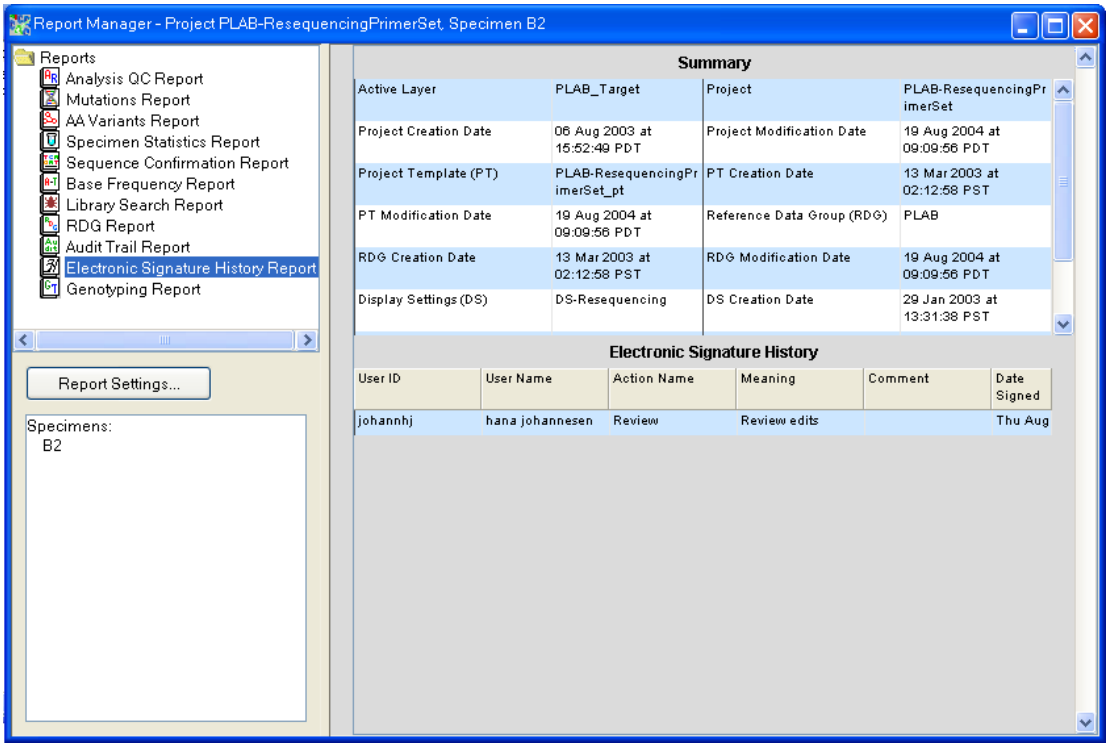


Figure 7-13 Electronic Signature History Report

Table 7-15 Parts of the Electronic Signature History Report

Table	Description
Summary	Displays project information and the specimens and samples in the report.
Electronic Signature History	Displays a record of the electronic signature history made to data in a project, if the electronic signatures feature is on.

Genotyping Report

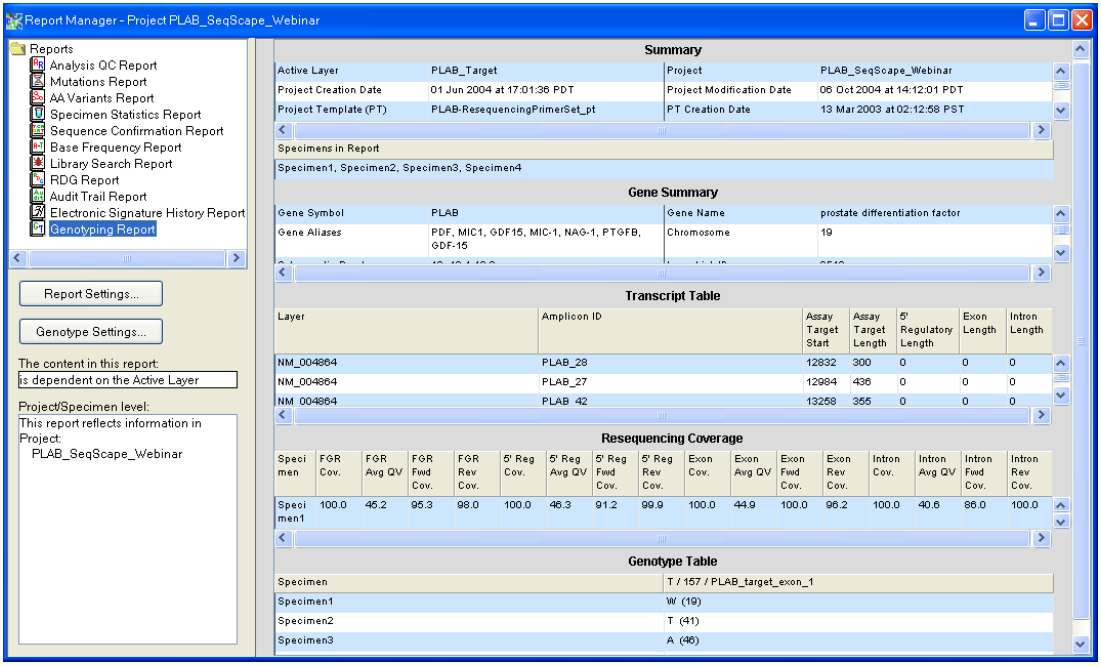


Figure 7-14 Genotyping Report

Table 7-16 Parts of the Genotyping Report

Table	Description
Summary	Displays project information and the specimens and samples in the report
Specimens_in_ Report	Lists the specimens included in the report
Gene Summary	(These fields are not used at this time.)

Table 7-16 Parts of the Genotyping Report *(continued)*

Table	Description
Transcript Table	Displays the layout of the intended amplicons. Lists all the transcripts with their accompanying amplicons in the RDG. • Layer: Name of the transcript being reported on • Amplicon_ID: Identity of an amplicon on this transcript (there are many rows of amplicons for each transcript) • Assay_Target_Start: Start of the assay region of the amplicon in reference segment coordinates • Assay_Target_ Length: Designed length of the assay region • 5′_Regulatory_Length: Amount of amplicon sequence covering the 5′ regulatory region • Exon_Length: Amount of amplicon sequence covering the exon region in the transcript • Intron_Length: Amount of amplicon sequence covering the intron region in the transcript

Table 7-16 Parts of the Genotyping Report *(continued)*

Table	Description
Resequencing Coverage	Displays specimen coverage statistics for the active layer. • Specimen: Name of specimen being reported on • FGR Coverage: Full gene region coverage. Percent of specimen consensus sequence covering the entire targeted region • FGR Avg QV: Average quality value of the consensus sequence across the target region • FGR Fwd Coverage: Percent of target region coverage by the specimen consensus sequence generated from forward reads • FGR Rev Coverage: Percent of target region coverage by the specimen consensus sequence generated from reverse reads • 5′ Reg Coverage: 5′ regulatory region coverage. Percent of the specimen consensus sequence covering the 5′ regulatory region • 5′ Reg Coverage Avg QV: Average quality value of the consensus sequence at the 5′ region • 5′ Reg Fwd Coverage: Percent of target region coverage by the specimen consensus sequence generated at the 5′ region for forward reads • 5′ Reg Rev Coverage: Percent of target region coverage by the specimen consensus sequence generated at the 5′ region for reverse reads • Exon Coverage: Exon coverage for a specific gene • Exon Avg QV: Average quality value of the consensus sequence at the exon region • Exon Fwd Coverage: Quality value of the consensus sequence at the exon region for forward reads • Exon Rev Coverage: Quality value of the consensus sequence at the exon region for reverse reads • Intron Coverage: Intron coverage for a specific gene • Intron Avg QV: Average quality value of the consensus sequence at the intron region • Intron Fwd Coverage: Quality value of the consensus sequence at the intron region for forward reads • Intron Rev Coverage: Quality value of the consensus sequence at the intron region for reverse reads
Genotyping Table	Displays genotypes of all specimens at selected positions in the reference.

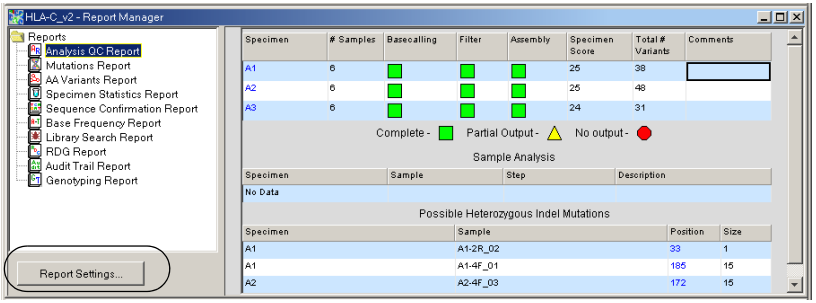
Customizing the Reports

Customizing Text Settings

You can customize the text and select the amino acid variant.

To customize the report settings:

1. Click the **Report Settings** button ion the Report Manager.



The Report Settings dialog box opens displaying the Report Display Settings tab.

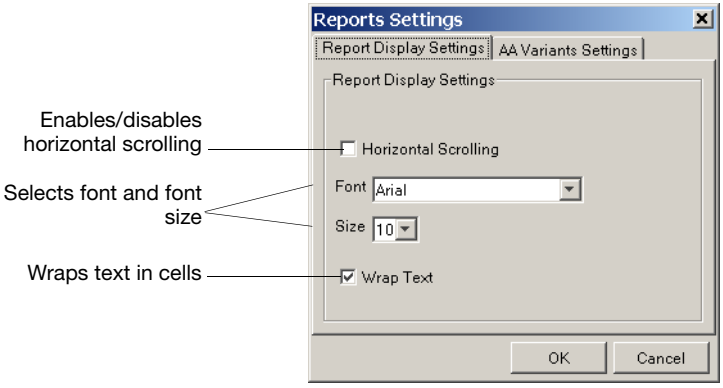


Figure 7-15 Report Display Settings tab of the Report Settings dialog box

2. Select or deselect **Horizontal Scrolling**. The default is off.

3. Select a font type and font size in the appropriate drop-down lists. The default font is Arial size 10.

4. Select or deselect **Wrap Text**. Examples of wrapped and unwrapped text are shown in Figure 7-16.

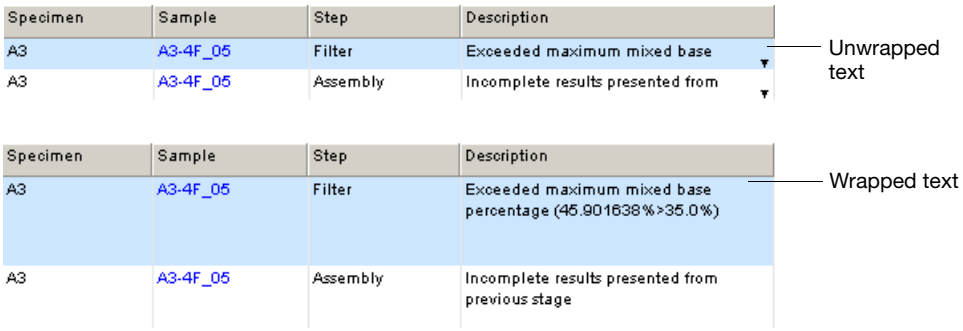
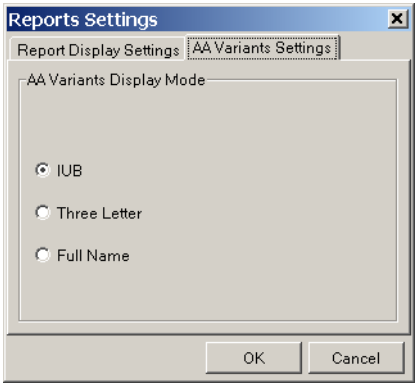


Figure 7-16 Examples of unwrapped and wrapped text

To set the display of AA variants:

1. Select the **AA Variants Settings** tab.

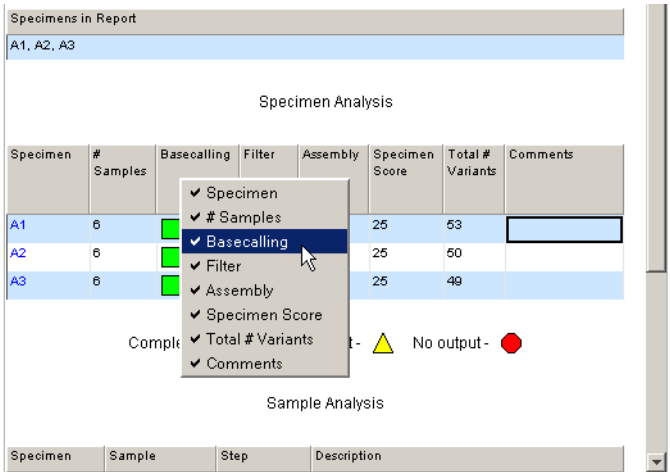


2. Select the display mode, then click **OK**.

Customizing the Data View

To customize the information displayed in a report:

1. Right-click any column heading of a table. A list of the column headings in the table is displayed.



2. To hide a column, deselect the column heading.

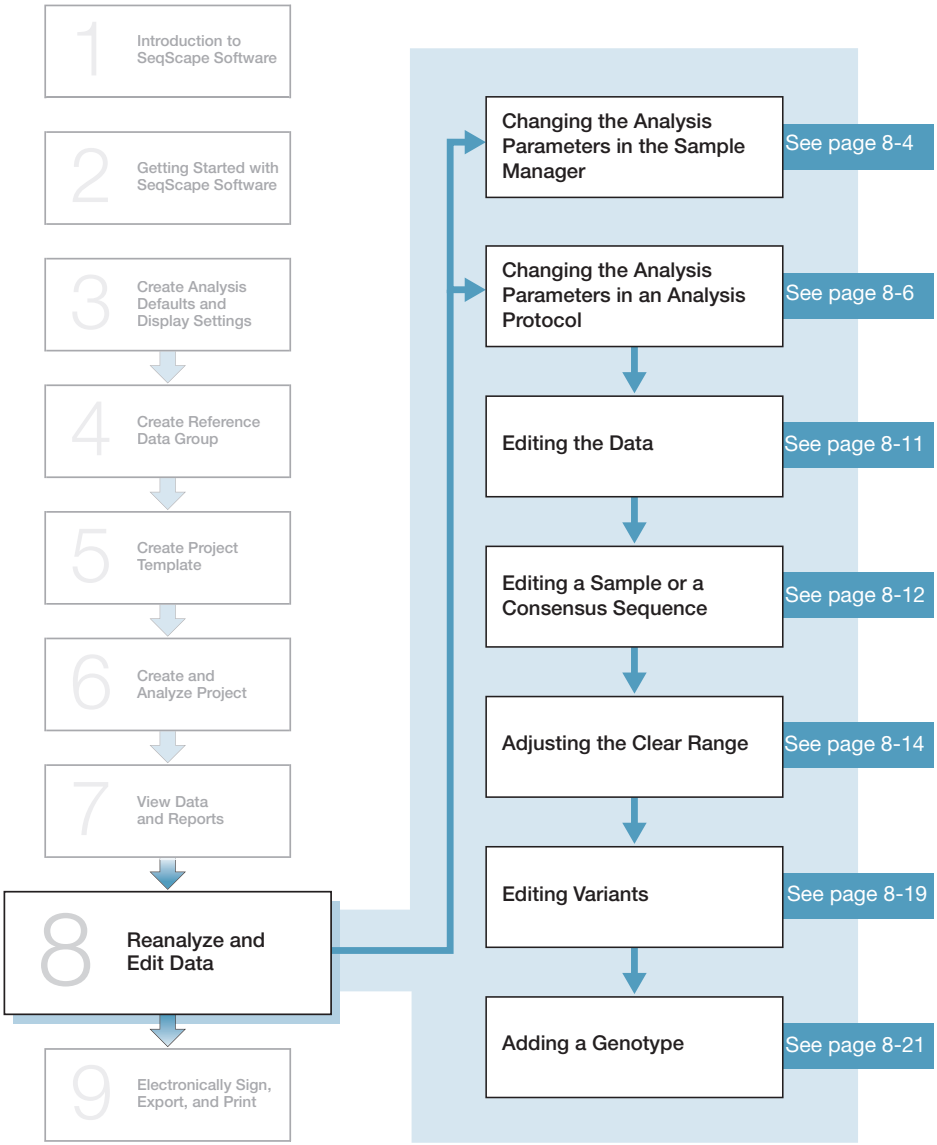
3. Repeat steps 1 and 2 to deselect additional headings.

4. To redisplay a column, right-click any column heading, then select the column heading.

5. To sort the data A to Z or Z to A in a Sample Details or Errors table column, double-click the column heading. Double-click again to sort in the opposite direction.

6. To customize the table header and footer information, see "Customizing Header and Footer Display" on page 9-22.

8 Reanalyzing and Editing the Data



About Analysis Parameters

Introduction

The analysis parameters (basecaller and DyeSet/Primer file) associated with every sample file are used when the sample files are analyzed.

Sometimes poor project results can be corrected or improved by changing certain analysis settings and applying the new settings to the affected samples.

Common examples of errors that affect basecalling are:

- Incorrect stop point selected
- Bad base spacing
- Poor quality data
- Incorrect basecaller and/or dyeset/primer used for basecalling
- Wrong peak 1 location and start point calculated by the software

Note: Refer to the *Sequencing Analysis Software User Guide* for instructions on defining a new peak 1 and the start and stop point locations.

Viewing Analysis Parameters in the Sample Manager

You can use the Sample Manager to display sample files and their current analysis information including the basecaller and DyeSet/Primer files (see Figure 8-1 on page 8-3). The analysis parameters can be modified and applied to samples. You can apply these changes to one sample, some samples, or all samples in the Sample Manager.

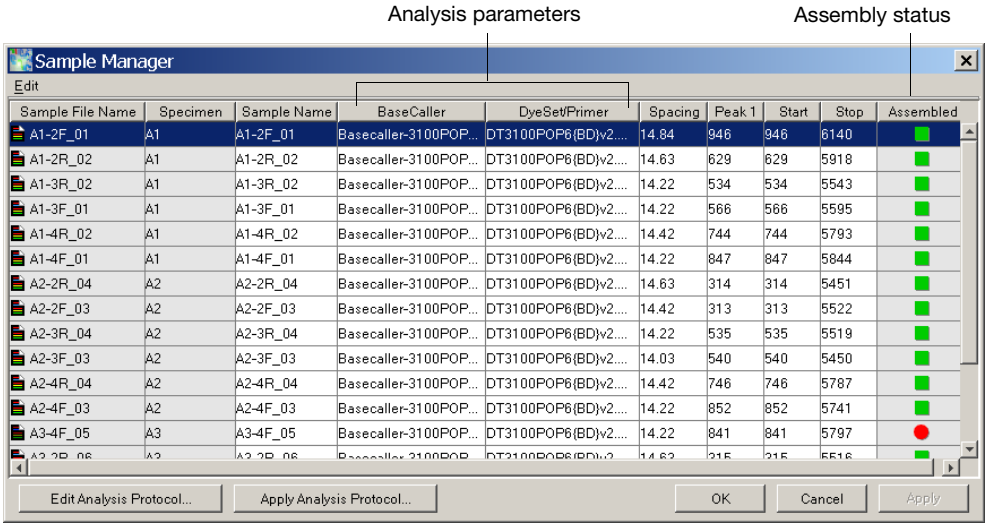


Figure 8-1 Sample files in the Sample Manager

The Sample Manager window information:

Column Heading	Description
Sample File Name	Information from the plate record and project. It cannot be changed in the Sample Manager.
Specimen	Information from the plate record and project. It cannot be changed in the Sample Manager.
Sample Name	Name of the sample, taken from the plate record. It can be changed.
Basecaller	Algorithm used to call the bases. It can be changed.
DyeSet/Primer	File that corrects for mobility shifts and color-code changes, depending on which chemistry was used. It can be changed. DyeSet/Primer files are sometimes known as mobility or .mob files. All mobility files have the extension *.mob*.

Column Heading	Description
Spacing	The number of scan points from the crest of one peak to the crest of the next peak. During basecalling, a spacing calibration curve is applied to the data to determine a base spacing value.
Peak 1	The first data point that is from the sample, not including primer peaks in dye primer chemistries. It is the reference point for the spacing and mobility corrections performed by the basecalling software.
Start	The raw data point where the basecalling starts in the sample file. The Start Point is normally the same as the beginning of the first base peak.
Stop	Specifies the last raw data point to be included in the basecalling. If the default Stop Point is used, this endpoint is the last data point in the file.
Assembled	Displays the assembly status of the sample. A green box indicates assembled, a red circle indicates not assembled.

Changing the Analysis Parameters in the Sample Manager

Adding Samples to the Sample Manager

To add samples to the Sample Manager:

1. Open the project of interest.
2. Select a layer in the Active Layer drop-down list.
3. In the navigation pane:

To add all …	Select the …
Samples in a project	Project icon
Samples in a specimen	Specimen
Selected samples in a segment	Segment
Selected sample(s)	Sample(s)*

* Use the Shift key to select contiguous samples, or use the Ctrl key to select noncontiguous samples.

4. Select **Analysis > Sample Manager**.

   The selected files are displayed in the Sample Manager.

Changing Basecaller and DyeSet/Primer Files

Note: Use the basecaller and DyeSet/Primer tables in Appendix B to select the correct combination of files.

To change the basecaller and/or DyeSet/Primer file:

1. Select **Analysis > Sample Manager** and then select the sample you want to change.
2. To edit the Analysis Protocol, click **Edit Analysis Protocol...** in the lower left corner of the window.
   a. In the Basecaller drop-down list, select a new basecaller.
   b. In the DyeSet/Primer drop-down list, select a new DyeSet/Primer file.
3. To change multiple samples, use the Fill Down function.
4. To apply a different Analysis Protocol, click **Apply Analysis Protocol**. Choose an Analysis Protocol that will be applied to the set of selected samples and click **OK**.
5. Click **Apply**.
6. Click **OK**.
7. Click  (Analyze).

Changing the Analysis Parameters in an Analysis Protocol

Editing an Analysis Protocol

To edit an analysis protocol:

1. Add the samples to the Sample Manager (see "Adding Samples to the Sample Manager" on page 8-4).
2. In the Sample Manager, click **Edit Analysis Protocol**.
3. If a sample-level HIM has been detected, in the Filter tab, you can select **Skip if sample-level HIM is detected**. This selection allows the sample to proceed to assembly after re-analysis. (See the discussion of HIM Detection that follows this procedure.)

   Note: For this selection to be applicable, the following conditions must apply for the sample:

   – A sample-level HIM has been detected and reported in the Analysis QC Report.
   – The minimum clear range and the sample score have values greater than zero.

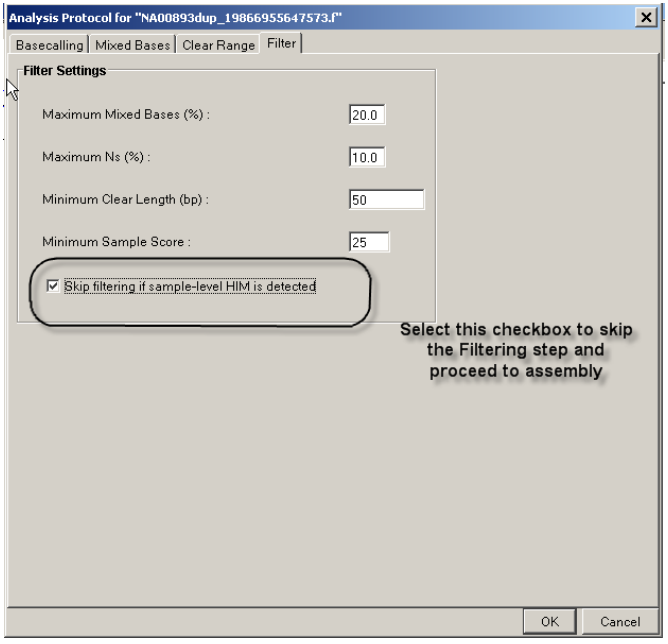


Figure 8-2 Filter tab of the Analysis Protocol Editor

4. Refer to "Creating an Analysis Protocol" on page 3-2 for descriptions of all the Analysis Protocol Editor tabs.

5. Click **OK** to save the protocol and close the Analysis Protocol dialog box.

HIM Detection

Heterozygous insertion/deletion mutations (HIMs) are detected at two stages in SeqScape software v2.5.

In the first stage, HIMs are detected at the sample level after the basecaller has called pure and mixed bases and has assigned quality values. The HIMs are identified in each individual sequence trace and reported in the Analysis QC Report.

In the second stage of HIM detection, an HIM that is present in both the forward and reverse trace is carried over to the consensus sequence and is reported at the specimen-consensus level in the Mutations Report.

Example of HIM Detection

The Initial Analysis

Figure 8-3 on page 8-8 shows that Specimen 1 has a forward and reverse trace. After analysis, the forward trace (NA 00893dup_19866955647573.f) appears in the unassembled mode in the project view.

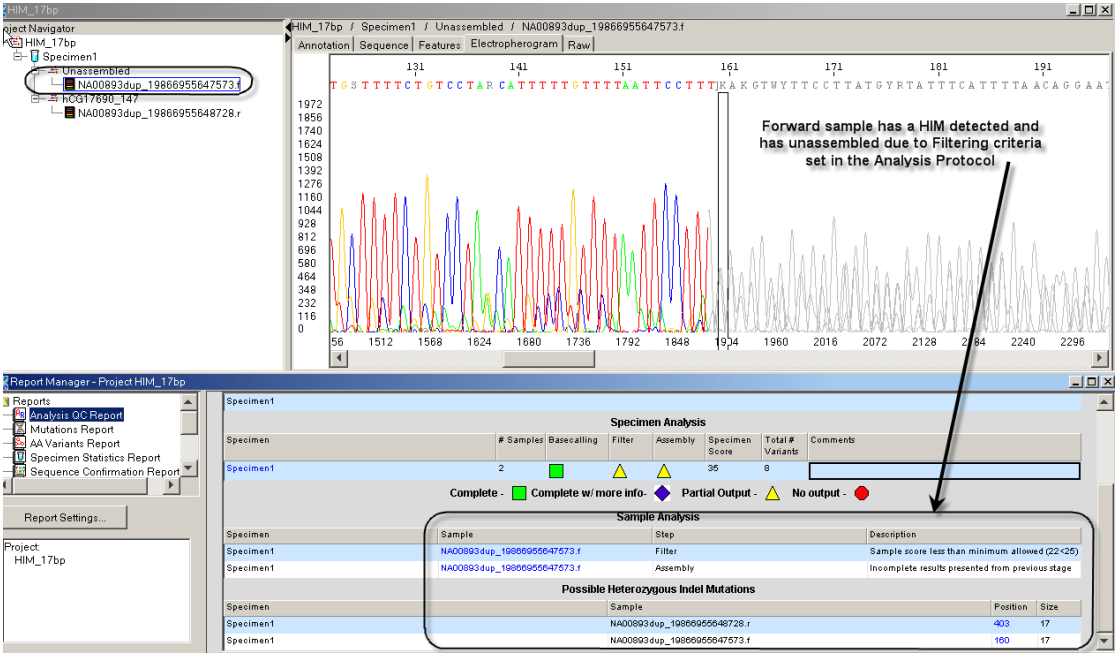


Figure 8-3 Analysis QC Report

Depending on the sequence quality and the criteria specified for filtering the data prior to assembly, the samples may or may not be assembled.

The Analysis QC Report shows that:

- An HIM is detected in this sample
- The sample has failed assembly because the sample score (the average quality value of the bases in the clear range) is less than the filter set criteria (value of 20 < 25) in the analysis protocol.

If you use the hyperlinks in the Analysis QC Report to look at the sequence trace, the reported HIM location may lie either within the clear range or outside the clear range (represented as the region of grayed-out bases). In Figure 8-3, the HIM is located at position 160, just outside of the clear range. As reported, the same HIM has been detected in the reverse trace for specimen 1, which has assembled.

Re-Analysis Without the Filtering Step

By selecting the same forward sample (NA 00893dup_19866955647573.f) and editing its as analysis protocol, the filtering step was bypassed for the forward trace and the data was re-analyzed.

Figure 8-4 displays the results for the project after the data were re-analyzed. The results show:

- An HIM in both the forward and reverse trace is carried over to the consensus sequence and reported at the specimen level in the Mutations Report.
- The sample is assembled as shown in the Project Navigator (Figure 8-4).

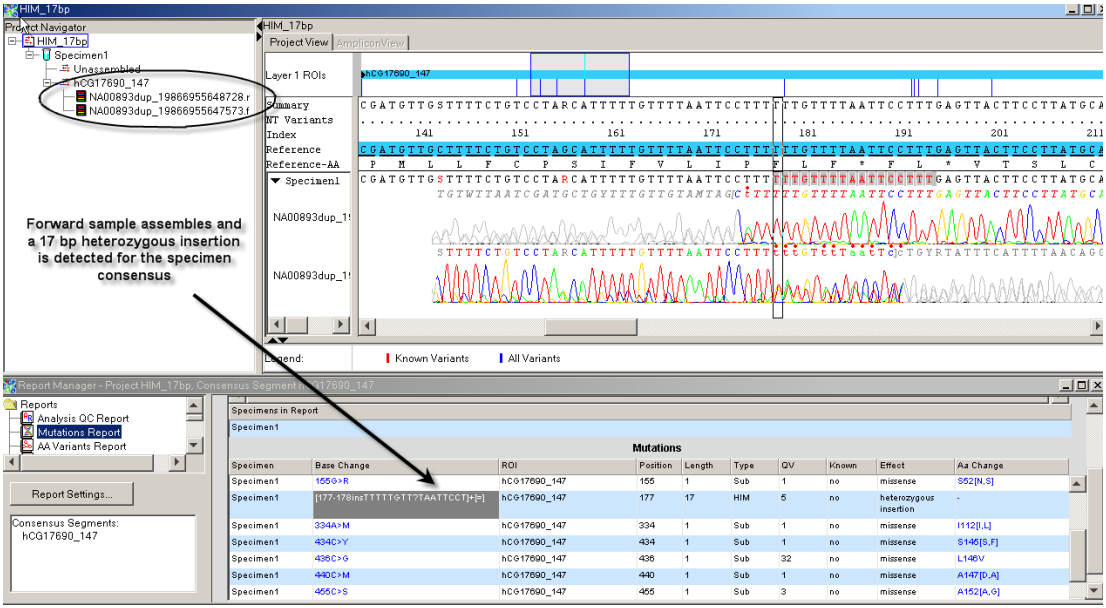
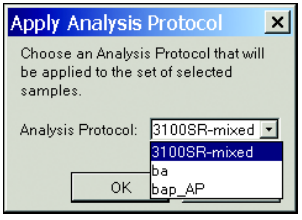


Figure 8-4 Mutations Report displaying an HIM in the consensus sequence

Applying the Analysis Protocol

To apply an analysis protocol:

1. Select the samples in the Sample Manager to apply the new settings to.
2. Click **Apply Analysis Protocol**.



3. Select a protocol from the Analysis Protocol drop-down list.
4. Click **OK**.

   The spacing, peak 1, and start and stop points change to zero.
5. Click **Apply**.

   The Assembled indicator changes from green (assembled) to red (unassembled), and the Analysis button becomes active.
6. Click  .

Editing the Data

About Sequence Editing

To edit a sequence, you can:

- Adjust the clear range
- Add, delete, or change a base in a sample
- Add or delete a space in a sample
- Add, delete, or change a base in a specimen consensus
- Add or delete a space in a specimen consensus
- Add or delete a space in a reference

You can edit sequences within a project. The change is immediately reflected in the consensus sequence. You can also edit the consensus sequence. In this case, all the samples change to reflect the consensus edits. You can edit consensus sequences when viewing the data in the Specimen view or in the Project view.

Note: An edited base change or insertion appears in lowercase to distinguish it from an unedited base. This applies to both user edits and consensus-caller edits. See "Editing Bases with Quality Values" on page A-11 for more information on editing bases with QVs.

When to Edit the Data

After analysis is complete, and you generate the analysis reports, depending on the results in the reports, you may want to:

- Adjust the clear range for a sample (see "Adjusting the Clear Range" on page 8-14)
- Edit a base or space in a sample or specimen (see "Editing a Sample or a Consensus Sequence" on page 8-12)

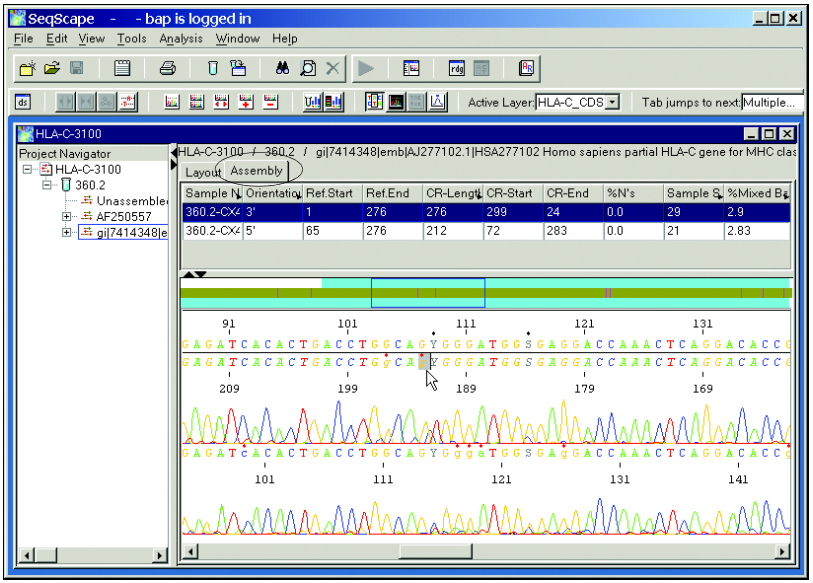
Editing a Sample or a Consensus Sequence

Editing a Consensus Sequence in the Segment View

To insert, delete, or change a base in the Specimen view:

1. Select the Segment icon in the navigation pane in the Project window.

   The Specimen view opens in the project document window.



2. Select the **Assembly** tab, then select a layer in the Active Layer drop-down list.

3. To change or delete a consensus base, click the base you want to edit, then delete or change the base.

4. To insert a base in the consensus sequence, click between two bases, then insert the bases.

Note: The changed bases appear in lowercase.

Note: If the audit feature is enabled, you must enter a reason for each base change, base insertion, and base deletion.

Editing Sample Bases

You can edit sample bases in the same manner as consensus bases. However, only the sample whose base is edited and the consensus sequence are affected by the changes.

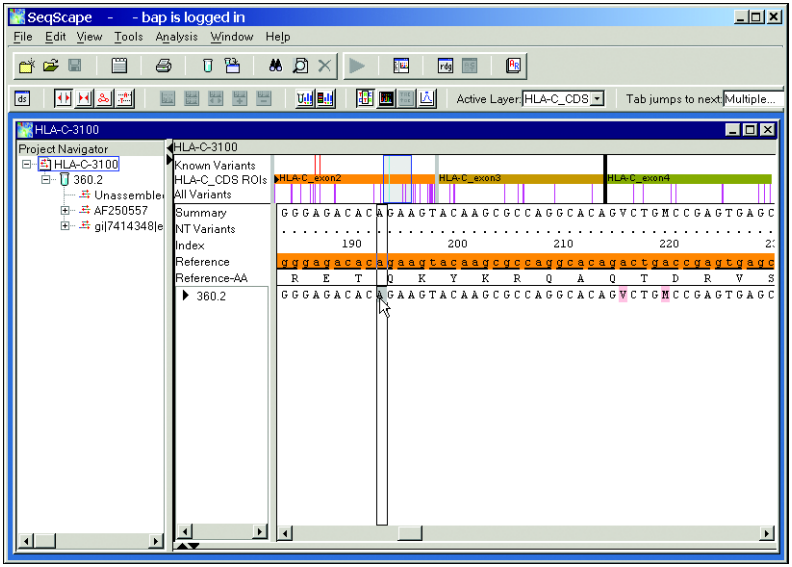
Editing a Consensus Sequence in the Project View

Note: Any changes are reflected in the sample sequences within the specimen and in the summary.

To insert, delete, or change a base in the Project view:

1. Select the project of interest in the navigation pane in the Project window.

   The project view opens, displaying the consensus sequences for each specimen.



2. To change or delete a consensus base, click the base you want to edit, then delete or change the base.

3. To insert a base in the consensus sequence, click between two bases, then insert the base.

   Note: The changed bases appear in lowercase.

4. To delete a space, click the space to select, then press **Delete**.

5. To insert a space, click where you want to insert a space, then press the dash key or space bar.

Note: If the audit feature is enabled, you must enter a reason for each base change, base insertion, and base deletion.

Adjusting the Clear Range

About the Clear Range

Sample data usually has unreadable or otherwise unusable sequence at the beginning and end of the data. Inclusion of this data causes errors in the alignments and erroneous variant detection.

The clear range is the area of continuous sequence that is the most error free. In SeqScape® software, the clear range is set automatically for all samples during the analysis based on the Analysis Settings for that sample. You can modify the clear range on a per-sample basis.

IMPORTANT! If you do not select Use Reference Trimming in the Analysis Settings, you should manually set the clear range to remove any sample data that lies 5′ of the 5′end or 3′of the 3′end of the reference, if needed. Any sample data that are outside the reference is not aligned.

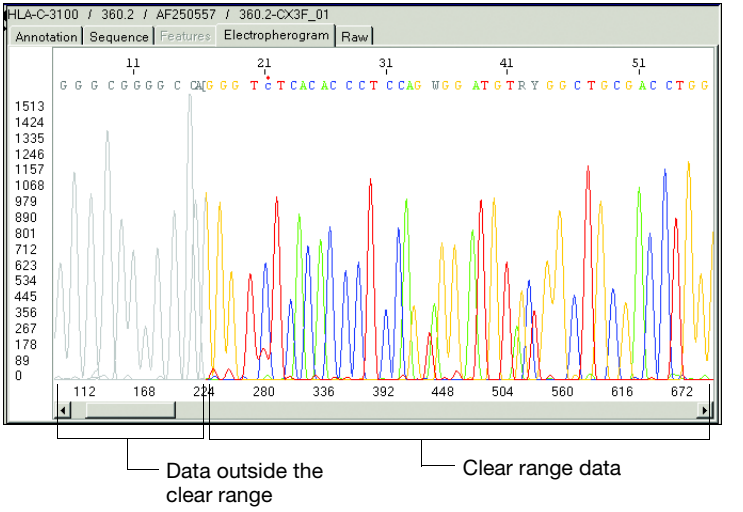


Figure 8-5 Clear range data

After changing the clear range, the specimen is automatically reassembled, then realigned and recompared to the reference.

The three methods for changing the clear range involve using the:

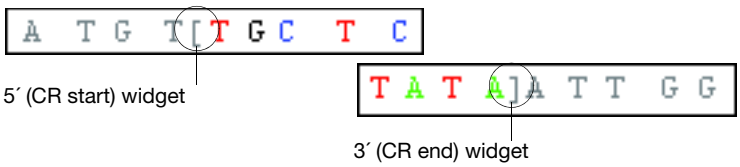
- Clear range widgets
- Mouse
- Set Clear Range dialog box

Using the Clear Range Widget

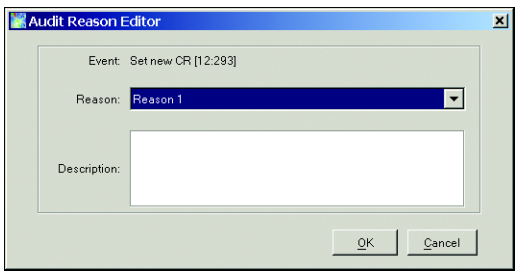
To use the Clear Range widget to adjust the clear range:

1. Open a sample from within a project.
2. Select the **Electropherogram** tab.
3. Locate and select the 5′ (CR start) or 3′(CR end) widget.

   The widget turns from gray to black, when selected.



4. Drag the widget along the bases to the right or left, as desired, then release the cursor.
5. If the audit feature is enabled, an Audit Reason Editor opens.



6. Complete the Audit Reason Editor dialog box, then click **OK**.

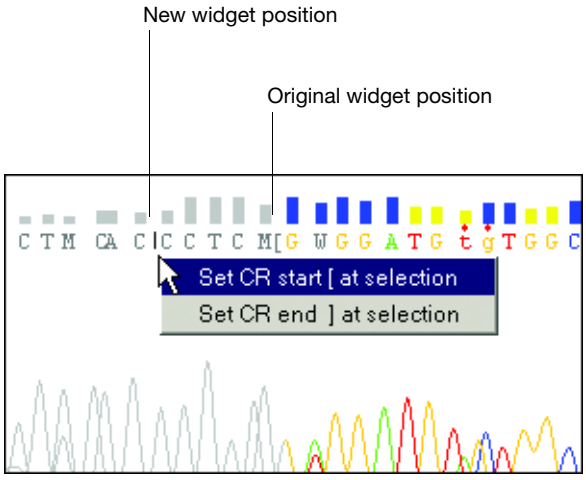
   The new clear range is displayed.

7. Repeat the process to define a new clear range for the opposite end.

Using the Mouse

To use the mouse to adjust the clear range:

1. Open a sample from within a project
2. Select the **Electropherogram** tab.
3. Right-click between two bases where you want to move the 5′ (CR start) or 3′(CR end) widget.



4. Do one of the following:

If you are moving the…	Then select…
CR start widget	**Set CR start [at selection**
CR end widget	**Set CR end] at selection**

5. If the audit feature is enabled, select a reason in the Audit Reason Editor, then click **OK**.

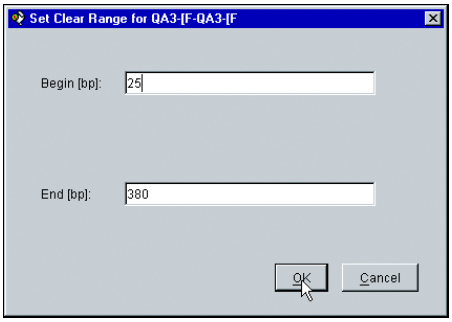
   The new clear range is displayed.

6. Repeat the process to define a new CR widget position for the opposite end.

Using the Set Clear Range Dialog Box

To use the dialog box to adjust the clear range:

1. Open a sample from within a project.
2. In the Electropherogram view or Specimen view, determine your new beginning and ending base numbers.
3. Select **Tools > Set Clear Range**.



4. Enter the values determined in step 2, then click **OK**.
5. If the audit feature is enabled, select a reason in the Audit Reason Editor, then click **OK**.

   The new clear range is displayed.

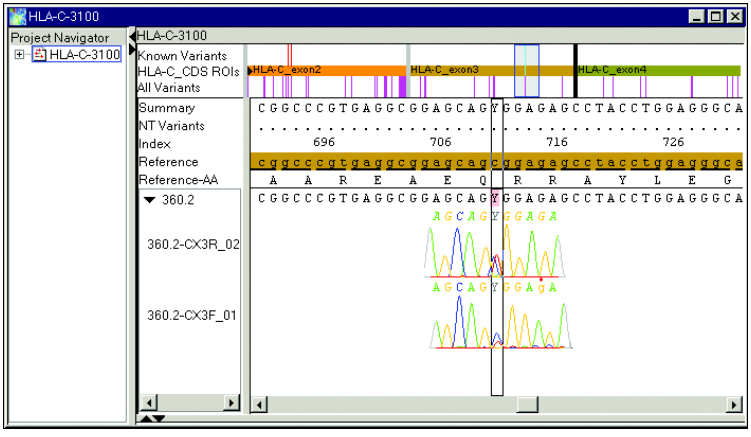
Editing Variants

After you clean up errors in the sequences, you can review and edit the variants. There are two methods to review variants.

Method 1 To view and edit variant data:

1. Open a project of interest.
2. Click a summary base.
3. Click the triangle next to the specimen name to view the electropherogram snippets.
4. Edit bases or spaces in the specimen consensus sequences or the displayed sample data.
5. Press **Tab** to move to the next variant or press **Shift-Tab** to move to the previous variant to view or edit more positions.

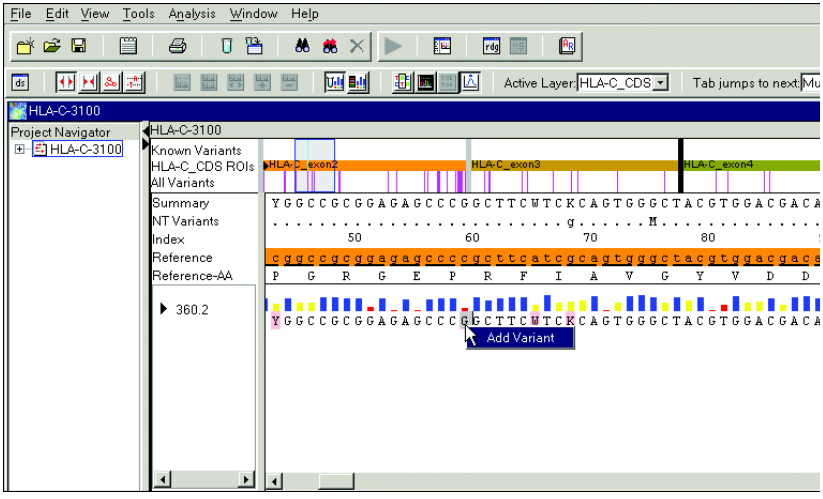
   Note: Pressing Ctrl+Z centers the selected column in the display, even if snippets are not showing.



8

Method 2 To view and edit variant data:

1. Open the project of interest.
2. Select **Analysis > Report Manager**.
3. In the navigation pane, select the report you want to view.
4. Select **Window > Tile**.
5. Review the positions by selecting a base change in the Mutations table. This adjusts the alignment view to the correct position in the alignment.
6. To add an unknown variant to the RDG, right-click the unknown variant position in a consensus sequence in the project alignment, then click **Add Variant**.



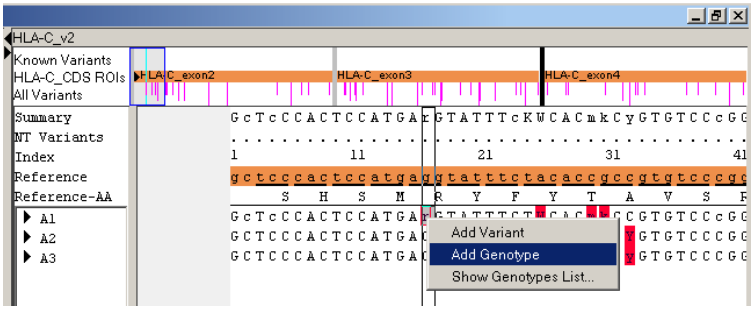
Saving Your Data

When you finish, select **File > Save Project** or click (Save Project).

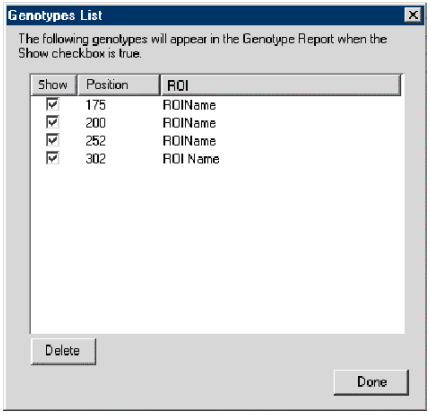
Adding a Genotype

To add a genotype to the Genotype Report:

1. In the Project View, select a specimen base, right-click the base, then select **Add Genotype**.

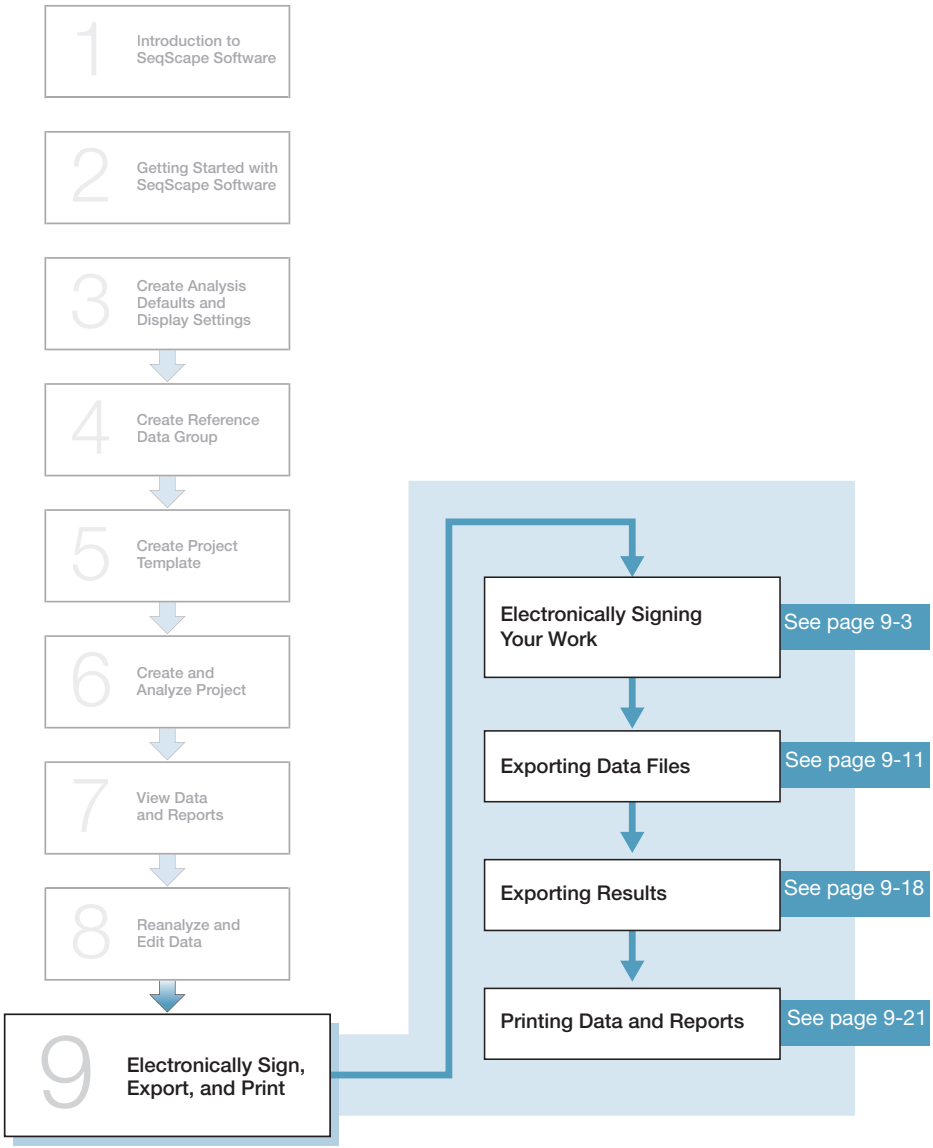


2. To view the genotype list, select the specimen base, right-click, and then select **Show Genotypes List**.



3. Select the **Show** checkbox to display the position in the Genotyping Report. See "Electronic Signature History Report" on page 7-39.

9 Electronic Signing, Exporting, and Printing Data and Reports



Section 9.1 Electronic Signatures

In This Section

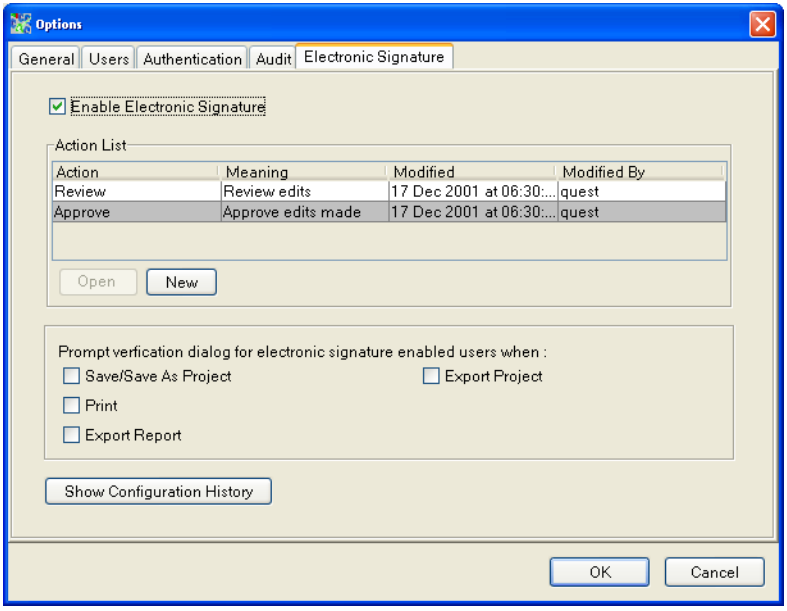
SeqScape Software v2.5 (and higher) allows you to electronically sign when saving, printing and/or exporting an analysis report. You can also sign at any other time. This feature is turned off by default.

Enabling Electronic Signatures

To use the electronic signature feature, you must enable it for the application *and* for each user who you want to have signature privileges.

To set up your system for electronic signatures:

1. Log in as an Administrator.
2. Select **Tools > Options**.
3. Select the **Electronic Signatures** tab.



4. Select **Enable Electronic Signature**.

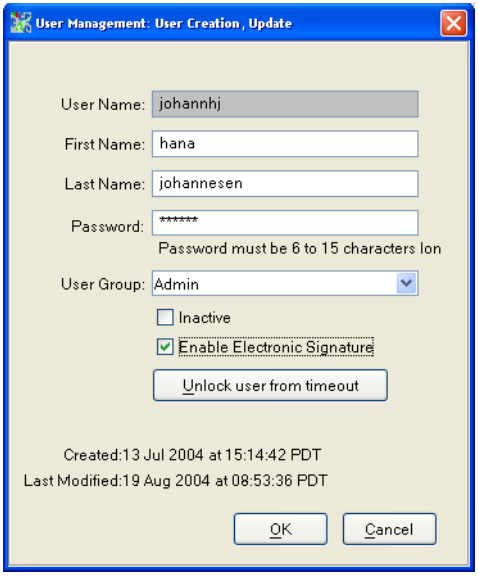
5. Select the actions for which you want to require electronic signatures.

   Every time an electronic-signature-enabled user performs one of these actions, a signature prompt is displayed.

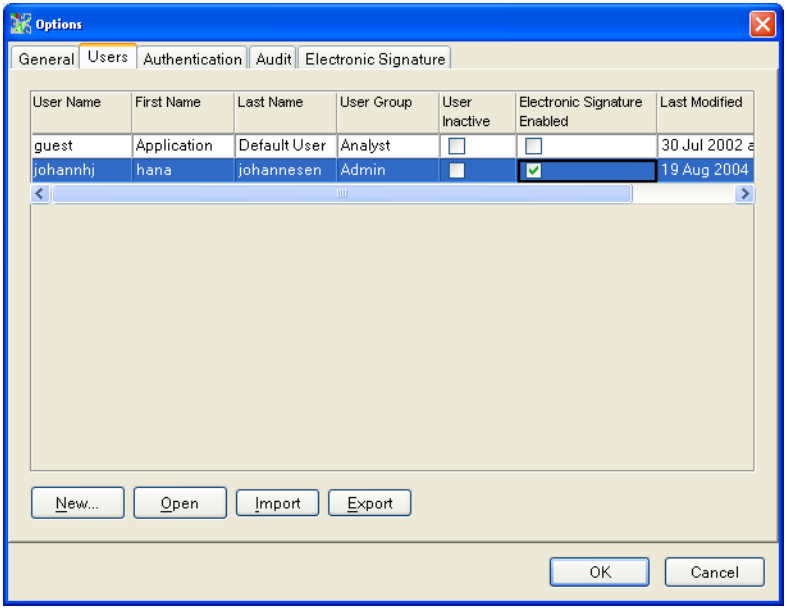
6. Select the **Users** tab.

7. Select a user, then click **Open**.

8. In the User Management dialog box, select **Enable Electronic Signature**, then click **OK**.



Note: In the Users tab, the Electronic Signature Enabled is selected.



9. Repeat for each user who should have electronic signature privileges.

10. Click **OK** to save.

Electronically Signing Your Work

After your system has electronic signatures enabled, you can sign off as specified in the Electronic Signature tab.

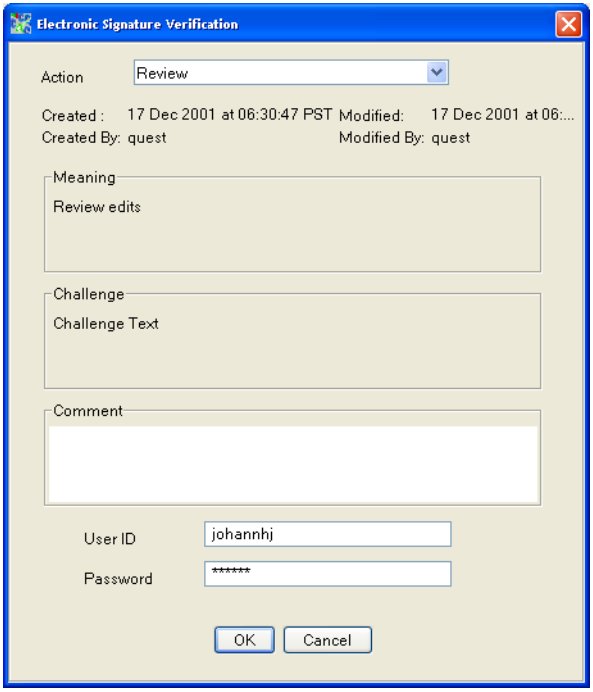
To electronically sign your work:

1. Perform any of the actions specified for electronic signatures.

   or,

   Click on a row header in the Sample Manager to select a sample, then select **Tools** > E**lectronic Signature** > **Sign**.

   The Electronic Signature Verification dialog box opens.



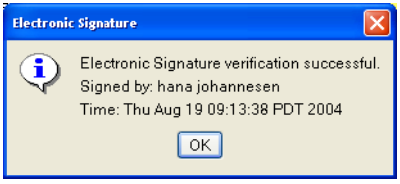
2. In the dialog box, select an action from the list.

3. Enter any comments, your user ID, and your password.

   Note: If you do not have electronic signature privileges, another user who does have them may sign with their user ID and password.

4. Click **OK**.

   If the signature is correct, the following message is displayed.



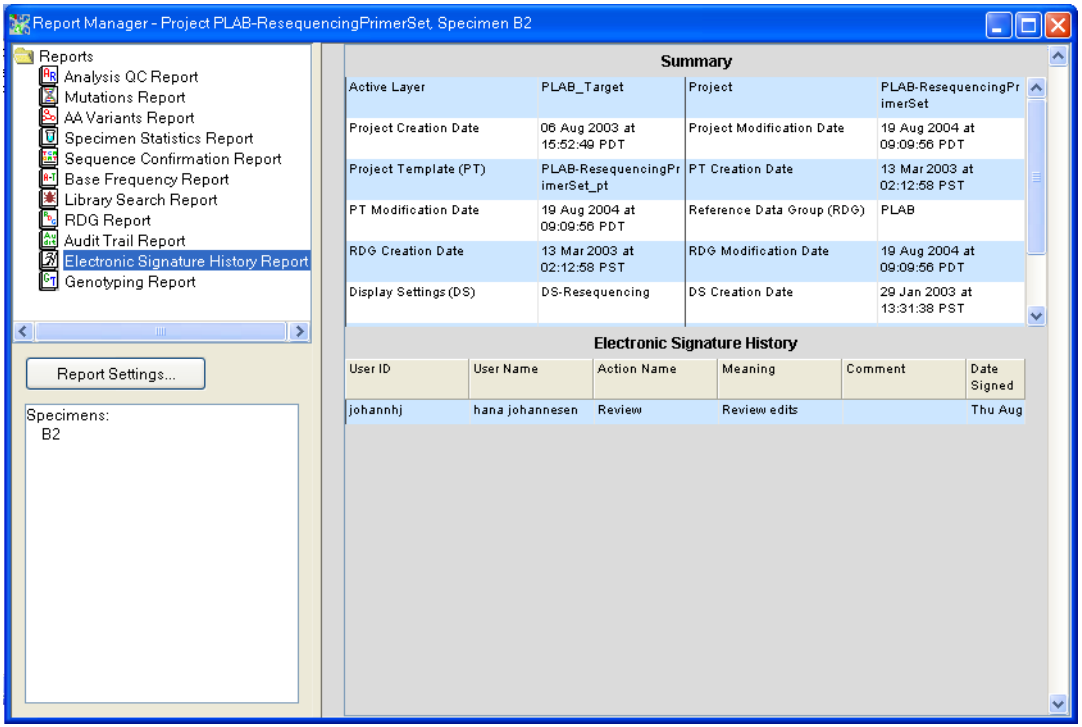
5. Click **OK** to close the message.

Viewing, Verifying, and Modifying Electronic Signatures

Viewing the Electronic Signature History

To view the electronic signature history for a project:

1. In the main tool bar, click  to open the Report Manager.
2. Select the **Electronic Signature History Report**.



3. (Optional) Perform the following tasks:
   - Export the report – Select **File** > **Export** > **Report**.
   - Display Challenge Text – Click **Report Settings**, select the **Electronic Signature Settings** tab, then select the **Show Challenge Text** check box.

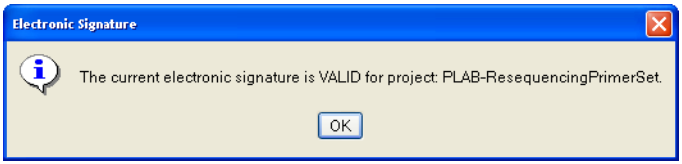
Verifying a Project

If you have electronically signed a project, you can verify whether or not it has changed since you signed.

To determine if a project has changed since the last signature:

1. Select a project, then click **Tools > Electronic Signature > Verify**.

   If the project has not changed, the following message is displayed.



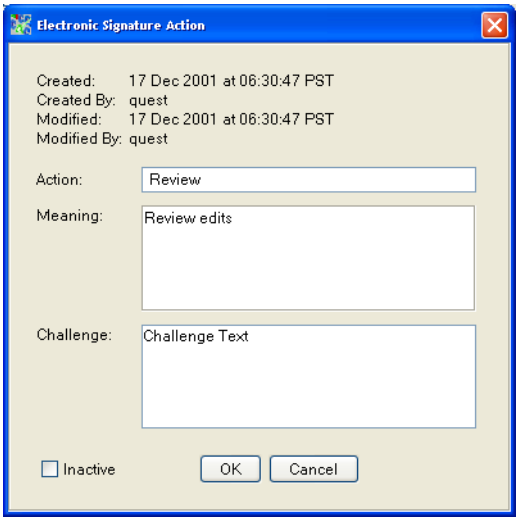
Modifying the Actions for an Electronic Signature

Each time you electronically sign, you must select an action from the list. The software comes with two built-in actions, Review and Approve. Administrators can modify them or add their own.

To add or modify the actions for electronic signatures:

1. Log in as an Administrator.
2. Select **Tools > Options**.
3. Select the **Electronic Signatures** tab.
4. To modify an action, select the row, then click **Open**.

   The Electronic Signature Action dialog box opens.



5. Make any changes, then click **OK**.

   The fields for an electronic signature action are:

   - **Action** – The name of the action.
   - **Meaning** – The meaning of the action.
   - **Challenge** – The legal implications of the action.

   If you select **Inactive**, the action appears shaded in gray in the list of actions in the Electronic Signatures tab in the Options dialog box. In addition, the action no longer appears in the list of actions for an electronic signature.

Section 9.2 Exporting

In This Section

Exporting Data Files

File Names

The default file name uses the project name and the report type. Do not use the following characters in any file name:
\ / : * ? “ < > | & and space

Format Options

You can export a project, specimen, segment, or sample file. Table 9-1 summarizes the available format options for files that you export. Header and footer information is not incorporated in any data file.

Note: Only one data file can be exported at a time.

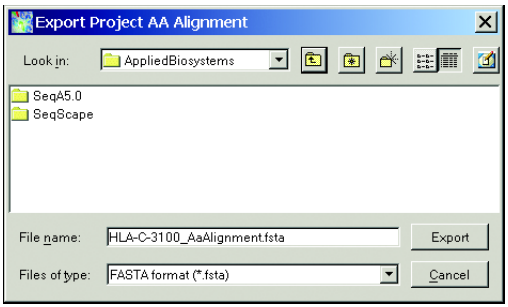
Table 9-1 Export and file format options

Export Option	File Format Options
Project	
Project Alignment-Nucleotides	FASTA
Project Alignment-Amino Acids	
Specimen and Segment	
Consensus Sequence	FASTA, SEQ, or QUAL
Aligned Sample Sequence	FASTA
Sample	
Sample Sequence File	FASTA, SEQ, AB1, or PHD

Exporting a Project Alignment

To export a project alignment:

1. Open the project of interest.
2. In the navigation pane, select the project icon.
3. Select **File > Export > Project Alignment-Nucleotides** or **Project Alignment-Amino Acid**.



4. Complete the Export Project dialog box:
   a. Select a folder location to store the project view.
   b. Change the file name, if desired.

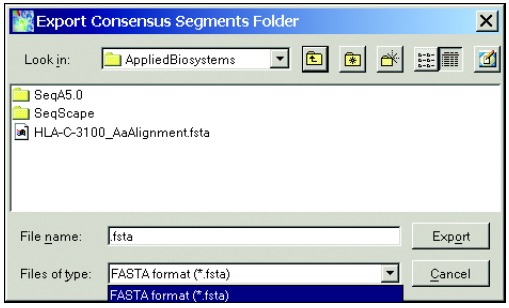
      Note: The default file name uses the project name with the element type suffix and the FASTA extension.

   c. Click **Export**.

Exporting a Specimen

To export a specimen:

1. Open the project of interest.
2. In the navigation pane, select a Specimen icon.
3. Select **File > Export > Consensus Sequence** or **Aligned Sample Sequence**.



4. Complete the Export Consensus dialog box:

   a. Select a folder location to store the file.

   b. Use the table below to decide whether to change the file name.

      Note: The default file name uses the project name with the element type suffix and the FASTA extension.

If the number of segments in a project is …	Then …
One	Change the file name, if desired.
Two or more	Do *not* type a file name. **Note:** The individual segment names are used. Any name you type is ignored.

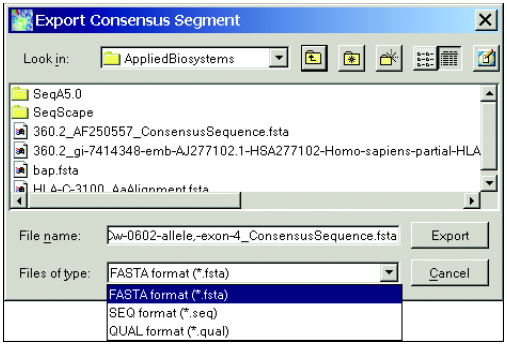
   c. For the Consensus Sequence option, select a file format in the Files of type drop-down list.

   d. Click **Export**.

Exporting a Segment

To export a segment:

1. Open the project of interest.
2. In the navigation pane, select a segment icon.
3. Select **File > Export > Consensus Sequence** or **Aligned Sample Sequence**.



4. Complete the Export dialog box:
   a. Select a folder location to store the file.
   b. Change the file name, if desired. The default file name uses the segment name and the FASTA extension.
   c. Select a file format in the Files of type drop-down list.
   d. Click **Export**.

Replacing ? with Another Character When Exporting a Consensus Sequence

You can replace any question marks (?) in a consensus sequence replaced with any other character when exporting a consensus sequence.

Question marks are typically seen in cases of a discontiguous consensus sequence. When you export your consensus sequence, you can replace the question marks with another character if your downstream application does not recognize the ? character.

To replace ? with another character when exporting a consensus sequence:

1. Select **Tools** > **Options**.

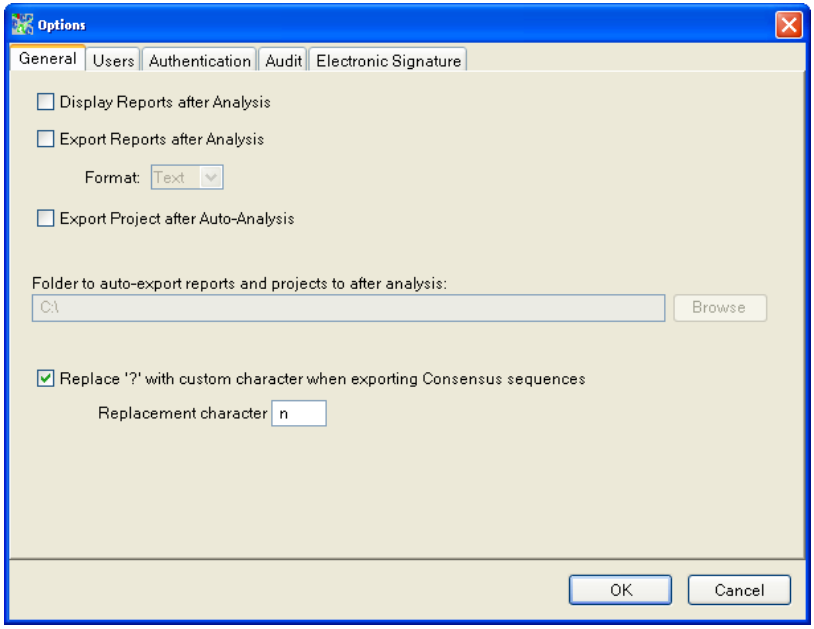


Figure 9-1 General tab of the Options dialog box

2. In the General tab, select the **Replace "?" with custom character when exporting Consensus sequences**.

   The results of this replacement are shown in Figure 9-2.

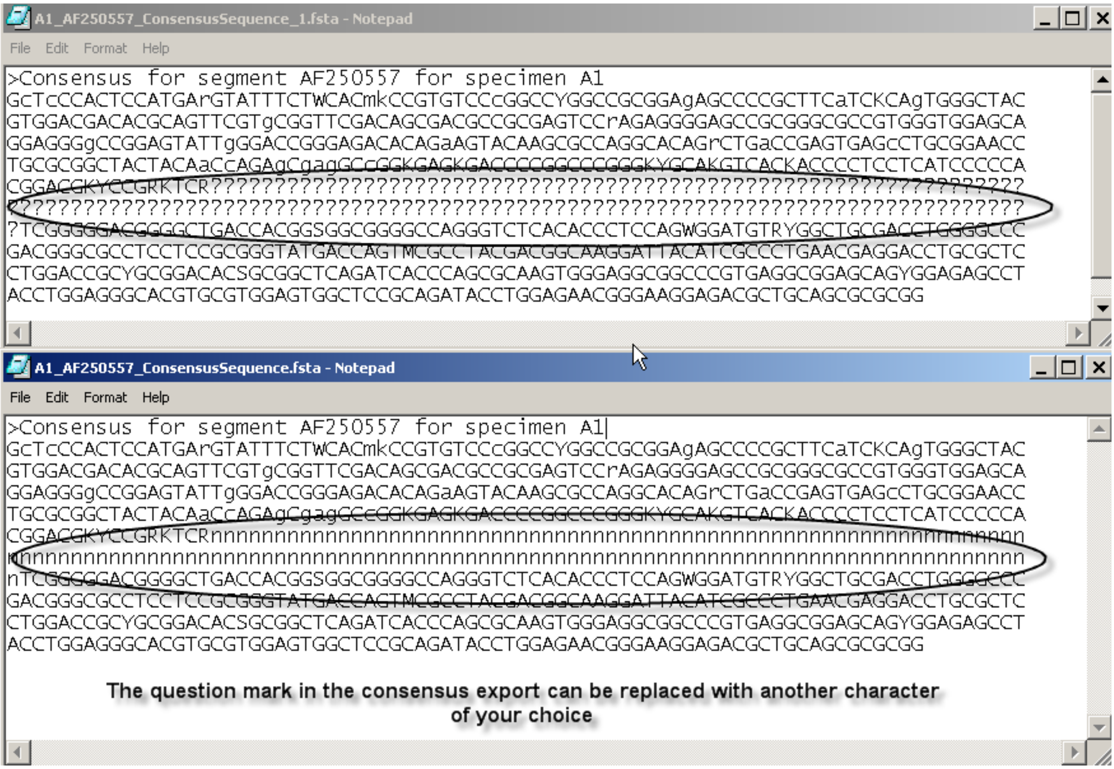
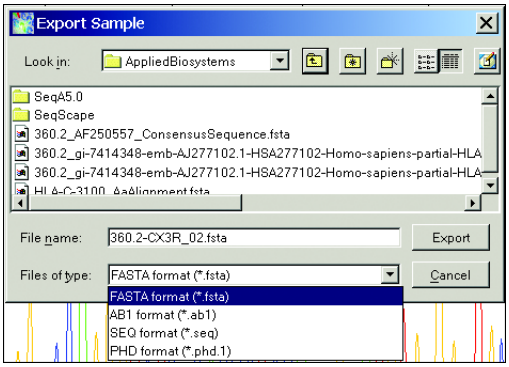


Figure 9-2 Question marks replaced by the letter *n* in the consensus sequence

Exporting a Sample

To export a sample:

1. Open the project of interest.
2. In the navigation pane, select a sample icon.
3. Select **File > Export > Sample Sequence File**.



4. Complete the Export dialog box:
   a. Select a folder location to store the file.
   b. Change the file name, if desired. The default file name uses the sample name and the FASTA extension.
   c. Select a file format in the Files of type drop-down list.
   d. Click **Export**.

Exporting Reports

File Names

The default file name uses the project name and the report type.

Do not use the following characters in a file name:
\ /: * ? > | and space

Format Options

You can export generated reports as text or in portable document format (pdf), HTML, or XML (Table 9-2).

Note: When choosing between HTML and XML, use HTML for standard display and XML for scripting applications.

Table 9-2 File formats and corresponding application options

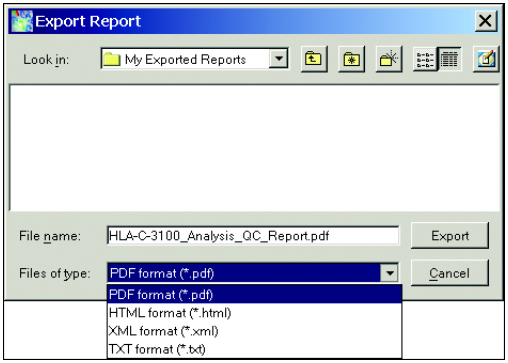
File Format	Open with...
PDF (default)	Adobe® Acrobat® Reader™
HTML*	A web browser or any software that is able to display HTML files
XML	A web browser
TXT	Notepad, Wordpad, Microsoft® Word, or any text-compatible software

*When exporting the report as HTML, a folder is automatically created that may contain more than one HTML file. The file that uses only the report name contains all the data from the report.

Exporting a Report

To export a report:

1. Open the project of interest, then click  .
2. In the navigation pane, select a report type.
3. Customize the report, if desired. (See "Customizing the Reports" on page 7-43.)
4. Select **File > Export > Report**.

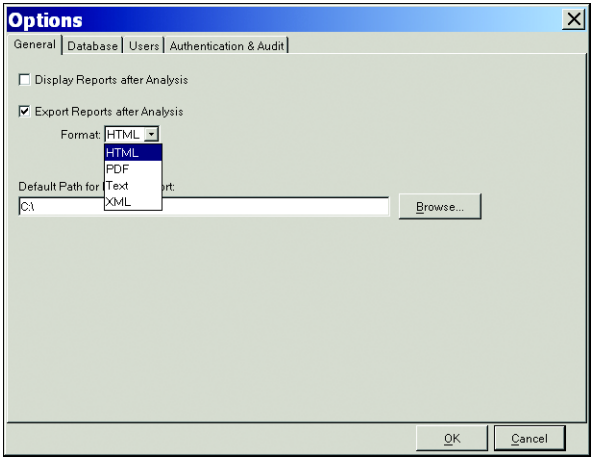


5. Complete the Export Report dialog box:
   a. Select a folder location to store the report.
   b. Change the file name of the report, if desired. The default file name uses the project name, the report type, and the .pdf extension.
   c. Select a file format in the Files of type drop-down list.
   d. Click **Export**.

Exporting All Reports Automatically

To set up for automatic exporting of reports:

1. Select **Tool > Options**.



2. Complete the General tab of the dialog box:
   a. Select **Display Reports after Analysis**, if desired.
   b. Select **Export Reports after Analysis**, then select an export format from the drop-down list.
   c. Define a default location to save the exported files.
   d. Click **OK**.

Section 9.3 Printing Data and Reports

In This Section

About Printing

You can print any viewable screen in a WYSIWYG (what you see is what you get) manner within the SeqScape® software. You can print project, specimen, segment, and sample views, as well as the reports for a project.

Table 9-3 describes what prints when you select a project, specimen, segment, or sample to print.

Table 9-3 Printing options

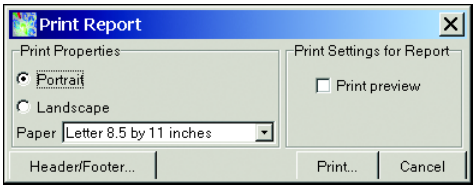
Object Selected	Option Selected	
	Only the Visible Data	**All Data**
Project	Prints visible Electropherogram view	Prints only summary, not the electropherogram
Specimen	Prints only visible data	N.A.
Segment	Layout tab: Prints only visible data Assembly tab: Prints only visible data	Layout tab: N.A. Assembly tab: Prints all data (applies only to the assembled region)
Sample file	Prints only visible data	Prints all data (applies only to the Electropherogram view)

Customizing Header and Footer Display

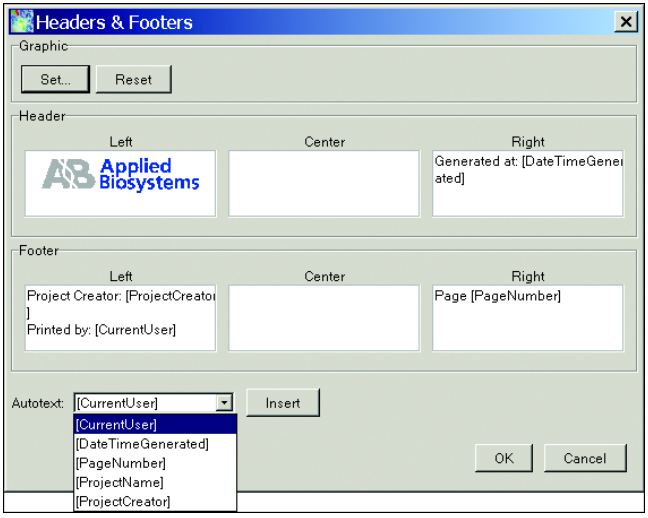
Default header and footer information is included in all exported and printed reports and in printed data views. However, headers and footers are not included in exported data files.

To customize the header/footer display in printed and exported reports:

1. Select **File > Print**.



2. Click **Header/Footer**.



3. To change the graphic, if desired:
   a. In the Graphic section, click **Set**.
   b. In the dialog box, locate, then select a graphic file.
   c. Click **OK**.

Note: The graphic is displayed in the Headers & Footers dialog box and in the upper left corner of printed or exported reports.

4. To change the header and/or footer information, do one of the following:
   - Type text into any of the header and/or footer text boxes.
   - Use the autotext variables from the Autotext drop-down list. (Insert the cursor in a text box, select an autotext option in the drop-down list, then click **Insert**).
   - Use a combination of typing text and using the autotext variables.

5. Save the changes:
   a. Click **OK** to close the Header & Footer dialog box.
   b. In the Print Report dialog box, click **Print**, then click:
      - **OK** to save the changes and print.

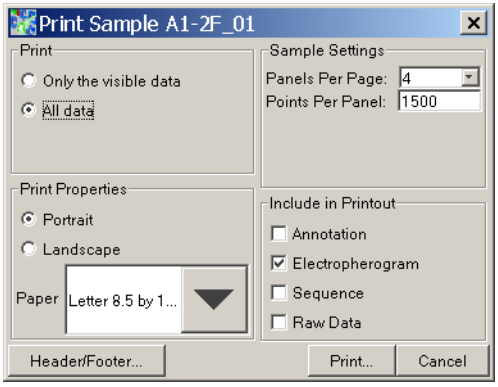
      or,
      - **Cancel** to save the changes without printing.

Printing Views

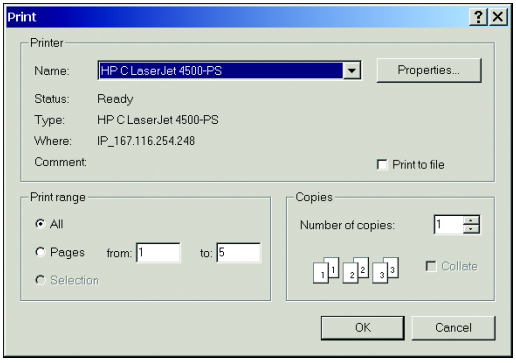
Printing Views of a Sample File

To print different views of a sample file:

1. Open the project of interest.
2. In the navigation pane, select a sample to print.
3. Select **File** > **Print** or click  .



4. Complete the Print Sample dialog box:
   a. In the Print section, select **All data**.
   b. In the Print Properties section, select the paper orientation and size.
   c. In the Sample Settings section, select a value in the Panels Per Page drop-down list. The range is 1 to 4, and the default is 4.
   d. Select a value in the Points Per Panel value box. The range is 100 to 12000, and the default is 1500 (about 120 bases).
   e. In the Include a Printout section, select the views you want to print.
   f. If you are printing a Segment assembly, type a new value in the Bases per panel field in the Print Settings for Project section.
   g. Click **Print**.



5. Select a printer, then click **OK**.

   Both print dialog boxes close and printing begins.

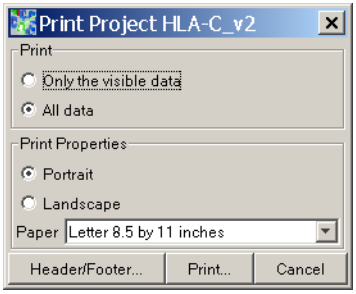
Printing Various Views of a Project

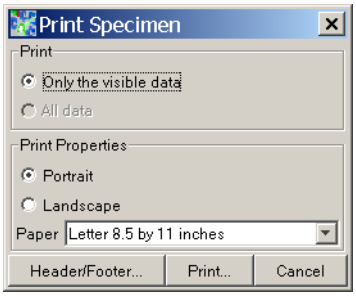
To print different views of a project:

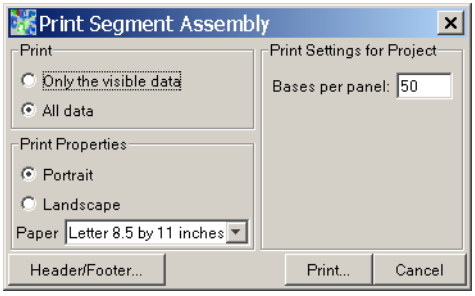
1. Open the project of interest.
2. In the navigation pane, select a view (Project, Specimen, or Segment) to print.

3. If you are using WYSIWYG, scroll to the area of the view you want to print.

4. Select **File > Print** or click .







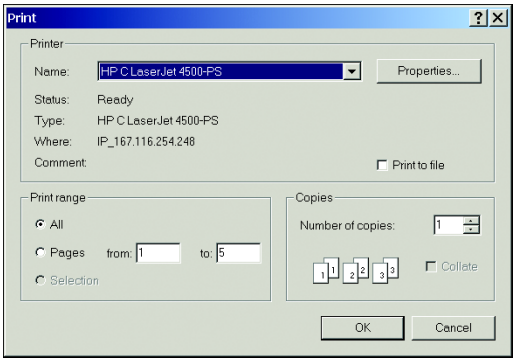
5. Complete the dialog box:

   a. In the Print section, select **Only the visible data** (WYSIWYG) or **All data** (if available).

   b. In the Print Properties section, select the paper orientation and size.

c. If you are printing a Segment assembly, type a new value in the Bases per panel field in the Print Settings for Project section.

d. Click **Print**.



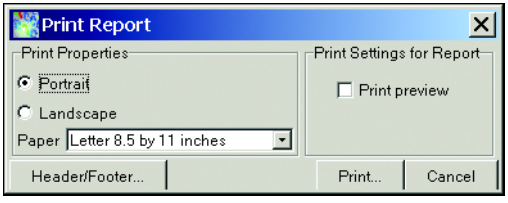
6. Select a printer, then click **OK**.

   Both print dialog boxes close and printing begins.

Printing a Report

To print a report:

1. Open the project of interest, then click  .
2. In the navigation pane, select a report type.
3. Customize the report, if desired (see "Customizing the Reports" on page 7-43).
4. Select **File > Print**.

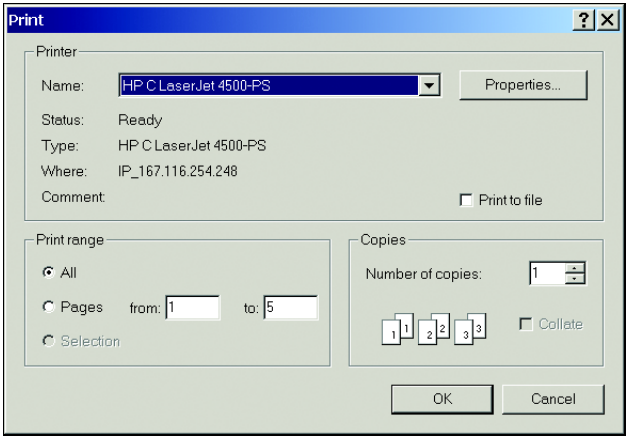


5. Complete the Print Report dialog box:
   a. In the Print Properties section, select the paper orientation and size.
   b. In the Print Settings for Report section, select **Print preview**, if desired.
   c. Click **Print**.

6. Use the following table to determine your next step:

If the Print preview option was …	Proceed to step …
Not selected	7
Selected	8

7. The Print dialog box opens:



Select the printer and define the page range, then click **OK**.

8. The Report Preview dialog box opens. Use the command buttons as described in Table 9-4 on page 9-29.

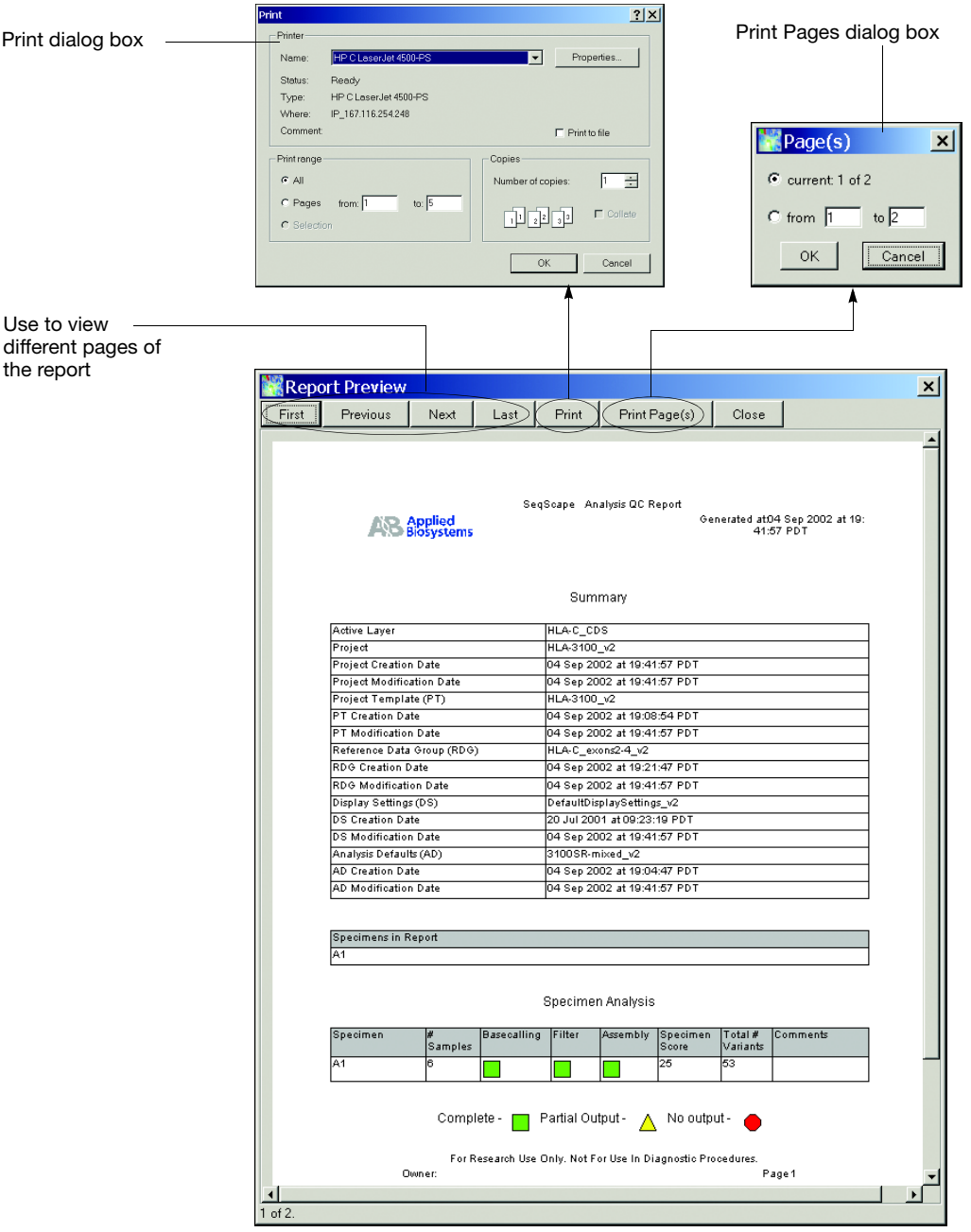


Figure 9-3 Report Preview dialog box

Table 9-4 Report Preview button functions

Button	Function
First, Previous, Next, and Last	Displays the various pages in a report (only one page is visible at a time).
Print	Opens the Print dialog box. Select a printer, then click **OK** to print the report.
Print Pages	Opens the Page(s) dialog box. Set the page range, then click **OK**. In the Print dialog box, click **OK** to print the report. **Note:** Page(s) dialog box settings override the settings in the standard Print dialog box.
Close	Closes the preview window without printing the report.

A

Sample and Consensus Quality Values

In This Appendix

Types of Quality Values (QVs)

Table A-1 summarizes the types of QVs and where they are displayed.

Table A-1 Quality value types

Quality Value Type	Definition	Location
Sample QV	A per-base estimate of basecaller accuracy.	• Sample view • Specimen view • Project view
Sample Score	The average quality value of the bases in the clear range sequence for that sample.	Specimen Statistics report
Consensus QV	A per-base estimate of the accuracy of the consensus-calling algorithm.	• Specimen view • Project view
Consensus Score	The average quality value of the bases in the consensus sequence for that specimen.	• Analysis QC report • Specimen Statistics report
Mutation QV	A per-base estimate of basecaller accuracy.	Mutations report
QV for deletion mutation	Average of the quality values for the bases to the left and right of the deletion.	Mutations report
FGR Avg QV	Average quality value of the consensus sequence across the target region	Genotyping report
5′ Reg Coverage Avg QV	Average quality value of the consensus sequence at the 5′ region	Genotyping report
Exon Avg QV	Average quality value of the consensus sequence at the exon region	Genotyping report
Intron Avg QV	Average quality value of the consensus sequence at the intron region	Genotyping report

Sample Quality Values

Sample Quality Values

A sample quality value (SQV) is a per-base estimate of the basecaller accuracy. There are two types of basecallers that generate SQVs:

- KB – An algorithm that identifies mixed or pure bases, and generates sample quality values.
- ABI – Algorithm used in ABI PRISM® Sequencing Analysis Software v3.7 that identifies pure bases. Subsequently, the ™ software identifies mixed bases and generates sample quality values.

KB™ Basecaller and ABI algorithms can produce slightly different SQVs.

Interpreting the Sample Quality Values

Per-base SQVs are calibrated on a scale corresponding to:

$$QV = -10\log_{10}(Pe)$$

Where *Pe* is the probability of error of the basecall.

The KB™ Basecaller basecaller produces a QV range of 1 to 99, with 1 being low confidence and 99 being high confidence. See Table A-2, “Quality Values and Probabilities of Error,” on page A-4 for the probability of basecall errors for QVs ranging from 1 to 99.

Mixed base calls yield lower SQVs than pure base calls.

The typical QV range for pure base data is 25 to 50.

Quality Values for mixed base positions range from 1-20.

The size and color of QV bars for QV 50 to 99 are identical. To view the actual QV number, see “Displaying Quality Values” on page A-6.

Sample Score

A sample score is generated from SQVs. It is the average quality value of the bases in the clear range sequence for a sample.

Table A-2 Quality Values and Probabilities of Error

Quality Value = $-10\log_{10}(Pe)$					
where *Pe* is probability of error					
KB basecaller generates QVs from 1 to 99					
Typical high quality pure bases will have QV 20- 50					
Typical high quality mixed bases will have QV 10-20					
Size and color of QVs bars are identical for QVs 50-99					
QV	***Pe***	**QV**	***Pe***	**QV**	***Pe***
1	79%	21	0.790%	41	0.0079%
2	63%	22	0.630%	42	0.0063%
3	50%	23	0.500%	43	0.0050%
4	39%	24	0.390%	44	0.0039%
5	31%	25	0.310%	45	0.0031%
6	25%	26	0.250%	46	0.0025%
7	20%	27	0.200%	47	0.0020%
8	15%	28	0.150%	48	0.0015%
9	12%	29	0.120%	49	0.0012%
10	10%	30	0.100%	50	0.0010%
11	7.9%	31	0.079%	60	0.0001%
12	6.3%	32	0.063%	70	0.00001%
13	5.0%	33	0.050%	80	0.000001%
14	4.0%	34	0.040%	90	0.0000001%
15	3.2%	35	0.032%	99	0.00000012%
16	2.5%	36	0.025%		
17	2.0%	37	0.020%		
18	1.6%	38	0.016%		
19	1.3%	39	0.013%		
20	1.0%	40	0.010%		

Consensus Quality Values

A consensus quality value (QV) is a per-base estimate of the accuracy of the consensus-calling algorithm. If the SQVs are generated from the KB™ Basecaller basecaller, then the KB™ Basecaller consensus-calling algorithm is used to generate the QVs. If the SQVs are generated from an ABI basecaller and ™ software, then the ™ consensus-calling algorithm is used to generate the QVs.

The KB™ Basecaller basecaller and ™ consensus-calling algorithms can produce slightly different consensus QVs.

Interpreting the Consensus Quality Values

The degree of certainty of either consensus-calling algorithm is reflected by the per-base consensus QVs. A consensus QV is derived from a number of factors:

- How large a quality-value discrepancy exists between calls from the individual sample sequence strands
- The possible redundancy of calls from strands in the same orientation
- The possibility that the basecaller missed a mixed base

The possible values for the QVs range from 1 to 50. Higher numbers indicate calls that the algorithm determined with a measure of confidence, lower numbers indicate calls that might require user inspection to verify the correct answer. The consensus quality values are roughly calibrated to follow the same scale as the per-base sample quality values.

Consensus Score

A consensus score is generated from consensus QVs. It represents the average quality value of the bases in the consensus sequence for a specimen.

Displaying Quality Values

QVs are displayed as bars above each base in a sample (Figures A-1 and A-2). The height and color of a bar indicates its value. The taller the bar, the higher the QV. The color of a bar, which is associated with its value, is editable in the Display Settings.

Note: QV bar height and color are identical for QVs 50 to 99.

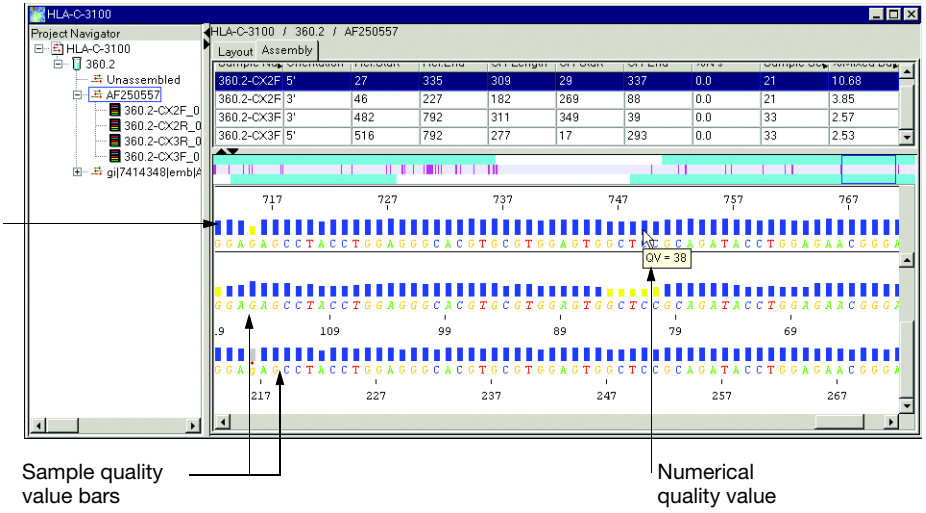


Figure A-1 Example of QV bars in the Specimen view

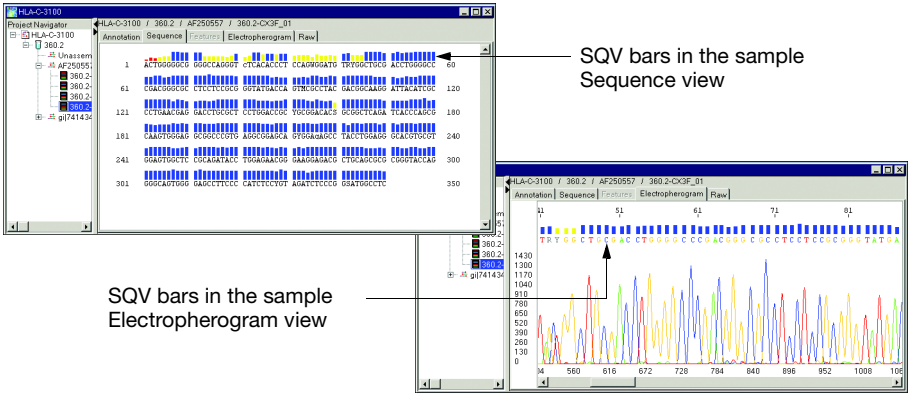


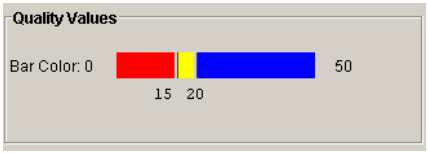
Figure A-2 Examples of SQVs in the Sample view

Customizing the Quality Value Display

You can modify the low, medium, and high ranges and the color associated with a QV.

To modify the QV display:

1. Select **Analysis > Display Settings** or click  .
2. Select the **Bases** tab.
3. In the **Quality Values section,** place the pointer between two colors (it becomes a double-headed arrow), then click the slider on the color bar and drag it to left or right to the desired value.



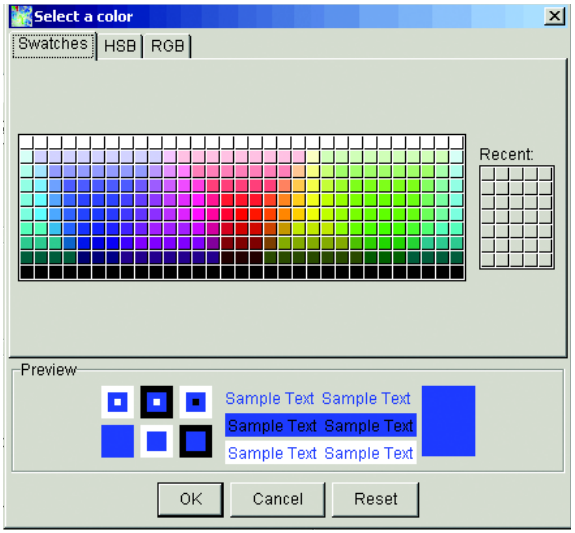
Use the criteria in the table below to define what values represent low, medium, and high ranges for your project. The Quality Values below refer to Pure Base Quality Values. Mixed Base Quality Values range from 0-20.

QV Bar	Default Color and Range	Set the range to identify data that ...
Low	Red 0 to 14	Are not acceptable
Medium	Yellow 15 to 19	Need manual review
High	Blue 20 or higher	Are acceptable

4. Change the colors that represent low, medium, and/or high QVs, if desired:

   a. Select the color in the Bar Code you want to change.

      The Select a color dialog box opens.



   b. Select a new color in the Swatches tab, or use the HSB or RGB tabs to define a new color.

   c. Click **OK**. The color dialog box closes.

5. Do one of the following:

   - Click **OK** to save the changes to the samples you are working with.
   - Click **Save to Manager As** to save the changes to the SeqScape® Manager.

Displaying the Quality Bars and Values

If you do not see the QV bars when viewing samples or a consensus in a project, then follow the procedures below to display QV bars and values.

To view quality bars and values:

1. Open a project, then open a specimen of interest.
2. Select the segment of interest, then select the **Assembly** tab.
3. To view sample QVs, select **View > Show/Hide sample QV** or click  .
4. To view consensus QVs, select **View > Show/Hide consensus QV** or click  .
5. To view a numerical value for a particular bar, place the cursor over the bar for 2 sec. The value is automatically displayed.

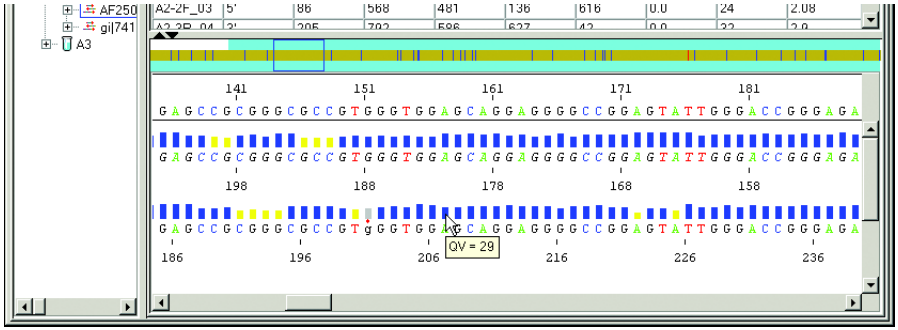


Figure A-3 Displaying the value of a sample QV bar

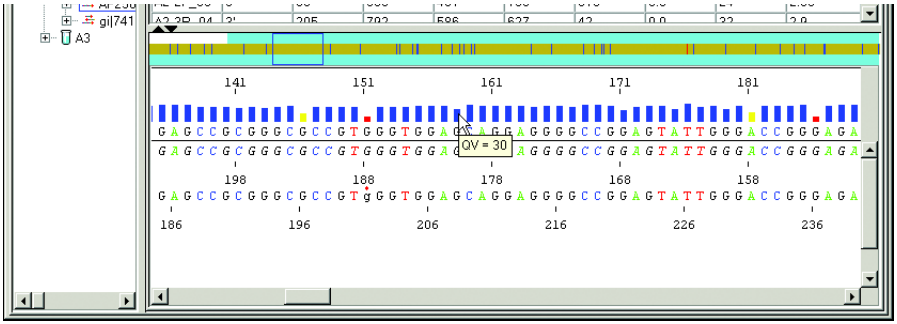


Figure A-4 Displaying the value of a consensus QV bar

Editing Bases with Quality Values

Changing, deleting, and inserting a base affect the consensus or sample QVs displayed.

If …	Then …
The consensus-caller calls a base not present in all the samples	The new base is in uppercase in the consensus sequence and in lowercase in the samples that did not contain that basecall with a red dot.
You change a base	The new base is in lowercase and the SQV has the same value but is displayed as a gray bar.
You change a base back to the original call	The base appears in uppercase and the quality value bar color is restored.
You insert a base	The inserted base appears in lowercase and it has no SQV.
You delete a base	The quality value for the base disappears.
You reinsert a deleted base	The reinserted base appears in lowercase and it has no SQV.

Cumulative Quality Value Scoring in Reports

Quality values and scores are also displayed in several reports. To view the reports, select **Analysis > Report Manager** or click  .

Analysis QC Report

Consensus scores in an Analysis QC report are shown as an average quality value across the consensus sequence for each specimen.

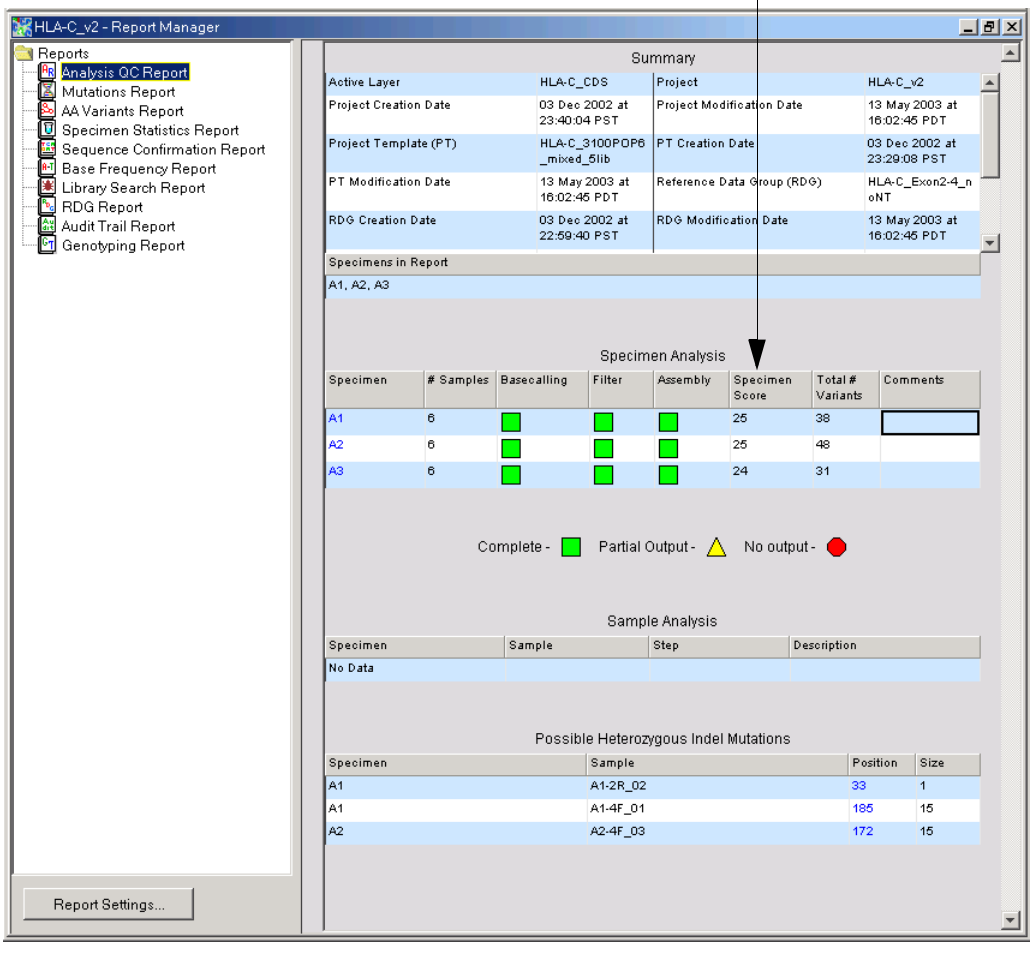


Figure A-5 Analysis QC Report

Mutations Report

QVs for each mutation, and the average QV for the bases to the left and right of the deletion are provided in Mutations report.

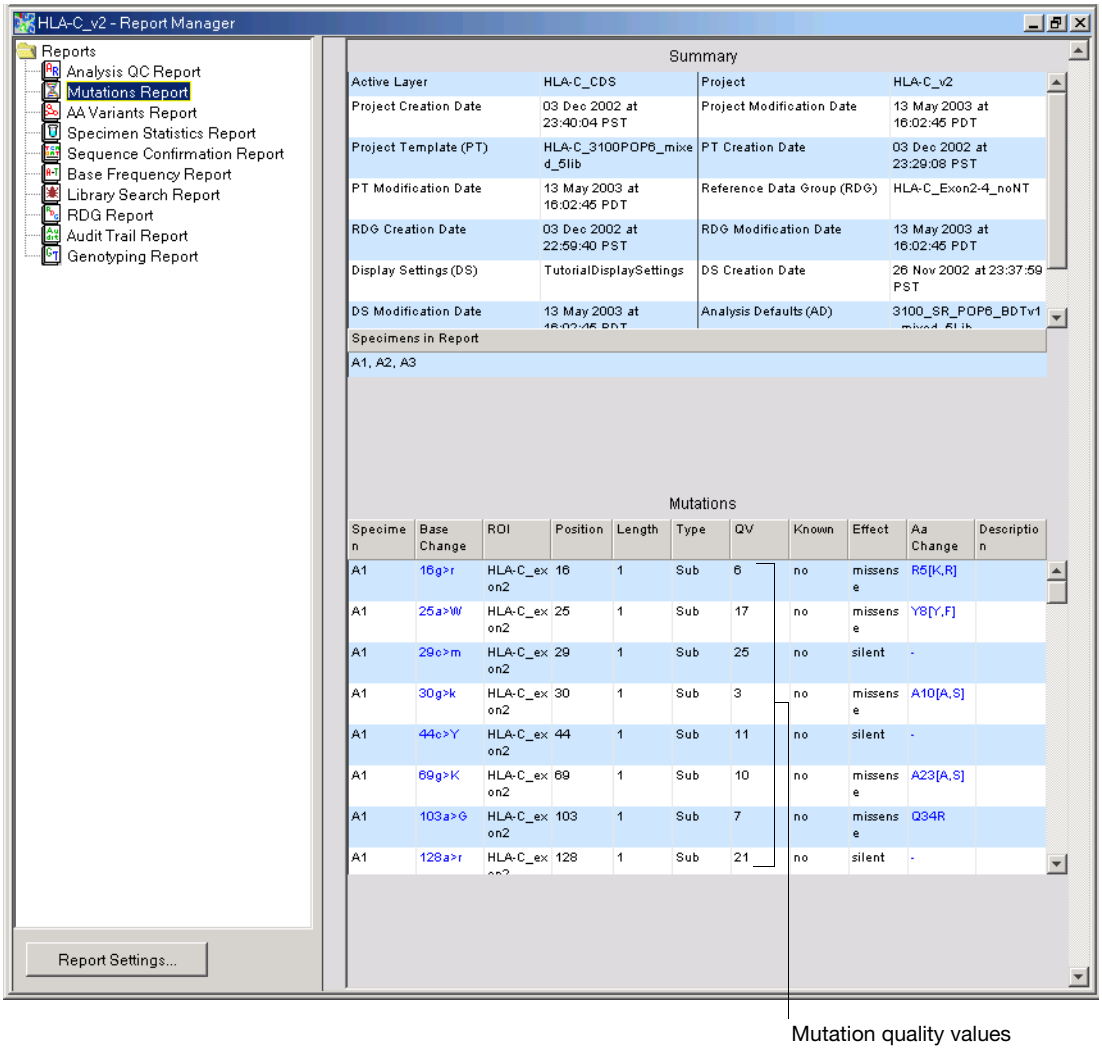


Figure A-6 Mutations Report

Specimen Statistics Report

The Specimen Statistics table of this report displays the average consensus QV score for a segment in the Segment Score column.

The Sample Results table displays the average sample QV for the bases in the clear range in the Sample Score column.

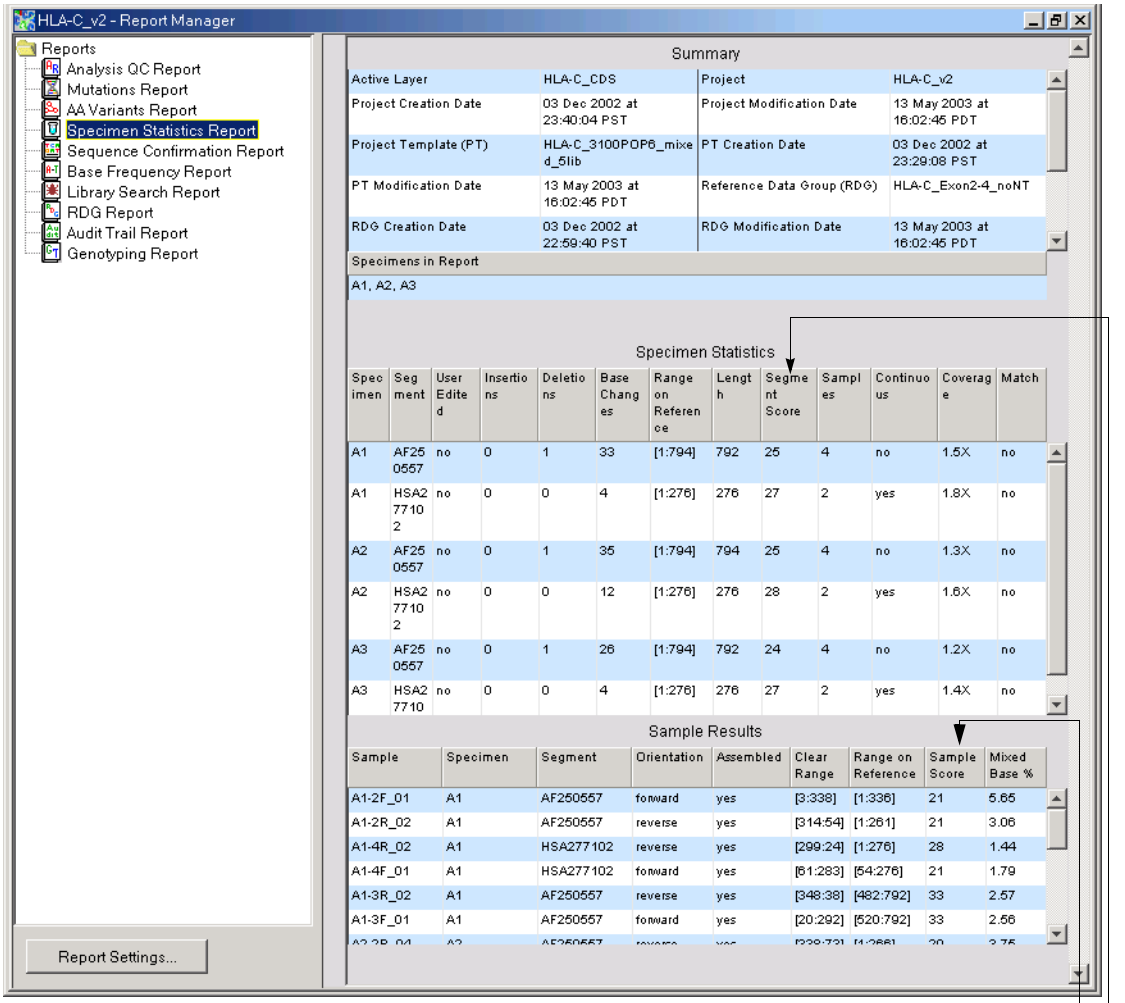


Figure A-7 Specimen Statistics Report

Genotyping Report

The Resequencing Coverage table lists several QVs.

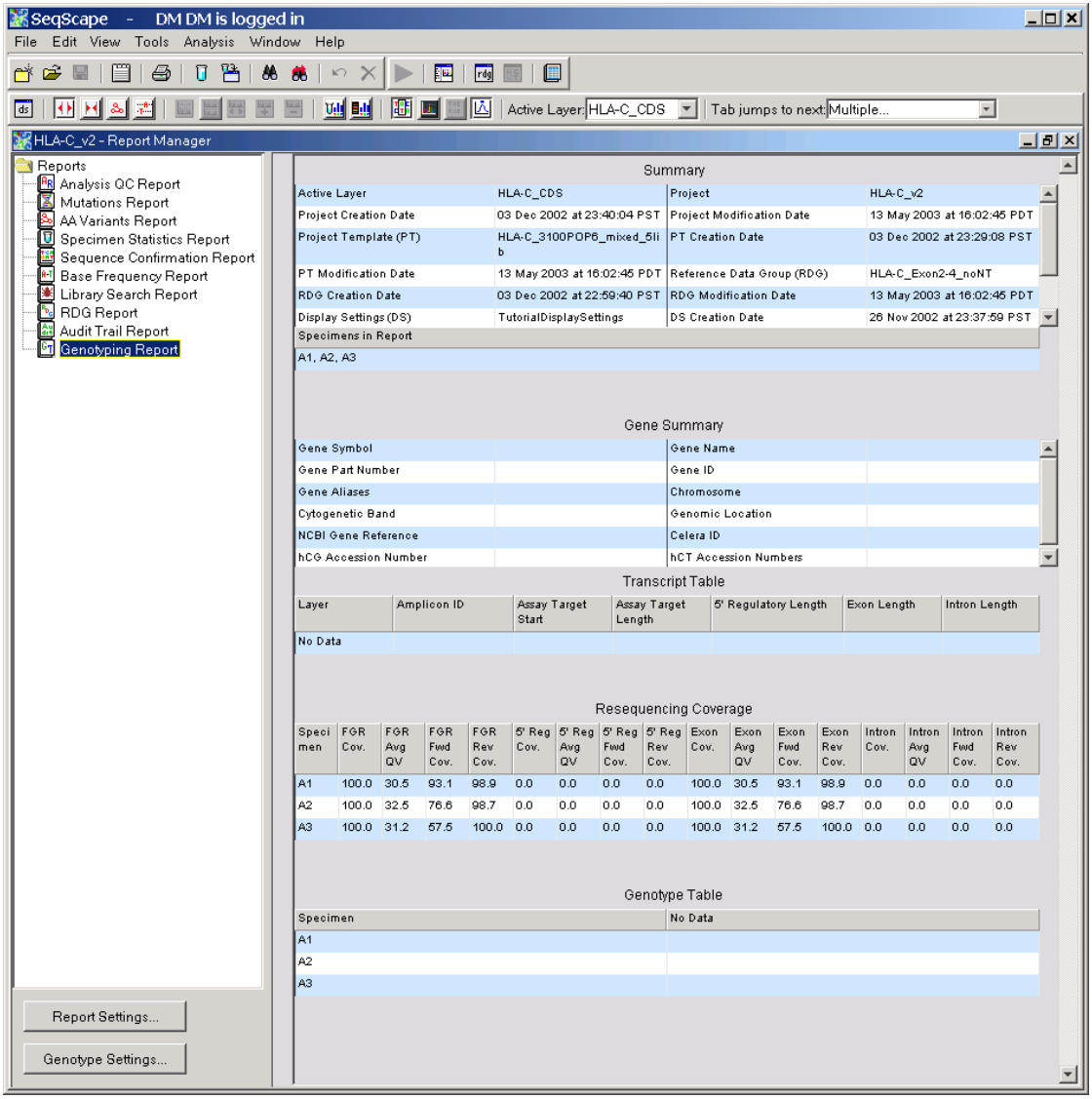


Figure A-8 Genotyping Report

B Basecallers and DyeSet/Primer Files

In This Appendix

Definitions and Naming

Basecaller Algorithm

A basecaller is an algorithm that determines the bases within a sequence during analysis. There are two types of basecallers:

- KB™ Basecaller – An algorithm that calculates mixed or pure bases and sample quality values.
- ABI Basecaller – An algorithm used in earlier versions of Sequencing Analysis and SeqScape® software.

DyeSet/Primer File

The DyeSet/Primer file compensates for the mobility differences between the dyes and primers and corrects the color code changes due to the type of chemistry used to label the DNA. DyeSet/Primer files are sometimes referred to as mobility files.

DyeSet/Primer File-Naming Conventions

DyeSet/Primer files use the following name convention:

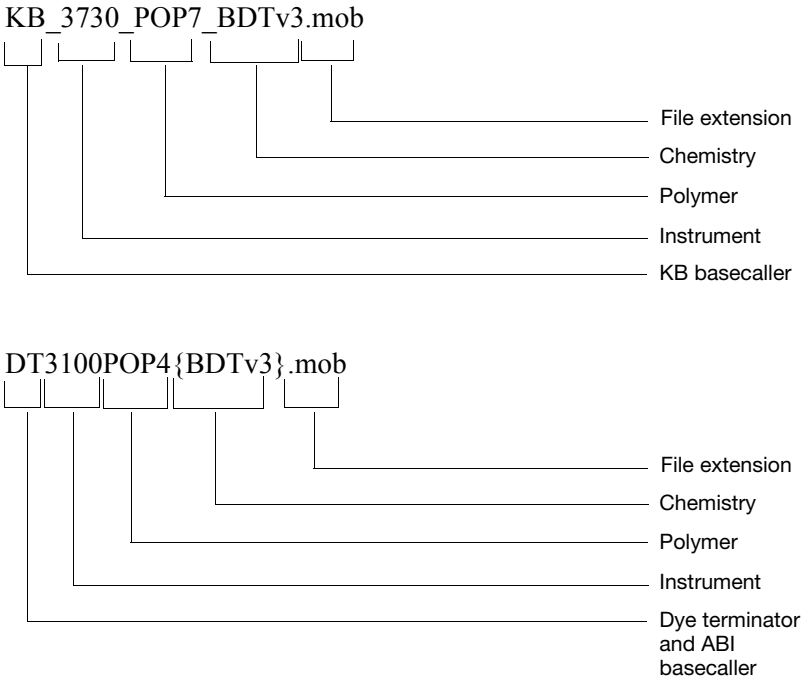


Figure B-1 Examples of DyeSet/Primer file naming convention

The DyeSet/Primer file names use a combination of characters to indicate the basecaller, instrument, chemistry, and polymer type as described below.

Abbreviation	For Runs Using …
Basecaller	
KB	KB™ Basecaller basecaller
DP	Dye primer chemistry and the ABI basecaller
DT	Dye terminator chemistry, and the ABI basecaller
Type of Polymer or Gel	
5%LR	% Long Ranger in the gel (377 instrument only)
POP4	POP-4™ polymer
POP6	POP-6™ polymer
POP7	POP-7™ polymer
Chemistry	
BDTv3	BigDye® v3.0 and 3.1 Terminator
{BDv3}	
BDTv3direct	BigDye® Direct Cycle Sequencing
{BDv1}	BigDye® v1.0 and 1.1 Terminator
{BD}	
{-21M13}	Dye primer chemistry – the -21M13 primer is labeled
{M13Rev}	Dye primer chemistry – the M13Rev primer is labeled

DyeSet/Primer Files Included

In addition to DyeSet/Primer files for 310, 3100, 3130/3130xl, 3500/3500xl, and 3730/3730xl, the following DyeSet/Primer files have been added.

- KB_3130_POP7_BDTv3direct.mob
- KB_3730_POP7_BDTv3direct.mob
- KB_3500_POP7_BDTv3direct.mob
- KB_3500_POP7_BDTv3.mob
- KB_3500_POP7_BDTv1.mob
- KB_3500_POP6_BDTv3.mob

- KB_3500_POP6_BDTv1.mob (improved to support 19.5 kV RapidSeq50 and 16.9 kV FastSeq50 run modules)

Basecaller and DyeSet/Primer Compatibility

The DyeSet/Primer file must match the chemistry and basecaller type that you are using. DyeSet/Primer files are filtered based on the selected basecaller.

IMPORTANT! If you select the DyeSet/Primer file, then select a basecaller file, no filtering of the basecaller list occurs. If you select a KB™ Basecaller DyeSet/Primer file and an ABI basecaller for analysis, or a DT DyeSet/Primer file and an KB™ Basecaller basecaller for analysis, error messages are displayed (see Figures B-2 and B-3).

Figure B-2 Error message in the Sample Manager

Figure B-3 Error message in the analysis protocol

310 Genetic Analyzer Files

Note: 47 cm capillary array length = 36 cm read length
61 cm capillary array length = 50 cm read length

Table B-1 310 Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
KB Basecalling			
BigDye® Terminator v1.0 and v1.1	47	KB.bcp	KB_310_POP4_BDTv1_36Rapid.mob KB_310_POP4_BDTv1_36Std.mob
	47	KB.bcp	KB_310_POP6_BDTv1_36Rapid.mob
	61		KB_310_POP6_BDTv1_50Std.mob
BigDye® Terminator v3.0 and v3.1	47	KB.bcp	KB_310_POP4_BDTv3_36Rapid.mob KB_310_POP4_BDTv3_36Std.mob
	47	KB.bcp	KB_310_POP6_BDTv3_36Rapid.mob
	61		KB_310_POP6_BDTv3_50Std.mob
ABI Basecalling			
BigDye® Terminator v1.0 and v1.1	47	Basecaller-310POP4.bcp	DT310POP4{BD}v2.mob
	47	Basecaller-310POP6.bcp	DT310POP6{BD}.mob
	61		DT310POP6{BD-LR}v3.mob

Table B-1 310 Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry *(continued)*

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
dRhodamine Terminator	47	Basecaller-310POP4.bcp	DT310POP4{dRhod}v1.mob
	47	Basecaller-310POP6.bcp	DT310POP6{dRhod}v2.mob
	61		
BigDye® Terminator v3.0 and v3.1	47	Basecaller-310POP4.bcp	DT310POP4{BDv3}v2.mob
	47	Basecaller-310POP6.bcp	DT310POP6{BDv3}v2.mob
	61		

Table B-2 310 Basecaller and DyeSet/Primer Files Used for Dye Primer Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
ABI Basecaller			
BigDye® Primer v1.0 and v1.1	47	Basecaller-310POP4.bcp	DP310POP4{BD-21M13}v1.mob DP310POP4{M13Rev}v1.mob
	47	Basecaller-310POP6.bcp	DP310POP6{BD-21M13}v1.mob DP310POP6{M13Rev}v1.mob
	61		

Table B-2 310 Basecaller and DyeSet/Primer Files Used for Dye Primer Chemistry *(continued)*

BigDye® Primer v3.0 and v3.1	47	Basecaller-310POP4.bcp	DP310POP4{BDv3-21M13}v1.mob DP310POP4{BDv3-M13Rev}v1.mob
	47	Basecaller-310POP6.bcp	DP310POP6{BDv3-21M13}v1.mob DP310POP6{BDv3-M13Rev}v1.mob
	61		

377 DNA Sequencer Files

Table B-3 377 Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry

DNA Sequencing Chemistry	WTR (cm)/Scan Rate (scans/hr)	Basecaller	DyeSet/Primer
ABI Basecalling			
• BigDye® Terminator v1.0 and v1.1 • dGTP BigDye® Terminator	36/2400	Basecaller-377.bcp	DT377{BD}.mob
	36 & 48/1200	Basecaller-377LR.bcp	
dRhodamine Terminator	36/2400	Basecaller-377.bcp	DT377{dRhod}.mob
	36 & 48/1200	Basecaller-377LR.bcp	
• BigDye® Terminator v3.0 and 3.1 • dGTP BigDye® v3.0 Terminator	36/2400	Basecaller-377.bcp	DT377{BDv3}v2.mob
	36 & 48/1200	Basecaller-377LR.bcp	DT377LR{BDv3}v1.mob

Table B-4 377 Basecaller and DyeSet/Primer Files Used for Dye Primer Chemistry

DNA Sequencing Chemistry	WTR (cm)	Basecaller	DyeSet/Primer
ABI Basecalling			
BigDye® Primer v1.0 and v1.1	36/2400	Basecaller-377.bcp	DP377-5%LR{BD-21M13}.mob DP377-5%LR{BD-M13Rev}.mob,
	36 & 48/1200	Basecaller-377LR.bcp	
BigDye® Primer v3.0 and 3.1	36/2400	Basecaller-377.bcp	DP377{BDv3-21M13}v1.mob DP377{BDv3-M13Rev}v1.mob
	36 & 48/1200	Basecaller-377LR.bcp	

3100 Genetic Analyzer Files

Table B-5 3100 Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
KB Basecalling			
BigDye® Terminator v1.0 and v1.1	36: ultra rapid	KB.bcp	KB_3100_POP4_BDTv1.mob
	50: std read		
	80: long read		
	36: rapid read	KB.bcp	KB_3100_POP6_BDTv1.mob
	50: std read		
BigDye® Terminator v3.0 and v3.1	36: ultra rapid	KB.bcp	KB_3100_POP4_BDTv3_.mob
	50: std read		
	80: long read		
	36: rapid read	KB.bcp	KB_3100_POP6_BDTv3.mob
	50: std read		

Table B-5 3100 Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry *(continued)*

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
ABI Basecalling			
• BigDye® Terminator v1.0 and v1.1 • dGTP BigDye® Terminator	36: ultra rapid	Basecaller-3100POP4UR.bcp	DT3100POP4LR{BD}v1.mob
	80: long read	Basecaller-3100POP4_80cmv3.bcp	
	36: rapid read	Basecaller-3100POP6RRv2.bcp	DT3100POP6{BD}v2.mob
	50: std read	Basecaller-3100POP6SR.bcp	
• BigDye® Terminator v3.0 and 3.1 • dGTP BigDye® v3.0Terminator	36: ultra rapid	Basecaller-3100POP4UR.bcp	DT3100POP4{BDv3}v1.mob
	80: long read	Basecaller-3100POP4_80cmv3.bcp	
	36: rapid read	Basecaller-3100POP6RRv2.bcp	DT3100POP6{BDv3}v1.mob
	50: std read	Basecaller-3100POP6SR.bcp	
dRhodamine Terminator	36: ultra rapid	Basecaller-3100POP4UR.bcp	DT3100POP4{dRhod}v2.mob
	80: long read	Basecaller-3100POP4_80cmv3.bcp	
	36: rapid read	Basecaller-3100POP6RRv2.bcp	DT3100POP6{dRhod}v2.mob
	50: std read	Basecaller-3100POP6SR.bcp	

Table B-6 3100 Basecaller and DyeSet/Primer Files Used for Dye Primer Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
ABI Basecalling			
BigDye® Primer v1.0 and v1.1	36: rapid read	Basecaller-3100POP6RRv2.bcp	DP3100POP6{BD-21M13}v1.mob DP3100POP6{BD-M13Rev}v1.mob
	50: std read	Basecaller-3100POP6SR.bcp	
BigDye® Primer v3.0 and 3.1	36: rapid read	Basecaller-3100POP6RRv2.bcp	DP3100POP6{BDv3-21M13}v1.mob DP3100POP6{BDv3-M13Rev}v1.mob
	50: std read	Basecaller-3100POP6SR.bcp	
BigDye® v3 Primer (All primers)	36: ultra rapid	Basecaller-3100POP4UR.bcp	DP3100POP4{BDv3}v1.mob
	80: long read	Basecaller-3100POP4_80cmv3.bcp	

3100-*Avant* Genetic Analyzer Files

Table B-7 3100-*Avant* Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
KB Basecalling			
BigDye® Terminator v1.0 and v1.1	36: ultra rapid	KB.bcp	KB_3100_POP4_BDTv1.mob
	50: std read		
	80: long read		
	36: rapid read	KB.bcp	KB_3100_POP6_BDTv1.mob
	50: std read		
BigDye® Terminator v3.0 and v3.1	36: ultra rapid	KB.bcp	KB_3100_POP4_BDTv3_.mob
	50: std read		
	80: long read		
	36: rapid read	KB.bcp	KB_3100_POP6_BDTv3.mob
	50: std read		

Table B-7 3100-*Avant* Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry *(continued)*

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
ABI Basecalling			
BigDye® Terminator v1.0 and v1.1	36: ultra rapid	Basecaller-3100APOP4UR.bcp	DT3100POP4LR{BD}v1.mob
	80: long read	Basecaller-3100APOP4_80cmv3.bcp	
	36: rapid read	Basecaller-3100APOP6RRv2.bcp	DT3100POP6{BD}v2.mob
	50: std run	Basecaller-3100APOP6SR.bcp	
BigDye® Terminator v3.0 and v3.1	36: ultra rapid	Basecaller-3100APOP4UR.bcp	DT3100POP4{BDv3}v1.mob
	80: long read	Basecaller-3100APOP4_80cmv3.bcp	
	36: rapid read	Basecaller-3100APOP6RRv2.bcp	DT3100POP6{BDv3}v1.mob
dRhodamine Terminator	36: ultra rapid	Basecaller-3100APOP4UR.bcp	DT3100POP4{dRhod}v2.mob
	80: long read	Basecaller-3100APOP4_80cmv3.bcp	
	36: rapid read	Basecaller-3100APOP6RRv2.bcp	DT3100POP6{dRhod}v2.mob
	50: std run	Basecaller-3100APOP6SR.bcp	

Applied Biosystems® 3130/3130*xl* Genetic Analyzer Files

Table B-8 3130 Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
KB Basecalling			
BigDye® Terminator v1.0 and v1.1	36: ultra	KB.bcp	KB_3130_POP4_BDTv1.mob
	50: std read		
	80: long		
	36: rapid read	KB.bcp	KB_3130_POP6_BDTv1.mob
	50: std read		
	36: ultra	KB.bcp	KB_3130_POP7_BDTv1.mob
	50: fast std		
	80: long		

Table B-8 3130 Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry *(continued)*

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
BigDye® Terminator v3.0 and v3.1	36: ultra	KB.bcp	KB_3130_POP4_BDTv3_.mob
	50: std read		
	80: long read		
	36: rapid read	KB.bcp	KB_3130_POP6_BDTv3.mob
	50: std read		
	36: ultra rapid	KB.bcp	KB_3130_POP7_BDTv3.mob
	50: fast std		
	80: long		
BigDye® Direct	36: ultra	KB.bcp	KB_3130_POP7_BDTv3direct.mob
	50: fast std		

3700 DNA Analyzer Files

Table B-9 3700 Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
ABI Basecalling			
BigDye® Terminator v1.0 and 1.1	50	Basecaller-3700POP6.bcp	DT3700POP6{BD}v5.mob
		Basecaller-3700POP5LR.bcp	DT3700POP5{BD}v3.mob
BigDye® Terminator v3.0 and 3.1	50	Basecaller-3700POP6.bcp	DT3700POP6{BDv3}v1.mob
		Basecaller-3700POP5LR.bcp	DT3700POP5{BDv3}v1.mob
dRhodamine Terminator	50	Basecaller-3700POP6.bcp	DT3700POP6{dRhod}v3.mob
		Basecaller-3700POP5LR.bcp	DT3700POP5{dRhod}v1.mob

Table B-10 3700 Basecaller and DyeSet/Primer Files Used for Dye Primer Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
ABI Basecalling			
BigDye® Primer v1.0 and v1.1	50	Basecaller-3700POP6.bcp	DP3700POP6{BD-21M13}v3.mob DP3700POP6{BD-M13Rev}v2.mob
		Basecaller-3700POP5LR.bcp	DP3700POP5{BD-21M13}v1.mob DP3700POP5{BD-M13Rev}v1.mob
BigDye® Primer v3.0 and v3.1	50	Basecaller-3700POP6.bcp	DP3700POP6{BDv3-21M13}v1.mob DP3700POP6{BDv3-M13Rev}v1.mob
		Basecaller-3700POP5LR.bcp	DP3700POP5{BDv3-21M13}v1.mob DP3700POP5{BDv3-M13Rev}v1.mob

Applied Biosystems® 3730/3730*xl* DNA Analyzers Files

Table B-11 3730/3730*xl* Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry

DNA Sequencing Chemistry	Capillary Array Length (cm)	Basecaller	DyeSet/Primer
KB Basecalling			
BigDye® v3.0 and v3.1 Terminator	all lengths	KB.bcp	KB_3730_POP7_BDTv3.mob
			KB_3730_POP6_BDTv3.mob
BigDye® v1.0 and v1.1 Terminator	all lengths		KB_3730_POP7_BDTv1.mob
			KB_3730_POP6_BDTv1.mob
BigDye® Direct	all lengths		KB_3730_POP7_BDTv3direct.mob
ABI Basecalling			
BigDye® Terminator v1.0 and v1.1	36: rapid read	Basecaller-3730POP7RR.bcp	DT3730POP7{BD}.mob
	36: std read	Basecaller-3730POP7SR.bcp	
	50: long read	Basecaller-3730POP7LR.bcp	
BigDye® Terminatorv3.0 and v3.1	36: rapid read	Basecaller-3730POP7RR.bcp	DT3730POP7{BDv3}.mob
	36: std read	Basecaller-3730POP7SR.bcp	
	50: long read	Basecaller-3730POP7LR.bcp	

Applied Biosystems® 3500/3500*xl* Genetic Analyzers Files

Table B-12 3500/3500*xl* Basecaller and DyeSet/Primer Files Used for Dye Terminator Chemistry

DNA Sequencing Chemistry	Capillary Array Length (50 cm)	Basecaller	DyeSet/Primer
KB Basecalling			
BigDye® Terminator v1.0 and v1.1	ShortReadSeq RapidSeq FastSeq StdSeq	KB.bcp	KB_3500_POP7_BDTv1.mob
	RapidSeq FastSeq StdSeq	KB.bcp	KB_3500_POP6_BDTv1.mob
BigDye® Terminator v3.0 and v3.1	ShortReadSeq RapidSeq FastSeq StdSeq	KB.bcp	KB_3500_POP7_BDTv1.mob
	RapidSeq FastSeq StdSeq	KB.bcp	KB_3500_POP6_BDTv1.mob
BigDye® Direct	ShortReadSeq RapidSeq FastSeq StdSeq	KB.bcp	KB_3500_POP7_BDTv3direct.mob

C

Frequently Asked Questions

In This Appendix

Upgrading FAQs

There are two versions of SeqScape® Software v3 available. See "SeqScape® Software Versions" on page 2-4.

Table C-1 Upgrading questions and answers

Question	Answer
How does SeqScape® Software 3 differ from earlier versions?	Refer to Chapter 1, "Introduction to SeqScape® Software."
What happens to my data when I upgrade my SeqScape® software?	See "Upgrading from SeqScape® Software v2.x" on page 2-7.
What happens to the data that I created with the demo version when I upgrade to the full version of SeqScape® software?	You need to export the data generated with the 30-day demo before it expires, install SeqScape® software v2.5, then import the data into the full version of the SeqScape® software v2.5. To export, select the desired object from the SeqScape® Manager window, then click **Export**. To import the object into the full copy of the SeqScape® software v2.5, use the SeqScape® Manager.

Training and Documentation FAQs

Table C-2 Training and documentation questions and answers

Question	Answer
How do I train myself on SeqScape® software?	The best way to train yourself on SeqScape® software is to use the training movie and the software tutorial included in the software package. A printed tutorial is included with all SeqScape® software versions except for the demo version. The tutorial is also available online (see below). Additionally, workflows are available for v2.5.
Where can I get find resources/documentation on SeqScape® software?	All documentation for SeqScape® is available at **Start** > **Programs** > **Applied Biosystems** > **SeqScape**. You can also find documentation about SeqScape® software at: lifetechnologies.com.

SeqScape® Software Basics FAQs

Table C-3 SeqScape® Software Basics Questions and Answers

Question	Answer
What is SeqScape® software?	SeqScape® software is a resequencing software tool designed to identify nucleotide variants, amino acid variants, and sequences in a library that match each consensus sequence.
SeqScape® software can be used for which sequencing application?	SeqScape® software can be used for SNP discovery and validation, mutation analysis and heterozygote identification, sequence confirmation for mutagenesis or clone-construct confirmation studies, and the identification of genotype, allele and haplotype from a library of known sequences.

Table C-3 SeqScape® Software Basics Questions and Answers *(continued)*

Question	Answer
What is the overall workflow for analyzing and reviewing data?	All analysis in SeqScape® software occurs in a project. Analysis and review of the data requires that you: 1. Create an analysis defaults and display settings (or use one previously created). 2. Create an RDG (or use one previously created). 3. Create a project template (or use one previously created). 4. Create and analyze a project by adding sample files to a project template. 5. Review the results (view/edit the data and view the reports). 6. Export/print the results and reports.
What happens in the SeqScape® software when I click  (Analyze)?	Each time you click the Analyze button, the software performs the following on specimens that have not been analyzed: 1. Basecalls and determines quality values 2. Identifies mixed bases. (This step can be bypassed if data was previously basecalled.) 3. Trims low-quality sequence ends 4. Filters (omit poor-quality sequences) 5. Assembles sequences to the reference 6. Generates a consensus sequence for each specimen 7. Aligns consensus sequence to the reference sequence 8. Compares each consensus to the reference 9. Searches the allele Library for matches to each consensus sequence. (This step can be bypassed if you do not need to identify allele matches.) 10. Generates reports

General SeqScape® Software FAQs

Table C-4 General Questions and Answers

Question	Answer
Instruments – What Life Technologies instruments can I use to generate data for SeqScape® software?	SeqScape® software analyzes sequence files generated from the 310, 377, 3100, 3100-*Avant*, 3700, and Applied Biosystems® 3130/3130*xl*, 3730/3730*xl,* and 3500/3500*xl* instruments. The software also accepts text sequences in FASTA format.
Instruments – Can SeqScape® software be used to analyze data that was generated on instruments other than Life Technologies instruments?	No. Sequencing data generated on platforms other than Life Technologies platforms are not compatible with SeqScape® software.
Sample files – What instrument sample files can I use with the KB™ Basecaller basecaller?	You can use the KB™ Basecaller basecaller in SeqScape® software to analyze sequencing sample files generated from the 310, 3100/3100-*Avant* Genetic Analyzers and Applied Biosystems® 3130/3130*xl*, 3730/3730*xl,* and 3500/3500*xl* instruments.
How can I share my work with someone at a different site? What should I send them?	All sample files, analysis parameters, reference sequence, and analysis results are saved in every SeqScape® project file. You can share these files with anyone who has the software by exporting the project or data objects, then importing them into the software on the computer of the other person. There is no link between the SeqScape® software installed on different computers. You can also share project templates, which contain the reference sequence and analysis parameters. Colleagues can then analyze sample files of his or her choice using the project templates to create a new project. The analysis is identical to your own analysis with the same project template.

Table C-4 General Questions and Answers *(continued)*

Question	Answer
Files – What are the file types/formats accepted by SeqScape® software?	**Reference Sequences** — SeqScape® software accepts the following file formats for reference sequences: • Genbank File Format (with a .fcgi, .cgi, or .gb extension) • .txt (text) file format • .ab1 file format • .fsta (FASTA) file format • .seq file format • Aligned sequences in .fsta (FASTA) format* *The imported sequence shows a summary of all the sequences in the file by substituting the IUPAC codes for bases where there is a discrepancy in the sequences. **Data Sequences**— SeqScape® software accepts the following file formats for data sequences used for analysis: • .ab1 files (previously basecalled or not basecalled) • .txt (TEXT) files • .fsta (FASTA) files • .seq files **Nucleotide Variants** – SeqScape® software accepts the following file formats for nucleotide variants: • .fsta file containing a set of aligned sequences in FASTA format. • Tab-delimited text (.txt) file that lists one variant per line and eight column headings: Type, ROI, NT position, Reference, Variant, Style, Description, and Used by all ROIs. **Amino Acid Variants** – SeqScape® software accepts the following file formats for amino acid variants: • Tab-delimited text (.txt) file that lists one variant per line and the following seven column headings: Type, Layer, AA position, Reference, Variant, Style, and Description
Files – Can .scf files be analyzed in SeqScape® software?	No. SeqScape® software analyzes only sequencing data in .ab1 files or text sequences.

Table C-4 General Questions and Answers *(continued)*

Question	Answer
Files – How can I use sample files generated on the Macintosh® computer with SeqScape® software?	To use data files generated on a Macintosh computer with SeqScape® software, you must convert the files using the SCU (Sample Conversion Utility). This utility is available as a Macintosh application file on the SeqScape® Analysis software CD. The SCU must be loaded onto and launched from a Macintosh computer. For more information, see the Read Me file associated with the SCU.
Chemistry – What Life Technologies chemistries are supported?	• BigDye® Terminator v3.1 and v1.1 Cycle Sequencing Kit • BigDye® Primers and dRhodamine dyes • Applied Biosystems® BigDye® Direct Cycle Sequencing Chemistry
Computer – What are the computer requirements for SeqScape® software?	• CPU – 2 GHz or faster Intel processor • Memory – 2 GB • OS – Windows® 7 Professional, (SP1), 32-bit • Hard drive – 1 GB free space • Monitor – 1280 x 1024 pixel resolution for full screen display. Use Windows 7 default theme.
Computer – What can I do if SeqScape® is running slowly?	If SeqScape® software is running slowly, you can improve performance by archiving older projects. To archive projects, select **Tools** > **SeqScape Manager** > **Project** > **Export**.
Software – Does SeqScape® software support BioLIMS/Sequence Collector software?	No. SeqScape® software no longer supports BioLIMS/Sequence Collector software.
Software – How does SeqScape® software compare to MicroSeq® and ViroSeq® software?	SeqScape® software – Compares samples to a reference sequence MicroSeq® software – Identifies bacteria ViroSeq® software – Identifies genotype HIV-1 resistance mutations
Software – Do I need Sequencing Analysis software if I have SeqScape® software?	Sequencing Analysis software is a multi-purpose software used to analyze, edit, view, display, and print sequencing sample files. Sequencing Analysis software should be used in every laboratory for general troubleshooting and viewing of data. SeqScape® software is designed specifically for resequencing.

Table C-4 General Questions and Answers *(continued)*

Question	Answer
Data objects – Can I transfer data objects like the RDG, Display Settings, Analysis Protocols, etc. from one computer to another?	Yes. You can transfer data objects from one computer to another. To export the data object, to the SeqScape® Manager, select the object, then click **Export**. Send the exported file to the second computer, then launch SeqScape® software. Open SeqScape® Manager, then click **Import** to import the file.
Can I BLAST against a database?	Yes. To search a database using a sequence generated with SeqScape® software, in the Project view, export the NT alignment as an aligned FASTA file by selecting **File** > **Export**. Open the file in a text viewer, then cut and paste the sequence you want to search for in your BLAST query. Refer to Chapter 9, "Electronic Signing, Exporting, and Printing Data and Reports," for details on exporting.
Alignment – What alignment algorithms are used in SeqScape® software?	The sample assembly and specimen alignments are generated using a Smith-Waterman local sequence alignment algorithm using parameters appropriate for DNA sequencing.
Alignment – Can SeqScape® software perform just the alignment for samples?	Yes. To assemble and analyze sequences without basecalling, open the Analysis Defaults for the project, then select the **Specimen** tab, deselect **Basecall Samples**.

SeqScape® Manager FAQs

Table C-5 SeqScape® Manager Questions and Answers

Question	Answer
What is the SeqScape® Manager?	SeqScape® Manager allows you to import, export, create, and delete projects, project templates, reference data groups, analysis defaults, libraries, analysis protocols, and display settings. To access SeqScape® Manager, select **Tools** > **SeqScape Manager**.
What is an object?	An object is a named collection of data elements used to perform certain functions, for example, analysis protocol.
How do I create a new user?	You must log in as an Admin user, then: 1. Select **Tools** > **Options**. 2. Select the **Users** tab, then click **New**. 3. Enter the new user name (be sure to omit any spaces in the user name), then click **OK**. 4. To log in with the new name, exit the software, then relaunch it. 5. Log in with the new user name.
What is a project in SeqScape® software?	Projects contain sample data files grouped into specimens. A project is created using a project template.
What is a project template?	A project template is the mold from which projects are created. A project template contains analysis defaults, display settings, and a reference data group.
What is a specimen?	A specimen contains all the sample data from a single biological source. SeqScape® software assembles all sample data within a specimen and generates a consensus sequence. For example, a specimen contains forward-strand PCR products for exons 3, 4 and 5 of a gene and several reverse-strand PCR products for the same exons. The software generates a single consensus sequence representing exons 3,4 and 5 and compares it to the reference sequence. Do not mix products from different biological sources into a single specimen.

Table C-5 SeqScape® Manager Questions and Answers *(continued)*

Question	Answer
What is a layer?	A layer is a set of ROIs that are grouped together for the purpose of display, report or amino acid translation. The ROIs within a layer cannot overlap. Example: Your project may contain introns 1, 2, 3, 4, 5, 6 and exons 1, 2, 3, 4, 5. You can create a layer that contains only exons 1,4,5 or a layer containing intron1, exon1,intron2, exon2, intron3, exon3, and intron4. A layer can represent a transcript.
How do I generate a new layer?	To generate a new layer, click **New Layer** in the ROI tab of the RDG, then add the desired ROIs by selecting the **On Layer** check box in the ROI table. Also, select whether or not you want the ROI translation turned on in the layer.
Can I put samples from different individuals in the same specimen?	No. Each individual sample should be in a different specimen.
Can I mix samples from different biological sources?	No. You cannot analyze data from different biological sources in the same specimen.
What is a reference data group (RDG)?	The RDG is an essential part of the project template that contains all the analysis-specific information, including the reference sequence, translation codon table, known variants, RDG name, reference segments, regions of interest (ROI), layers, and the name of the associated allele libraries.
What are the new features of the extended RDG?	When using a Genbank file to create the reference sequence, the feature table of the Genbank file is pulled into the RDG, and each feature is listed out in the ROI (region of interest) table. The ROIs can be used to create new layers for sequence comparison. For example, if a Genbank file for a gene containing two exons and one intron is imported into the RDG, you can create a layer that includes only the two exons. When analysis occurs, the specimens are compared to the layer containing the two exons as well as to the reference backbone layer that includes the two exons and the intron. In addition, you can turn translation on or off for specific ROIs. A library containing aligned sequences can also be attached to a specific layer for comparison during analysis.
What is a reference sequence?	A reference sequence is the backbone sequence against which the software compares the consensus segments. A reference sequence contains continuous or discontinuous sequences made up of one or more reference segments

Table C-5 SeqScape® Manager Questions and Answers *(continued)*

Question	Answer
What is a reference segment?	A reference segment is a contiguous section of the reference backbone within the reference sequence that corresponds to a single contiguous DNA sequence.
What is the reference backbone?	The reference backbone is the entire reference sequence that can consist of one or more reference segments. The backbone is the first layer of the RDG, which cannot be modified.
What does splitting the reference mean and how do I split it?	Creating a split can represent that the sequences are not contiguous; One side of the split may contain Exon3 and the other side may represent Exon8. In the ROI tab, click the base position where you want to split the reference segment, then select **Split Reference Segment**.
Where can I find information on the ROI tab?	You can find information by clicking **Info** on the bottom left of the ROI tab within the RDG.
Can the ROI contain negative numbering?	Yes. The ROI can contain negative numbering. You can assign a negative number to an ROI by entering the number into the ROI table of the RDG.
Can the reference sequence contain negative numbering?	No. The reference backbone sequence cannot contain negative numbering. However, individual ROIs within the reference backbone can contain negative numbers.
How do I save GenBank files?	After finding the desired sequence at the NCBI website, select the check box to the left of the accession number. At the top of the page next to Display, select **GenBank**, then select **Send to File**. The file is saved to the specified location and can then be imported into the RDG. The file can have a .gb, .fcgi, or .cgi extension.
How can I designate part of my sequence as untranslated (intronic region)?	You can designate part of the sequence as untranslated. First, select the desired section of the sequence in the ROI tab, then click **Add ROI**. The region appears in the ROI table. Select the layer where you want the ROI to appear, then deselect the **Translate** check box.
How can I change the number of the first base in the reference sequence? How can I reset the first codon?	You can designate the first base/codon in the reference segment pane of the ROI tab. This is the pane that shows selected reference sequence. Click the box on the top left of the pane, then enter the desired number.
How can I change the translation frame?	You can change the translation frame in the ROI tab of the RDG.

Table C-5 SeqScape® Manager Questions and Answers *(continued)*

Question	Answer
Can I use an implicit reference sequence in SeqScape® software?	No. SeqScape® software does not support the use of an implicit reference sequence. However, you can use .ab1 files and genbank files as reference sequences.
What is a reference break?	A reference break is a break in the reference sequence between two reference segments where the reference is not contiguous.
What is a translation codon table?	A table that translates amino acid and genetic codes. Refer to Appendix D, "Translation Tables."
What is a known variant?	An AA variant or NT variant that has been previously identified in the reference.
What is a region of interest (ROI)?	An ROI is a region on the reference segment with special numbering properties used for display. The numbering for the ROI is continuous, always increases from left to right, and does not have to correspond to the numbering on the reference segment. The ROI can also contain negative numbers. ROIs can be grouped into layers for display or translation purposes
How can I configure a reference segment and the ROIs within it?	After you import a reference sequence into the RDG, use the ROI tab to reconfigure a reference segment and to add ROIs.
What if I do not have variant information?	Variants are not necessary to create a reference data group. If you do import variants, they must be in a tab-delimited text file format or FASTA alignment of sequences.
File import – What kinds of files can I import into SeqScape® software?	ABI sample files, tab-delimited text, and FASTA file format can be imported into the software.
File import – Can I import multiple individual text files into one specimen rather than .ab1 files?	Yes. You can import .seq format files or FASTA format files as sample files. To see these files in the import dialog box, you need to deselect **Show .ab1 Samples File Only**. The files can then be analyzed like normal files (except no basecalling occurs).
File import – Do my sample files need to be imported in the same reading frame as the reference sequence?	No. The imported sample files do not need to be in the same reading frame as the reference sequence.

Table C-5 SeqScape® Manager Questions and Answers *(continued)*

Question	Answer
File format – Can analyzed data be used in SeqScape® software?	Yes. Analyzed data can be used. However, if the data are in the ABI data format (not FASTA), any prior analysis, results, and edits are overwritten when the files are reanalyzed using SeqScape® software.
File format – What is FASTA format? How can I convert non-FASTA files into the correct format?	A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (>) symbol in the first column. **Note:** When creating a file in Microsoft® Word, be sure to save it in text-only format (line breaks are OK, but spaces are not OK). >HumMitoCamb from 15871 to 450 (hard return) aatactcaaatgggcctgtccttgtagtataaactaatacaccagtcttgtaaaccggagatg aaaaccttttccaaggacaaatcagagaaaaagtctttaactccaccattagcacccaaag ct (hard return)
What are Analysis Settings?	The analysis settings specify the basecalling, mixed base settings, clear range, and filter settings.
What is Clear Range?	Clear range specifies the range of usable sample sequence to be included in the consensus.
Can the Clear Range be modified within a project? Does changing the Clear Range require that the data be re-analyzed?	You can change the Clear Range for the entire project by applying a new Project Template (with a modified Analysis Protocol), in which case all samples must be re-analyzed and any sample basecall/edits are lost. You can reset the Clear Range for an individual sample by modifying the sample's Analysis Protocol setting. In this case, the specimen containing affected sample(s) must be re-analyzed. If only the Clear Range tab was modified in the Analysis Protocol, the analysis pipeline is started from the Clear Range determination onward, so basecalls are not overwritten. You can also change the sample Clear Range within the sample file. This will not require re-analysis. To change the clear range in this way, right-click a sample sequence and select **Set CR... at selection**, then click-drag the CR bracket to reset the Clear Range. You can also select **Tools** > **Set Clear Range** to reset the sample clear range.
What are Filter Settings?	Filter settings specify the maximum percentage of mixed-bases allowed, maximum Ns allowed, minimum clear range length, and the minimum sample score for each sample. Samples failing the filter checks are not included in the analysis.

Table C-5 SeqScape® Manager Questions and Answers *(continued)*

Question	Answer
What are Display Settings?	Display settings control the font styles and colors for bases, electropherogram display and axis scale, display view for variants, and display views for nucleotide translation.
Can I export consensus sequences?	Yes. Select the desired consensus sequence in the Project Navigator, then select **File** > **Export**.

Library FAQs

Table C-6 Library Questions and Answers

Question	Answer
What are the requirements of the library?	The library search feature is designed for use with libraries of alleles, genotypes, or haplotypes where all the sequences in the library are variations of the same sequence. This feature is not designed for searching against a library of diverse sequences. The library must have the following characteristics: • All library sequences must be pre-aligned and equal in length. • All library sequences must be variations of the same sequence (variations must be less than 50%). • All library sequences must cover the same regions as the layer that is associated with that library (for example, for a layer that contains exons 2,3,7 and 8, a valid library should have sequences from exons 2, 3, 7 and 8. A library with sequences covering exons 2, 3, 4,5, 7 and 8 would be invalid). A library is associated with its specific layer.
Is there a minimum/maximum number of Library Search match returns that I can define?	No. There is no maximum. However, it does not make sense to request more than the number of alleles in the library. The minimum should be 1. You can set the number of library matches to return in the Analysis Defaults > Specimen tab of the project.
What is the difference between a haploid and diploid library?	In a haploid library, all the sequences are pure base sequences. A diploid library contains both mixed base and pure base sequences. A haploid library returns two possible matches, while a diploid library returns one possible match.
Can I add a library to an open project from within the RDG properties button and see the search results instantly, or must I re-analyze the project for the library search to be initiated?	The library is automatically searched immediately after loading a new library for the active layer and after closing the RDG dialog box.

Table C-6 Library Questions and Answers *(continued)*

Question	Answer
What is the function of the Library Identification pane? How can I view the Library Identification pane?	You can use the Library Identification pane to display the crucial positions identified among the set of library matches returned against the selected specimen consensus sequence. To view this pane, click a base in a specimen consensus sequence in the Project Assembly view. You can adjust the height of the pane by using the split bar. The crucial position columns are hyperlinked to the specimen consensus sequence base positions that are highlighted by the column selector in the Project Assembly view.
What is a crucial position error?	A crucial position error is a polymorphic position that occurs in all the allelic matches. It is the position that makes each allele unique to one another.
What is a constant position error?	A constant position error is a position in a specimen consensus sequence that is different from the corresponding position in all the allelic matches. All the allelic matches have the same base for that particular position.

Mutation, Variant, HIM, and HFM Detection FAQs

Table C-7 Mutation, Variant, HIM, and HFM Detection Questions and Answers

Question	Answer
Does SeqScape® software account for heterozygous indel mutations (HIMs)? For example, a sequence with an insertion of three bases.	Yes. SeqScape® software shows possible HIM location and identity in the Analysis QC Report.
Can the SeqScape® software separate the HIM sequence traces?	No. SeqScape® software cannot separate HIM sequence traces. However, SeqScape® software reports how many bases were deleted or inserted.

Data Analysis FAQs

Table C-8 Data Analysis Questions and Answers

Question	Answer
How do I begin analysis?	Click  (Analyze) in the toolbar or select **Analysis** > **Analyze**.
What does a red line across a specimen indicate?	The strike through symbols indicate that analysis needs to be performed.
Can the SeqScape® software handle gaps in sequence?	SeqScape® software automatically inserts gaps in the sample and consensus sequences if the gaps are necessary to produce clean sequence alignments. Gaps should be removed before importing sequences from FASTA-formatted files.
What does the Alignment Score mean in the Analysis Report?	The alignment score shows the number of characters that were inserted in each specimen consensus to create the project alignment. A lower alignment score indicates more similarity between the specimen consensus and the reference.
How does editing affect my data? What gets updated?	If you insert, delete, or change a base within a sample, the change is reflected in the consensus sequence. All samples change to reflect the consensus edits.
How can I distinguish between edited and non-edited data?	Edited bases are displayed as lowercase letters while unedited bases are displayed in uppercase letters.
What happens to my edited sequence when I start analysis?	After basecalling starts, all current edits are overwritten. Changes to the analysis settings that do not require re-basecalling of the sample preserve edits and the reference sequence.
What happens if I edit a consensus base?	The base changes to lowercase in the consensus, and the quality bar turns gray. All bases in the samples at that position that disagree with the new basecall are changed to agree with the new consensus base and are shown in lowercase with a gray quality bar.
How do I remove unwanted spaces in my samples?	To remove unwanted spaces in the sample, double-click the space, press the **Delete** key.
What can I do if I deleted too many bases?	Repeat the analysis.
Is there an option to basecall without generating quality values?	No. All the basecallers in SeqScape® software generate quality values. However, you do not have to display the quality values. You can hide the quality values by deselecting the confidence bar icons in the Views tab of the project's Display Settings.

Table C-8 Data Analysis Questions and Answers *(continued)*

Question	Answer
Can I assemble/analyze my samples without re-basecalling my samples so that I can conserve the existing basecalls?	Yes. To assemble and analyze your sequences without basecalling, open the Analysis Defaults for the project, select the **Specimen** tab, then deselect **Basecall Samples**.
How do I analyze samples in one project with different basecallers/dye set primer files?	To analyze samples in one project with different basecallers and dye set/primer files, select **Analysis** > **Sample Manager**, select the appropriate basecaller and dye set/primer files, then click **Apply**. The project must be re-analyzed for the changes to take effect. Samples and specimens with a red slash indicate an unanalyzed status. in the sample manager, you can also edit the analysis protocol for the individual samples or apply an analysis protocol.
What does a red line through the specimen icon indicate?	A red line indicates that analysis has not occurred. A red line may also appear if the analysis settings have been changed and the project requires re-analysis to apply the settings. Click the green arrow (run) button at the top of the window to start analysis.
How can I edit my specimen name?	Select the specimen, then select **Edit** > **Rename** or right-click the selected specimen, then select **Edit**.
How can I delete samples or specimens?	Select the item to be deleted, then do one of the following: Select **Edit** > **Delete**, click the **Delete** button on the toolbar, press the **Delete** key on the keyboard, or right-click the selected item, then click **Delete**.
What is the TraceTuner™ basecaller module?	The ABI basecaller contains an algorithm that assigns bases and invokes a second algorithm, the TraceTuner™ module. The TraceTuner™ module generates per-base sample quality values and identifies mixed bases.
What does a red dot mean in the analyzed project?	A red dot indicates a base that has been called by the consensus caller. The consensus caller edits this base in the relevant sample sequences of the specimen. The edited base appears in lowercase, and has a gray quality-value bar.
Can I change the settings of the tab jump key?	Yes, you can change the settings of the tab jump key in the Views tab of the project Display Settings. You can also change the tab jump key settings when the project is open by selecting “Tab jumps to next...” on the toolbar.

Analysis Reports FAQs

Table C-9 Analysis Reports Questions and Answers

Question	Answer
How can I access my reports?	Access all reports by clicking the **Report Manager** button in the toolbar or by selecting **Analysis** > **Report Manager**.
What are the different reports available in SeqScape® software?	SeqScape® software v2.5 can generate the following reports: • Analysis QC Report • Mutations Report • AA Variants Report • Specimen Statistics Report • Sequence Confirmation Report • Base Frequency Report • Library Search Report • RDG Report • Audit Trail Report • Electronic Signature History Report • Genotyping Report **Note:** For more information on the reports, see "Viewing the Reports" on page 7-24.
What is the Analysis QC Report?	The Analysis QC report provides a summary of the project's history. This report indicates the status of each specimen at each step of analysis. In addition, the Analysis QC report lists possible HIMs (heterozygous insertion/deletion mutations).
What does "Segment Score" mean in the Specimen Statistics report?	Segment Score gives an average of all the quality values within the clear range in that particular reference segment region.
What does "Coverage" mean in the Specimen Statistics and Sequence Confirmation reports?	Coverage gives a value for the number of samples in the consensus sequence.

Table C-9 Analysis Reports Questions and Answers *(continued)*

Question	Answer
Can I edit sequences within a project while reports are open and see the updated information in the reports instantly, or must I close and re-open the reports to see any changes?	Yes. The reports stay open, and the results are updated as edits are made.
Why are my sample files unassembled?	If you have samples in the unassembled node of a specimen, check the Analysis QC report to determine why the sample files were not assembled. The analysis QC report shows whether or not the sample assembled, as well as the reason for failure at a particular point in the analysis pipeline.

Quality Values FAQs

Table C-10 QV Questions and Answers

Question	Answer
What are quality values?	A quality value is an estimation of the certainty for a basecall in the sample (sample QV) or consensus (consensus QV).
Is there an option to basecall without generating quality values?	No. All of the basecallers in SeqScape® software generate quality values. However, you can choose to not display the quality values by deselecting the confidence bar icons in the Views tab of the project's Display Settings.
How is the basecaller quality value generated?	The basecaller quality value is generated by an algorithm that is designed to examine the certainty of basecalls. See, Appendix A, "Sample and Consensus Quality Values," for more information.
What is the quality value equation?	$QV = -10\log_{10}(PE)$, where PE is the probability of error.
How are sample quality values generated?	They are generated by a statistical algorithm which is calibrated to estimate the certainty of basecalls.
How is a sample quality value different from the sample score?	The sample score is the average quality value of the bases in the clear range sequence for that sample. A sample quality value is a per-base estimate of basecaller accuracy.
How does the consensus quality value differ from the consensus score?	The consensus score is the average quality value of the bases in the consensus sequence for that specimen. A consensus quality value is a per-base estimate of the accuracy of the consensus-calling algorithm.

Printing and Exporting Results FAQs

Table C-11 Printing and Exporting Results Questions and Answers

Question	Answer
Printing – What can I print in SeqScape® software?	You can print the views only for sample, specimen, segment, project, and complete reports. You can also print electropherograms, complete reports, and the visible data or all data for the project view.
File export – What can I export from SeqScape® software?	User information, projects, project alignments, project templates, reports, nucleotide and amino acid variants, and libraries can be exported from the software. Refer to Appendix E, "User Privileges."
File export – Can I export each consensus sequence individually?	Consensus sequences for a project can be exported as a group by using selecting **File** > **Export** in the Project view.
What format can I print/export reports in?	You can export reports in .pdf, .xml, .htm, or .txt file formats. You can print the exported reports or you can print an open report by selecting **File** > **Print**.
Can I export and print individual .ab1 sample files from the project?	Yes. To export and print individual .ab1 files from within the project, select the sample file in the Project Navigator view, then click a sample file and select **File** > **Export** > **Sample Sequence File**. The sample file can be exported in four formats: .seq, .fsta (FASTA), .phd.1 (PHD), and .ab1. You can print individual sample files by selecting **File** > **Print**.

Audit Trail, Security, and Access Control FAQs

Table C-12 Audit Trail, Security, and Access Control Questions and Answers

Question	Answer
What security and audit trail features are included in SeqScape® software?	SeqScape® software v2.5 (and higher) has the following security and audit trail features: • Three levels of user access • User lockout after a specified time frame has passed • Password expiration • Audit trail that can be created for base change, insertion, or deletion • Audit trail that includes time/date stamp and reason for change • User name that is displayed when logged into the software • Audit Trail report
What are the access control differences as you go from Admin to Scientist to Analyst?	**Administrators** can do everything that the application possibly lets you do. This includes the admin-specific tasks: creating users, viewing and changing user details, importing and exporting users, and changing the Authentication and Audit features that assist with 21 CFR 11 Part 11 requirements. **Scientists** can do everything except the admin-specific tasks. **Analysts** can open projects and import samples, but cannot affect other master objects. For example, an analyst cannot view, modify, import, or export project templates, RDGs, analysis defaults, etc. This includes changing the RDG or analysis settings in a project, although you can change the basecaller and dye set/primer files from within the Sample Manager. An Analyst is allowed to edit the project. The Analyst also cannot perform any admin-specific tasks. For more information, see Appendix E, “User Privileges.”
Does the audit trail function add User ID and a Time/Date stamp to each entry?	Yes. The Audit Trail report does include a user ID, user first and last names, and time/date stamp for each audit event. The Audit Trail report also includes the reason why the user modified the data, and it includes any comments entered.

KB™ Basecaller FAQs

The KB™ Basecaller is designed to reduce manual data review time, elongate the read length of high-quality bases in sequences, and thereby substantially reduce sequencing costs. This new algorithm accurately extracts more bases out of the sequencing data generated on current instrument and chemistry platforms provided by Life Technologies. KB™ Basecaller v1.4 supports all chemistries and run modules available on the 310, 3100/3100-*Avant*, Applied Biosystems® 3130/3130*xl* and 3500/3500*xl* Genetic Analyzers, and on the Applied Biosystems® 3730/3730*xl* DNA Analyzers.

Key Benefits of Using the KB™ Basecaller

Increased Length of Read

The KB™ Basecaller uses advanced algorithms to accurately extract more bases from the 3' and 5' ends of the sequence. Tests on genomic BAC samples indicate a measurable improvement of roughly 100 bases in length-of-read as compared to the same data analyzed by the ABI Basecaller and Phred software (v0.020425.c). The tests were performed on a data set generated by Life Technologies and several customer sites using 3730*xl* instruments. The gain in read length varies depending on the run module used to collect the data. The accuracy of start point estimation and the first 50 bases of called sequence are substantially increased. Typically, ~10 more correct calls on average are identified at the 5' end, as compared to the ABI Basecaller.

Provides Per-Base Quality Value Predictions Using Equation Standardized by Phred Software

The KB™ Basecaller assigns quality values to every basecall. The quality prediction algorithm is calibrated to return Q values that conform to the industry-standard relation established by the Phred software. The KB™ Basecaller and its output are, therefore, interchangeable in pipelines requiring Phred software or output.

Quality value calibration was performed using a controlled set of correct-sequence annotated sample files representative of production sequencing data generated on capillary electrophoresis platforms. Over 49 million basecalls were used to calibrate KB™ Basecaller v1.4 and over 24 million distinct basecalls were used to test the calibration.

Accuracy in Start Point Detection

Improved start-point detection contributes to better mobility shift corrections and greater basecalling accuracy in the first 50 bases. Because the KB™ Basecaller detects the start point accurately, you do not need to manually set start points for each sample.

Optional Detection of Mixed-Base with Quality Values

The KB™ Basecaller provides the option to detect mixed base positions and assign IUB codes and quality values to those positions. Quality values are assigned to mixed basecalls using an algorithm similar to that for pure bases.

The definition conforms to the Phred relation. Quality values for mixed bases are inherently lower than those of pure bases due to the higher error risk associated with interpreting more complex signals. Note that when using the ABI Basecaller or ABI Basecaller and Phred software, a separate analysis stage is required to determine mixed bases.

Increased Accuracy in Regions of Low Signal to Noise or Anomalous Signal Artifacts

The KB™ Basecaller increases the accuracy of sequence reads extracted from low-signal regions or in data partially contaminated by secondary sequence or by other sources of "chemistry noise."

Basecalling errors caused by anomalous chemistry and/or instrument signals (*e.g.*, dye blobs, fluorescent spikes) are substantially reduced. These artifacts are often found in otherwise high-quality "clear-range" data, resulting in the loss of high-quality bases downstream from the noise region. Tests indicate that KB™ Basecaller can better distinguish between target DNA peaks and the most common artifacts, thus allowing the basecaller to better "read through" the noise.

Analysis of Short PCR Products

The KB™ Basecaller has been tested for accuracy in basecalling and quality value estimation on PCR products as short as 100 bases. Although you can basecall products with fewer than 100 bases, such sample files were not tested.

Detection of Failed Samples

The KB™ Basecaller indicates gross sample quality. Each analysis is classified as "Success without warnings," "Success with warnings," or "Failure due to poor data quality." A common failure mode is no signal – i.e., insufficient detection of DNA peaks. For the failed samples, the KB™ Basecaller uses "NNNNN" as the sequence, signaling that the sample quality is very low and may need to be

omitted from further analysis. Failed samples are flagged in reports provided by the analysis software. Note that this behavior is different from the ABI Basecaller, which *always* attempts to call bases, resulting in sequences of many Ns.

Provide the Option to Trim Data Using Per-Base Quality Value

Software integrated with the KB™ Basecaller can automatically determine the clear range region by trimming the ends using the per-base quality values provided by the KB™ Basecaller. The parameters used for trimming are similar to those offered in other tools used by the genome community.

Provide per-sample quality value (QV) that facilitates determining quality of reads

Software with the KB™ Basecaller integrated uses the QV provided by the KB™ Basecaller to trim and also determine a sample score. The sample score is the average QV in the clear range, or in the entire read when no clear range is determined. This single number is a useful measure to determine the quality of the data. The sample score appears in reports generated by Sequencing Analysis Software, SeqScape® Software and/or MicroSeq® ID Software.

Optional Detection of PCR Stop

You can set the KB™ Basecaller to terminate basecalling at a PCR stop. Note that samples with enzymatic failure may have signal properties mirroring those in PCR stop conditions. The KB™ Basecaller may not be able to distinguish between these two cases.

Optional Assignment of Ns

By default, the KB™ Basecaller does not generate Ns; however, you can reassign Ns to bases with QV below a user-specified threshold.

Optional Generation of .Phd.1 files

.phd.1 files can be generated by auto-analysis or in analysis software. The .phd.1 files can be used for further analysis by down-stream software such as Phrap software.

Future Support of ABI and KB™ Basecaller

Although Life Technologies will continue to provide technical support for the ABI Basecaller, further development and defect fixes will be done only on the KB™ Basecaller. If you encounter a defect in the ABI Basecaller, please use the KB™ Basecaller instead. In future releases, ABI Basecaller support files will be removed from the software wherever there is duplicate support in the KB™ Basecaller.

New Features in KB™ Basecaller v1.4.1

- Support for Applied Biosystems® 3500/3500xl Genetic Analyzers
- Support for 3730/3730xl POP-6™
- Support for BigDye® Direct Cycle Sequencing Kit
- Support for fast 3500/3500xl POP-6™ BDTv1.1 run modules, RapidSeq50 collecting at least 450 bases in 65 minutes or less and FastSeq50 collecting at least 600 bases in 90 minutes or less
- Improved Quality Values

Features in KB™ Basecaller v1.3

- Ability to Interpolate between run voltages
- Support for high voltage 3730 POP-7™ BDTv3 TargetSeq module

Features in KB™ Basecaller v1.2

- Improvements over all earlier versions of the KB™ Basecaller (v1.0, v1.1, v1.1.1 and v1.1.2).
- Support for Applied Biosystems® 3130/3130*xl* Genetic Analyzers
- The .scf files generated using the KB™ Basecaller contain quality values
- Content of the “comment” block in phd1 output files conforms better to standards established by Phred

  Note: In the comment block, the lines labeled TRIM and TRACE_PEAK_AREA_RATIO always contain the following default values:

  - TRIM: -1 -1 -1.000000e+000
  - TRACE_PEAK_AREA_RATIO: -1.000000e+000

Comparison of the ABI and KB™ Basecallers

Table C-13 Comparison of the ABI and KB™ Basecallers

Question	ABI Basecaller	KB™ Basecaller
What does the software do?	• Processes raw traces. • Provides processed traces. • Provides AGCTN calls.	• Processes raw traces • Provides processed traces • Provides pure bases only *or* • Provides pure & mixed calls (R, Y, K, M, S, or W) • Provides quality values • Generates phd.1 and .scf files • Provides a sample score
What are the resulting basecalls?	One available option: • Mixed bases are assigned as Ns. Further processing (either manually or by additional software) is required to assign IUB codes to the Ns or pure bases.	Four available options: • Assigns A, C, G, or T and a Q value to each peak • Assigns A, C, G, or T and a Q value to each peak. Any peak with Q value below a defined threshold is reassigned an N • Assigns A, C, G, T, or a mixed base and a Q value to each peak • Assigns A, C, G, T, or a mixed base and a Q value to each peak. Any peak with Q value below a defined threshold is reassigned an N
How are failed samples handled? (no signals, chemistry failure)	Attempts to call all bases, so sample results in many Ns.	Assigns 5 Ns to the entire sample to indicate that the sample failed analysis Analysis report flags these files
How does the baseline appear in processed data?	Appears smoother.	Appears less smooth. (See the FAQ "Why does the baseline look less smooth when the data are analyzed with the KB™ Basecaller?" on page C-30.)
How are the data processed?	Uses ABI Basecaller to call bases on Windows OS.	Uses KB™ Basecaller to call bases and estimate QVs on Windows OS

Table C-13 Comparison of the ABI and KB™ Basecallers

Question	ABI Basecaller	KB™ Basecaller
What are the supported instruments and future developments?	310, 373, 377, 3100/3100-*Avant*, and 3700 and Applied Biosystems® 3130/3130*xl* and 3730/3730*xl* instruments. No longer under development.	310, 3100/3100-Avant, 3130/3130*xl,* 3500/3500*xl,* and 3730/3730*xl* instruments. Development is ongoing.

Differences Between the ABI and KB™ Basecallers

Table C-14 Differences between the ABI and KB™ Basecallers

Question	Answer
Can the KB™ Basecaller be used to basecall short PCR products?	The KB™ Basecaller has been tested for accuracy in basecalling and quality value estimation on PCR products as short as 100 bases. It may be possible to basecall products with less than 100 bases, but such sample files have not been tested. Samples significantly shorter than 100 bases may not contain enough signal information needed by the basecaller to process the sample file. SeqScape® Software analyzes sequence files generated from 310, 377, 3100, 3100-*Avant*, 3700 and Applied Biosystems® 3130/310xl, 3500/3500*xl,* and 3730/3730*xl* instruments. The software also accepts text sequences in FASTA format.
Why does the baseline look less smooth when the data are analyzed with the KB™ Basecaller?	Processed signals or traces provided by the ABI Basecaller will appear smoother than those provided by the KB™ Basecaller because each algorithm processes the signals somewhat differently. With the ABI Basecaller, only AGCT and Ns are assigned to each peak. Therefore, you must manually search for mixed bases or use a secondary software to complete the task. To facilitate this secondary process, the ABI Basecaller subtracts a more aggressive baseline estimate to present a cleaner baseline in the processed signals. Because the KB™ Basecaller can determine pure and mixed bases, there is no need for second-stage processing, which allows for less aggressive baseline subtraction. The processed traces will have a higher baseline. If you have mixed bases, turn on the mixed-base detection option and allow KB™ Basecaller to call mixed bases. Use the mixed base calls and the associated QVs to review mixed bases. Do not simply look at the baseline.
What is the signal-to-noise value for data analyzed with the KB™ Basecaller?	The KB™ Basecaller calculates signal-to-noise information and presents the data in the Annotation view and analysis report. The ABI Basecaller calculates only the signal intensity. The signal-to-noise ratio is more informative of data quality than the signal intensity value alone. Both properties are important in determining quality.

Table C-14 Differences between the ABI and KB™ Basecallers *(continued)*

Question	Answer
What are the scaling options available with the KB™ Basecaller?	With the KB™ Basecaller, you have two options for scaling data: • **True-profile scaling** – The processed traces are scaled uniformly so that the average height of peaks in the region of strongest signal is about equal to a fixed value (*e.g.*, 1000). The profile of the processed traces is similar to that of the raw traces. • **Flat-profile scaling** – The processed traces are scaled semi-locally so that the average height of peaks in any region is about equal to a fixed value (*e.g.*, 1000). The profile of the processed traces is flat on an intermediate scale (> about 40 bases). You should decide which option is better suited to your particular circumstances. The sequence and QVs called by the KB™ Basecaller are *independent* of the selected scaling option. Options for scaling data are not provided with the ABI Basecaller. The ABI Basecaller employs a scaling method closer to the "True profile" option than the "Flat profile" option
Will I get more "good" sample files using the KB™ Basecaller?	Our tests show that medium- and high-quality data yield more usable bases (*i.e.*, longer read length) when analyzed by the KB™ Basecaller as compared to results produced by the ABI Basecaller. For extremely poor-quality data, the KB™ Basecaller does not provide more bases but instead fails the samples, that is, no signal, extremely low signals, or extremely noisy signals. By calling a string of "NNNNN" for the failed samples (instead a sequence all containing low QVs), the KB™ Basecaller indicates that the sample is *unusable*.
Can the KB™ Basecaller analyze data generated on 373, 377 or 3700 instruments?	No. The KB™ Basecaller is not calibrated for this task. It is calibrated to basecall and estimate the basecall quality for specific combinations of instrument/polymer/chemistry/run condition that are currently supported on 310, 3100/3100-Avant, and Applied Biosystems® 3130/3130*xl,* 3730/3730*xl,* and 3500/3500*xl* instruments. There are no plans to include support for analysis of data from the 373, 377, or 3700 instruments.
How can I determine which basecaller was used to analyze each sample file?	The Annotation view for each sample file and the print header contain the basecaller name and version number. When displaying samples files, files analyzed by the KB™ Basecaller display QV-value bars above the electropherogram.

Table C-14 Differences between the ABI and KB™ Basecallers *(continued)*

Question	Answer
Are there any known incompatibilities when a sample file is analyzed with the KB™ Basecaller?	Life Technologies does not know of any incompatibility issues when a sample file (.ab1) is analyzed with the KB™ Basecaller and used in third-party software.

Processing Data with Phred Software and .phd1 Files FAQs

Table C-15 Processing Data with Phred Software and.phd1 Files FAQs

Question	Answer
Can I analyze sample files with the KB™ Basecaller and then reprocess it with Phred software?	In principle, yes, but this is not recommended. The resulting quality values from Phred software are not calibrated, that is, Phred can over- or under-predict quality in certain circumstances because it has not been trained on the type of processed electropherogram produced by the KB™ Basecaller. (Phred software has been trained using the ABI Basecaller to produce the processed traces.) In addition, because Phred software replaces (and ignores) the initial called sequence, reprocessing KB-analyzed samples with Phred software, on average, degrades the accuracy of the analysis in terms of actual sequence error. In this case, the analysis improvements provided by KB™ Basecaller are lost. **Note:** Our studies indicate that running Phred software on sample files processed by the KB™ Basecaller significantly *degrades* the quality of the results. Analysis with KB™ Basecaller can output .phd.1 files, which are interchangeable with any pipeline that currently depends on Phred software.

Table C-15 Processing Data with Phred Software and.phd1 Files FAQs *(continued)*

Question	Answer
Which Life Technologies software generates .phd.1 files?	The following software products have KB™ Basecaller integrated and can generate .phd.1 files: • 3100/3100-*Avant* Data Collection Software v3.0 • Applied Biosystems® 3730/3730*xl* Data Collection v2.0 or later • Applied Biosystems® 3130/3130*xl* Data Collection Software v3.0 • Sequencing Analysis Software v5.2 • SeqScape® Software v2.5 • MicroSeq® ID Software v1.0

Quality Values FAQs

Table C-16 Quality Values Questions and Answers

Question	Answer
How should I use quality values to review data?	When analyzing data with pure bases, Life Technologies recommends that you set Low QV = <15, Medium QV = 15 to 19, and High QV = 20+ (default). When reviewing data with pure bases, use the quality values to briefly review bases with QV >20. Pay close attention to bases with medium QVs because you may need to make edits. Quickly review low-QV bases, although most likely you will discard these bases from further analysis. When reviewing mixed bases, your quality values will be lower than pure bases. Review all mixed bases. In all cases, keep in mind that, by definition, the predicted probability of error for a particular basecall is equal to $10^{-q/10}$.
What are the differences in quality values between mixed bases and pure bases?	The definition of quality values is the same for pure and mixed bases. In both cases the probability of error for the associated basecall is $10^{-q/10}$. The distribution of quality values assigned to mixed bases, however, differs dramatically from that for pure bases. Typically, high-quality pure bases are assigned QVs of 20 or higher. Mixed base QVs range from 1 to 20; accurate mixed bases can have low quality values. The reason that a high quality mixed base can receive such low QVs is that the probability of error with more complex signals is higher. Do not discard mixed bases solely based on QVs. It is a good practice to review all mixed bases.

Table C-16 Quality Values Questions and Answers *(continued)*

<table>
<tr><th>Question</th><th>Answer</th></tr>
<tr><td>Can I trim my data using quality values?</td><td>Yes. When using data collection software, you can set trimming using QVs in the analysis protocols.

When using Sequencing Analysis, SeqScape®, or MicroSeq® ID software, you can set trimming using QVs in the Analysis settings.</td></tr>
<tr><td>Is there a table mapping each quality value and the corresponding probability of error?</td><td>The table below maps each quality value to the corresponding probability of error. For a more extensive table, look in the Help menu or the Sequencing Analysis or SeqScape® Software User Guides.

<table>
<tr><th>QV</th><th>Pe</th><th>QV</th><th>Pe</th></tr>
<tr><td>1</td><td>79.0%</td><td>35</td><td>0.032%</td></tr>
<tr><td>5</td><td>32.0%</td><td>40</td><td>0.010%</td></tr>
<tr><td>10</td><td>10.0%</td><td>41</td><td>0.0079%</td></tr>
<tr><td>15</td><td>3.2%</td><td>45</td><td>0.0032%</td></tr>
<tr><td>20</td><td>1.0%</td><td>50</td><td>0.0010%</td></tr>
<tr><td>21</td><td>0.79%</td><td>60</td><td>0.00010%</td></tr>
<tr><td>25</td><td>0.32%</td><td>99</td><td>0.00000000013%</td></tr>
<tr><td>30</td><td>0.10%</td><td></td><td></td></tr>
</table>
</td></tr>
<tr><td>Where can I see quality value bars and numbers?</td><td>Sequencing Analysis, SeqScape®, and MicroSeq® ID software allow you to display or hide quality value bars in displays and printouts. You can customize the color and range for low-, medium-, and high-quality values. For QV ≤ 50, the length of a bar is proportional to the corresponding quality value. Quality values above 50 have the same color and QV bar length as that defined for a QV of 50. To see the quality value for a particular base, position the cursor over the QV bar.

In SeqScape® Software and MicroSeq® ID Software, the per-base quality values also appear in the reports corresponding to bases identified as mutations.</td></tr>
</table>

Table C-16 Quality Values Questions and Answers *(continued)*

Question	Answer
Why are the quality value bars displayed in gray?	A quality value is assigned to a specific basecall. When you alter the basecall the quality value no longer applies to the new base. Therefore, it is displayed as a gray bar. Also, when you reassign Ns to bases below a certain QV, the QV bar is not applicable to the N basecall. Therefore, it is displayed as a gray bar
Are quality value bars printed for the Electropherogram or Sequence views?	You can show or hide QV bars when printing the Electropherogram or Sequence view of the sample file. QV bars cannot be printed if you print more than seven panels per page, due to space limitations. The actual quality value numbers cannot be printed.
Which Life Technologies software can display the quality values?	Sequencing Analysis Software v5.X, SeqScape® Software v2.X, and MicroSeq® ID Software v1.X can display quality values. Sequencing Analysis Software v3.X and SeqScape® Software v1.X can open and display the sample files with quality values, but the QVs are not displayed.
Can I view quality values provided by KB™ Basecaller with other software?	Quality value graphic views are customized for software provided by Life Technologies. The design allows for additional functionality such as clear range trimming and more streamlined editing.

Miscellaneous Basecaller FAQs

Table C-17 Questions about Ns, spacing values, and providing feedback

Question	Answer
When will I see Ns in samples analyzed by the KB™ Basecaller?	When using the KB™ Basecaller, you see the sequence “NNNNN” when the sample fails analysis. Omit this file from further analysis. The Analysis Report in Sequencing Analysis Software also flags these files. In addition to pure and mixed bases with QV bars, you can also see Ns and gray QV bars when you choose to reassign Ns to all bases before the user-specified QV threshold. This option allows you to analyze data with the KB™ Basecaller but share data with others who do not have software that can display quality values. This allows you to take advantage of the longer read length and more accurate basecalling provided by the KB™ Basecaller while still viewing data with software that does not display QVs.
Why does the spacing value sometimes appear in red?	When the ABI Basecaller fails to determine a spacing value for a sample file, it uses a default value of 12.00 for all run conditions. This number appears in red in the Sample Manager, and the Annotation view displays “-12.00.”
Why does the spacing value sometimes have a negative value?	When the KB™ Basecaller fails to determine a spacing value for a sample file, it uses a default value specific to the particular instrument/polymer/chemistry/run condition used to generate the sample file. This number appears in red in the Sample Manager and the Annotation view displays –1 times this value.
How do I provide feedback to the KB™ Basecaller product team?	Please send feedback information to your local Life Technologies applications support representative. You can also find technical support at lifetechnologies.com/support. Whenever possible, please include sample files and detailed instructions (including analysis settings) on how to reproduce your observation.

Conference References

References

1. B. Ewing and P. Green, Genome Research, 8:186-194, 1998.

D Translation Tables

In This Appendix

IUPAC/IUB Codes

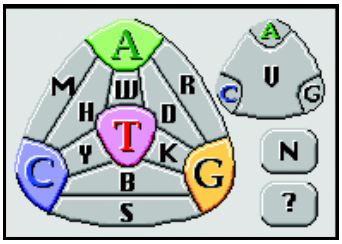
Table D-1 IUPAC/IUB Codes

Code	Translation
A	Adenosine
C	Cytidine
G	Guanosine
T	Thymidine
B	C, G, or T
D	A, G, or T
H	A, C, or T
R	A or G (puRine)
Y	C or T (pYrimidine)
K	G or T (Keto)
M	A or C (aMino)
S	G or C (Strong—3 H bonds)
W	A or T (Weak—2 H bonds)
N	aNy base
V	A, C, or G

Note: This chart is accessible from the Help menu.

IUPAC Diagrams

IUPAC



IUPAC heterozygous

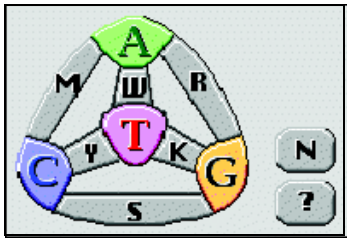


Figure D-1 IUPAC Diagrams

Complements

Table D-2 Complements

A	T	S	S
C	G	W	W
G	C		
T	A	B	V
		D	H
R	Y	H	D
Y	R	V	B
K	M	N	N
M	K		

Note: These charts are accessible from the Help menu.

Universal Genetic Code

Table D-3 Universal Genetic Codes

5' End	2nd Position				3' End
	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	OCH	OPA	A
	Leu	Ser	AMB	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G
Stop Codes: AMBer, OCHer, OPA					

Note: This chart is accessible from the Help menu.

Amino Acid Abbreviations

Table D-4 Amino Acid Abbreviations

Amino Acid	Three Letters	One Letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Any Amino Acid		X

Note: This chart is accessible from the Help menu.

E User Privileges

In This Appendix

This appendix contains a list of privileges for users of the three categories, Administrator, Scientist, and Analyst, when they use SeqScape® Software.

Tables of User Privileges

Table E-1 Access for Admin Level

Access		Description of access for users of Admin level only	Admin	Scientist	Analyst
Admin only access	1	Create User Accounts	Allowed	Not Allowed	Not Allowed
	2	Exporting/Importing User Accounts			
	3	Export a Project/PT/RDG/Library from the SeqScape® Manager			
	4	Import objects from outside the DataStore into the SeqScape® Manger			
	5	Install SeqScape® for an automated analysis system			

Table E-2 Access for Admin and Scientist Levels

		Description of access for users of Admin and Scientist levels	Admin	Scientist	Analyst
SeqScape® Manager	1	Delete an object from the SeqScape® manager	Allowed	Allowed	Not Allowed
	2	Delete a Project from the SeqScape® manager			
	3	Save As... an object in the SeqScape® manager			
	4	Create a new object in the SeqScape® Manager			
	5	Create a new Project Template in SeqScape® Manager			
	6	Configure analysis defaults in SeqScape® manager			
	7	Deleting entries from a library in the SeqScape® manager			
	8	Re-Configure an existing Project Template in the SeqScape® Manager			
Analysis Protocol & Settings	9	Creating an analysis protocol	Allowed	Allowed	Not Allowed
	10	Editing an existing analysis protocol			
	11	Apply an analysis protocol to a set of samples (project/sample/specimen)			
	12	Create new Primary Seq Analysis Protocols			
	13	Set Clear range determination in Analysis settings or analysis defaults			
	14	Set Mixed Base determination in Analysis settings or analysis defaults in a Project, PT/SS Manager			

Table E-2 Access for Admin and Scientist Levels *(continued)*

		Description of access for users of Admin and Scientist levels	Admin	Scientist	Analyst
RDG	15	RDG: Import Variants and Reference into an RDG from a set of aligned FASTA files	Allowed	Allowed	Not Allowed
	16	RDG general tab: configure an RDG in general tab			
	17	RDG ROI tab: Edit a Reference Data Group (RDG): configure Layers			
	18	RDG ROI tab: Edit a Reference Data Group (RDG): configure ROIs			
	19	RDG ROI tab: Edit a Reference Data Group to use an implicit reference			
	20	RDG ROI tab: adding/modifying a Reference Segment			
	21	RDG ROI tab: Change the Reference Segment index Base in an embedded RDG			
	22	RDG ROI tab: deleting a Layer			
	23	RDG ROI tab: deleting a Reference Segment			
	24	RDG ROI tab: deleting an ROI			
	25	RDG ROI tab: Import GenBank sequences into the RDG for automated Ref Segment and feature creation			
	26	RDG NT variants Tab: Edit NT variants in an RDG			
	27	RDG NT variants Tab: Import NT variants from a Tab Delimited Text into RDG			
	28	RDG AA variants Tab: Add amino acid variants to an RDG			
	29	RDG AA variants Tab: Edit AA variants in a RDG			

Table E-2 Access for Admin and Scientist Levels *(continued)*

		Description of access for users of Admin and Scientist levels	Admin	Scientist	Analyst
RDG	30	RDG AA variants Tab: Import AA variants from a tab delimited text file into RDG	Allowed	Allowed	Not Allowed
	31	RDG variant styles tab: configure an RDG in Variant Styles tab			
Library	32	Library: overwriting/appending sequences to an existing library	Allowed	Allowed	Not Allowed
	33	Library: editing sequence data in the library			
	34	Library: exporting data from the library as a Multi-FASTA file			
	35	Library: viewing/editing library types in the Library Type manager			
	36	Library: creating a new sequence library			
Other	37	Sets General Preferences in Options	Allowed	Allowed	Not Allowed
	38	Sets Sequence Collector (Database) Preferences in Options			
	39	Add NT or AA variants from any data view			
	40	Set specimen level analysis settings			
	41	Set project level analysis settings			

Table E-3 Access for Admin, Scientist and Analyst Levels

		Description of access for users of Admin, Scientist and Analyst levels	Admin	Scientist	Analyst
Reports	1	View Reports	Allowed	Allowed	Allowed
	2	View Reports with enabled links back to primary data			
	3	View Reports while editing project			
	4	Export all reports			
	5	Export all customized reports			
	6	Print all reports			
	7	View heterozygous frame shifts links from Mutations Report			
	8	Print a report from the reports manager			
Project View/Display	9	Move sample data from one Specimen to another	Allowed	Allowed	Allowed
	10	Display SQVs and CQVs			
	11	Re-order aligned Specimen consensi			
	12	Change active Layer view			
	13	Show/hide variants that result in silent mutations			
	14	Sort Summary Table in Specimen view			
	15	Display Sample and Consensus Scores			
	16	View Amino Acid tooltips for degenerate codons			
	17	View Amino Acid Alignment in Main Window			
	18	View Library Search Results in Alignment View Identification Pane			

Table E-3 Access for Admin, Scientist and Analyst Levels *(continued)*

		Description of access for users of Admin, Scientist and Analyst levels	Admin	Scientist	Analyst
Project View/Display	19	View electropherogram data as aligned peaks	Allowed	Allowed	Allowed
	20	View all objects in Project Navigator and Main Windows			
	21	View Specimen Layout			
	22	View Specimen-Segment Assembly tab			
	23	View Unassembled data in the Project Navigator and Specimen Views			
	24	View/Navigate through electropherogram snippets			
	25	View/Navigate Specimen Segment electropherogram data			
	26	View a Project/Navigate using the Overview pane			
	27	View Samples in the Sample Manager tab			
	28	View/Navigate alignments using the display toolbar buttons			
Project-Other Controls	29	Apply a new Project Template to an existing Project	Allowed	Allowed	Allowed
	30	Create a new Project from the SeqScape® Toolbar			
	31	Delete Samples in Project Navigator			
	32	Delete Specimens in Project Navigator			
	33	Export Project Alignment in FASTA format			
	34	Export Sample data in SEQ, FASTA or AB1 format			
	35	Export Specimen consensus or aligned sample sequences in FASTA format			

Table E-3 Access for Admin, Scientist and Analyst Levels *(continued)*

		Description of access for users of Admin, Scientist and Analyst levels	Admin	Scientist	Analyst
Project-Other Controls	36	Import a Text segment to a Text Specimen	Allowed	Allowed	Allowed
	37	Import Samples to Project			
	38	Import Samples to Project from Database (Sequence Collector)			
	39	Import/create a text-only Specimen			
	40	Open an embedded Settings Object inside a Project			
	41	Open an existing Project			
	42	Print wrapped nucleotide or amino acid Project Alignments			
	43	Save Project from the Menu or Toolbar			
	44	Search for text strings in any sequence data			

Table E-3 Access for Admin, Scientist and Analyst Levels *(continued)*

		Description of access for users of Admin, Scientist and Analyst levels	Admin	Scientist	Analyst
Editing	45	Generate an Audit Trail event	Allowed	Allowed	Allowed
	46	Project Alignment view: Change consensus basecalls			
	47	Project Alignment view: Insert or delete a space in a Reference			
	48	Project Alignment view: Insert or delete a space in a Specimen consensus			
	49	Project Alignment view: Insert/delete Consensus bases			
	50	Project Navigator: Rename Specimens			
	51	ROI tab: Rename Segments in RDG			
	52	Specimen view: Change a base in a sample			
	53	Specimen view: Change basecalls in the consensus			
Editing	54	Specimen view: Change the Clear Range for sample data	Allowed	Allowed	Allowed
	55	Specimen view: Insert or delete a base in a sample			
	56	Specimen view: Insert or delete bases in consensus			
	57	Undo base edits			
SeqScape® Manager	58	Open the SeqScape® Manager	Allowed	Allowed	Allowed
	59	Save any SeqScape® Manager Object			
Library	60	View the Libraries in the SeqScape® Manager	Allowed	Allowed	Allowed
	61	View results of Library search in the Project Alignment View			

Table E-3 Access for Admin, Scientist and Analyst Levels *(continued)*

		Description of access for users of Admin, Scientist and Analyst levels	Admin	Scientist	Analyst
Analysis Protocol and Settings	62	View the Analysis Protocol	Allowed	Allowed	Allowed
	63	Change basecaller settings in an existing Sample within a Project			
	64	Reconfigure Analysis Defaults inside a Project			
	65	Configure Display Settings in Project or SeqScape® Manager			
	66	Analyze data using the BGB without basecalling samples			
	67	Analyze data using the BGB			
	68	Indicate that specific Samples are not to be basecalled			

Table E-3 Access for Admin, Scientist and Analyst Levels *(continued)*

		Description of access for users of Admin, Scientist and Analyst levels	Admin	Scientist	Analyst
Other	69	Browse/Locate data in the file system	Allowed	Allowed	Allowed
	70	Exit SeqScape®			
	71	Sort items in columns in any table in SeqScape®			
	72	Install SeqScape® on a clean system			
	73	Upgrade SeqScape® Software v1.0 or v1.1 to v2.0			
	74	Uninstall SeqScape®			
	75	Launch SeqScape®			
	76	Configure a sample in Data Collection for automated import into SeqScape®			

F

Aligned Variant and FASTA File Format

In This Appendix

Tab-Delimited Files

You can import variants into the SeqScape® Software if they are in the format of a tab-delimited text file.

Creating a Variant Text File

SeqScape® software tab-delimited text files must conform to the following rules:

- One variant per line
- The following tab-delimited column headings:

NT Variant Headings	AA Variant Headings
Type	Type
ROI	Layer
NT position	AA position
Reference	Reference
Variant	Variant
Style	Style
Description	Description
Used by all ROIs	

An example is provided in Figure F-1.

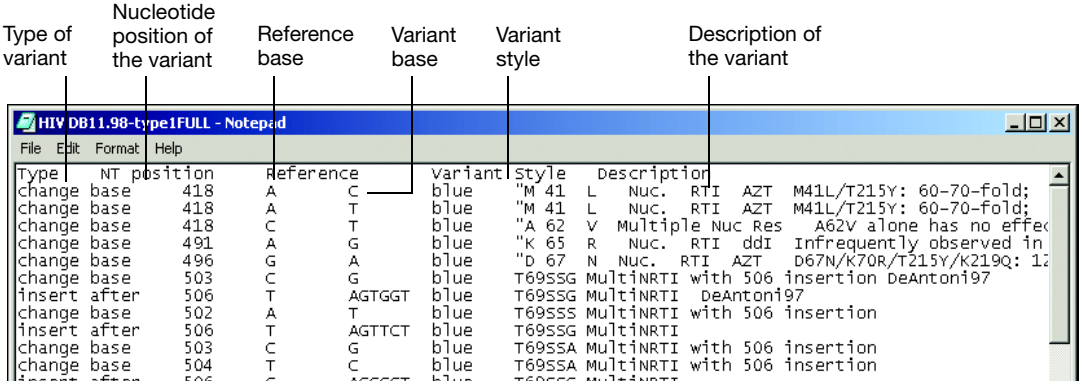


Figure F-1 Sample of Variant Tab-Delimited Text Format

FASTA File Format

Note: The information on FASTA was obtained from http://www.ncbi.nlm.nih.gov

FASTA Format Description

A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (>) symbol in the first column.

FASTA Format Example

An example sequence in FASTA format is as follows:

>HIV HXB2 Prt-RT1(1-320)
cctcaggtcactctttggcaacgacccctcgtcacaataaagataggggggcaactaaaggaag
ctctattagatacaggagcagatgatacagtattagaagaaatgagtttgccaggaagatggaaa
ccaaaaatgatagggggaattggaggttttatcaaagtaagacagtatgatcagatactcatagaa
atctgtggacataaagctataggtacagtattagtaggacctacacctgtcaacataattggaaga
aatctgttgactcagattggttgcactttaaattttcccattagccctattgagactgtaccagtaaaat
taaagccaggaatggatggcccaaaagttaaacaatggccattgacagaagaaaaaataaaag
cattagtagaaatttgtacagagatggaaaaggaagggaaaatttcaaaaattgggcctgaaaat
ccatacaatactccagtatttgccataaagaaaaaagacagtactaaatggagaaaattagtagatt
tcagagaacttaataagagaactcaagacttctgggaagttcaattaggaataccacatcccgca
gggttaaaaaagaaaaaatcagtaacagtactggatgtgggtgatgcatatttttcagttcccttag
atgaagacttcaggaagtatactgcatttaccatacctagtataaacaatgagacaccagggatta
gatatcagtacaatgtgcttccacagggatggaaaggatcaccagcaatattccaaagtagcatg
acaaaaatcttagagccttttagaaaacaaaatccagacatagttatctatcaatacatggatgattt
gtatgtaggatctgacttagaaatagggcagcatagaacaaaaatagaggagctgagacaacat
ctgttgaggtggggacttaccacaccagacaaaaaacatcagaaagaacctccattcctttggat
gggttatgaactccatcctgataaatggacagtacagcctatagtgctgccagaaaaagacagct
ggactgtcaatgacatacagaagttagtggggaaattgaattgggcaagtcagatttacccaggg
attaaagtaaggcaattatgtaaactccttagaggaaccaaagcactaacagaagtaataccacta
acagaagaagcagagctagaactggcagaaaacagagagattctaaaagaaccagtacatgg
agtgtattatga

IUB/IUPAC Codes

Sequences are expected to be represented in the standard IUB/IUPAC amino acid and nucleic acid codes, with the following exceptions:

- Lower-case letters are accepted and are mapped into uppercase
- In amino acid sequences, U and * (asterisk) are acceptable letters (see below)

Note: Although FASTA codes allow a hyphen or dash to represent a gap in nucleotide sequences, this practice is not acceptable for using FASTA format in SeqScape® software.

Before importing a sequence, any numerical digits or spaces in the sequence need to be either removed or replaced by appropriate letter codes (for example, N for unknown nucleic acid residue or X for unknown amino acid residue).

Supported Nucleic Acid Codes

Table F-1 Accepted Nucleic Acid Codes

Code …	Represents …
A	Adenosine
C	Cytidine
G	Guanine
T	Thymidine
U	Uridine
R	GA (purine)
Y	TC (pyrimidine)
K	GT (keto)
M	AC (amino)
S	GC (strong)
W	AT (weak)
B	GTC
D	GAT
H	ACT

Table F-1 Accepted Nucleic Acid Codes

Code …	Represents …
V	GCA
N	AGCT

Supported Amino Acid Codes

Table F-2 Accepted Amino Acid Codes

Code …	Represents …
A	Alanine
B	Aspartate or asparagine
C	Cystine
D	Aspartate
E	Glutamate
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
U	Selenocysteine
V	Valine
W	Tryptophan
Y	Tyrosine
Z	Glutamate or glutamine

Table F-2 Accepted Amino Acid Codes *(continued)*

Code …	Represents …
X	Any
*	Translation stop
-	Gap of indeterminate length

Library and BLAST Searching

In This Appendix

About the Search Feature

SeqScape® Software is designed to perform two levels of analysis:

- *Identification of nucleotide and amino acid variants:* The software compares each consensus sequence to the reference sequence and all differences will be reported as a variant. If a set of known variants was supplied with the reference sequence, the software will further classified all variants found as known or unknown variants.
- *Identification of sequences from a library that match each consensus sequence:* The software compares each consensus sequence to a library of sequences and top matches will be identified. In addition to the name of each sequence that match the consensus, the software reports the number and the location of each polymorphic position.

This library search feature is designed for use only with libraries of alleles, genotypes or haplotypes where all the sequences in the library are variations of the same sequence. This feature is *not* designed for searching against a library of diverse sequences.

To use this feature, you must design your library such that all library sequences have these characteristics:

- All library sequences must be pre-aligned.
- All library sequences must be equal in length.
- All library sequences must be variations of the same sequence (variations must be less than 50%).
- All library sequences must cover the same region as the Layer that is associated with that library. When using a layer that contains Exon 2, 3, 7, & 8, a valid library should have sequences from Exons 2, 3, 7, & 8, a library with sequences covering exons 2, 3, 4, 5, 7, & 8 would be invalid.
- A library is associated to its specific Layer.

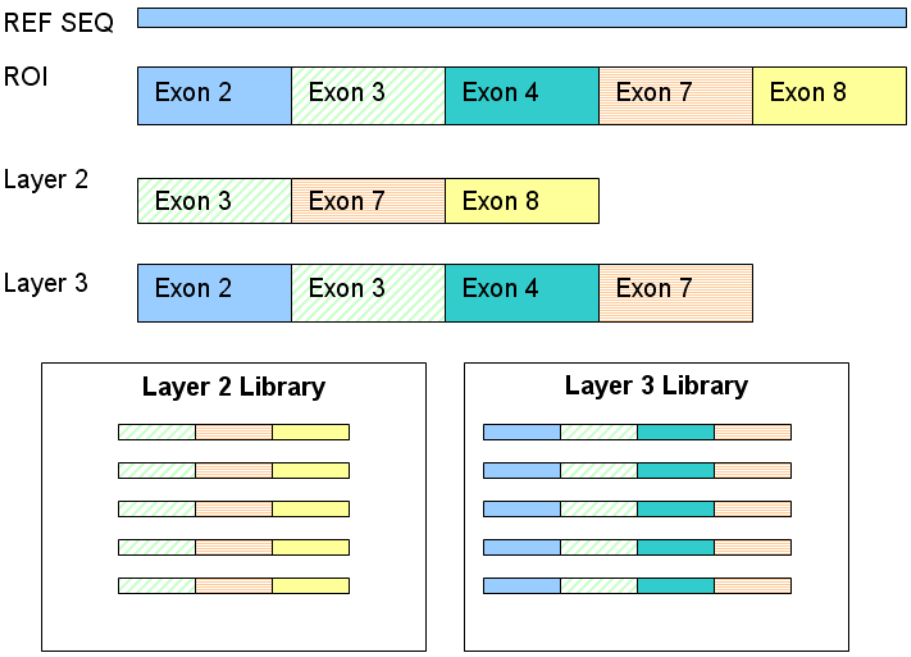


Figure G-1 When working with projects with multiple layers and multiple libraries, ensure that different layers (which would not have the same set of regions of interest) are not compared to the same library.

About Creating a Multi-Aligned FASTA File

If you have sequences of equal length that are already aligned, go directly to SeqScape® software to create your sequence library.

This section contains the instructions for two different methods of preparing your sequences before creating a library.

- **Method 1** – Using SeqScape® software to create a multi-aligned FASTA file
- **Method 2** – Using ClustalX software to create a multi-aligned FASTA file

About ClustalX Software

ClustalX software is a powerful multiple-sequence-alignment program available free of charge on the Internet. See: Jeanmougin *et al*. (1998) *Trends Biochem. Sci.* **23**, 403-5.

Method 1: Create a Multi-Aligned FASTA File Using SeqScape® Software

Create a Multi-Aligned FASTA File

To create a multi-aligned FASTA file using SeqScape® software:

Note: You can try this procedure by obtaining HLA sequences from GenBank database. The file names are listed below as bullet-point items.

1. Obtain the reference sequence:
   - gi|512471|emb|X76776.1|HSHLADMBG
2. Obtain files that contain the sequences that will be used in your sequence library (.txt or .fsta files):
   - gi-1045472-gb-U32663.1-HSU32663-Human-MHC-class-II-antigen-HLA-DM-beta-chain-(HLA-DMB)-gene
   - gi-1373022-gb-U31743.1-HSU31743-Human-HLA-DMB-variant-gene
   - gi-2315188-emb-Y14395.1-HSHLADMB3-Homo-sapiens-HLA-DMB-gene
   - gi-512471-emb-X76776.1-HSHLADMBG-H.sapiens-HLA-DMB-gene
   - gi-881918-gb-U16762.5-HSU16762-Human-(DMB)-gene
3. Launch SeqScape® software and create an RDG using the file obtained in step 1.
4. Create a project template.
5. Start a new project using the project template created.
6. Add the files obtained in step 2 and analyze.
7. Select **File** > **Export**.
8. Select **Project Alignment-Nucleotides**.

You can now use this exported file to create a library in SeqScape® software. See "Method 2: Create a Multi-Aligned FASTA File Using ClustalX Software" on page G-5.

Method 2: Create a Multi-Aligned FASTA File Using ClustalX Software

1. Obtain ClustalX Software

To obtain ClustalX software:

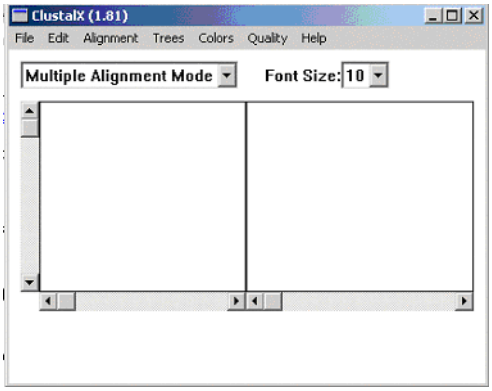
1. Launch an internet browser and search for "ClustalX" or go to: http://inn-prot.weizmann.ac.il/software/
2. Select ClustalX software from the list of software products available.
3. Select the ClustalX software for your operating system.
4. Download the software to your computer.

2. Create a Multi-Aligned FASTA File

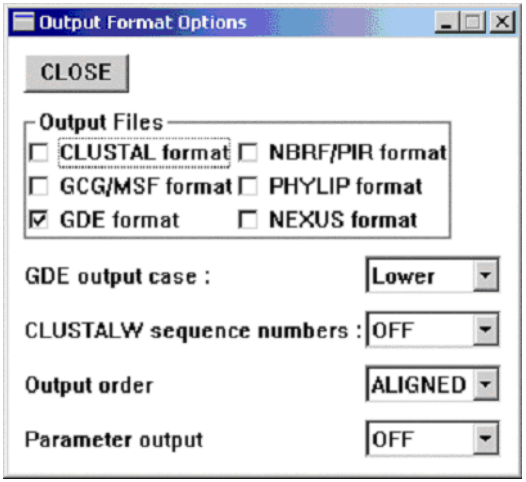
To create a multi-aligned FASTA file using ClustalX software:

Note: You can try this procedure by obtaining HLA sequences from ftp://ftp.ebi.ac.uk/pub/databases/imgt/mhc/hla/, then selecting "DRB_nuc.fasta."

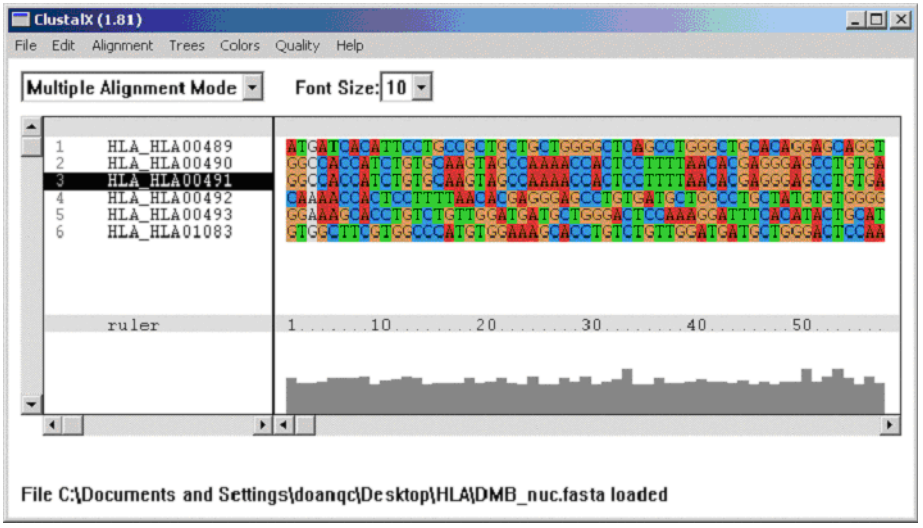
1. Open the folder that you downloaded.
2. Double-click the ClustalX icon.



3. Select **Multiple Alignment Mode**.
4. Select **Alignment** > **Output Format Options**.
5. Select **GDE format** and set the other parameters as indicated below.

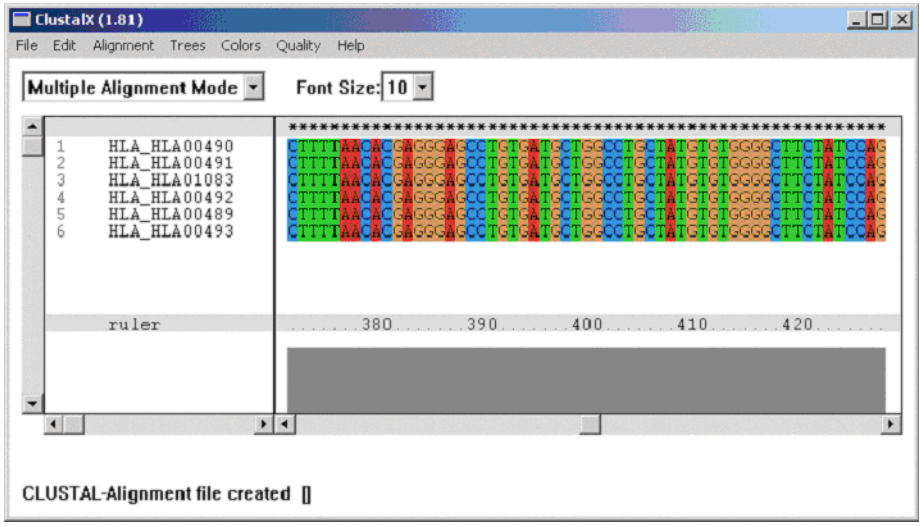


6. Select **File** > **Load Sequences**. To obtain a file, go to http://www.ebi.ac.uk/imgt/hla/, then select the **Download** tab.
7. Browse to the location of your multiple sequence file, then select the file.



8. Select **Edit** > **Select All Sequences**.
9. Select **Alignment** > **Alignment Parameters** > **Reset All Gaps Before Alignment**.
10. Select **Alignment** > **Do Complete Alignment**.

11. A dialog box prompts you to select the location where the results should be saved. Select the location.
12. Click the **ALIGN** button and wait for the alignment to be completed. Notice that dashes are inserted where bases were missing.



13. Navigate to the file that was created. The extension on this file is .gde.

3. Edit Characters in the .gde File

To edit characters in the .gde file from ClustalX using a text editor:

1. Launch a text editor, such as Microsoft® Word.
2. Open the .gde file that you just created.
3. Search for all "#" and replace with ">".
4. Save the file as a text document with the file extension .txt or .fsta.

   Note: The format of the sequences in this document is commonly referred to as a multi-aligned FASTA format.

You can now use this exported file to create a library in SeqScape® software. See "Method 2: Create a Multi-Aligned FASTA File Using ClustalX Software" on page G-5.

Creating Your Library in SeqScape® Software

To create your library in SeqScape® software:

1. Launch SeqScape® software.
2. Select **Tools** > **SeqScape Manager**.
3. Select the **Library** tab, then click **New**.
4. Enter a library name.
5. Select the **Entries** tab, then click **Import**.
6. Select **Display All File Types**, then navigate to and open the multi-aligned FASTA file.

   The software creates the library and displays the number of sequences added.

Software Warranty Information

Computer Configuration

Life Technologies supplies or recommends certain configurations of computer hardware, software, and peripherals for use with its instrumentation. Life Technologies reserves the right to decline support for or impose extra charges for supporting nonstandard computer configurations or components that have not been supplied or recommended by Life Technologies. Life Technologies also reserves the right to require that computer hardware and software be restored to the standard configuration prior to providing service or technical support. For systems that have built-in computers or processing units, installing unauthorized hardware or software may void the Warranty or Service Plan.

Glossary

ABI basecaller	An algorithm used in earlier versions of DNA Sequencing Analysis and SeqScape® Software.
administration	The functions of SeqScape® software relating to installing, removing, or updating the application.
aligned allele library	A collection of aligned sequences that are all variations of the same sequence. This is the only type of library supported in SeqScape® v2.0. An aligned allele library differs from a library of diverse sequences such as a library of different gene sequences, and is also different from a library of unaligned sequences.
alignment	The aligned reference sequence together with the aligned specimen consensus sequences.
alignment display	A table of IUB codes, space characters, blanks, and dots showing how the sequences within a project are aligned.
alignment score	The number of mismatches between the aligned reference and the aligned consensus sequence for a given specimen.
allele	An alternative form of a genetic locus.
analysis	The complete procedure that SeqScape® Software performs in a batch-wise manner on sample data.
analysis defaults	The default analysis settings that are stored in a project template.
analysis protocol	The default settings (basecalling, mixed base identification, clear range and trimming, and filtering) that govern sample analysis.

analysis settings	The parameters that govern the basecalling, trimming, filtering, and assembly of the analysis.
assembly	The set of aligned and overlapping sample data that result from the sequencing of one PCR product or clone.
Assembly view	Shows the specimen consensus sequence as well as the aligned sample sequences. Electropherograms and quality values can also be viewed.
basecaller	An algorithm that determines the bases within a sequence during analysis. There are two types of basecallers: KB™ Basecaller basecallers and ABI basecallers.
clear range	The region of sequence that remains after excluding the low-quality or error-prone sequence at both the 5′ and 3′ ends.
comparison	The relationship between the aligned specimen consensi and the reference sequence and the associated reference data.
consensus quality values	See quality values.
consensus caller	The analysis algorithm that is responsible for generating an accurate consensus sequence with per-base quality values.
consensus sequence	The output of the assembly from a biologically related group of samples.
constant position	A position in the library alignment that is identical for every allele in the library. See polymorphic position.
constant position error	A position in a specimen consensus sequence that corresponds to a constant position in the library and that disagrees with the library at that position.
contig	The set of aligned and overlapping sample data that results from the sequencing of one PCR product or clone. Also known as an assembly.
crucial position	A position in a specimen consensus sequence that differs among the set of matches returned after a library search.
display settings	The parameters that govern the display of the data and results.

DyeSet/Primer file	Files that compensate for the mobility differences between the dyes and primers and corrects the color-code changes due to the chemistry used to label the DNA. DyeSet/Primer files are sometimes referred to as mobility files.
export	Moving the data or settings from inside the SeqScape® Software Data Store to outside the SeqScape® Software Data Store either in .ctf or .txt format.
FASTA format	A standard text-based file format for storing one or more sequences.
filtered sample sequence	A sample that has been processed by the basecaller/factura/filter algorithms of the pipeline.
genotype library	A library where the allele sequences are either pure-base or mixed-base sequences. When searching against a genotype library, SeqScape® attempts to find the best matches to the consensus sequence without trying different allele combinations. **Note:** This term is not used by SeqScape® software.
haplotype library	A library where the allele sequences are completely pure-base sequences. When searching against a haplotype library, SeqScape® attempts to combine haplotypes two at a time to find the best genotype match to the consensus sequence.
HIM	Heterozygous insertion/deletion (indel) mutation
IUB/IUPAC	International Union of Biochemistry/International Union of Pure and Applied Biochemistry. More information can be found at: http://www.chem.qmw.ac.uk/iubmb/misc/naseq.html#300.
KB™ basecaller	An algorithm that calculates mixed or pure bases and determines sample quality values.
layout view	Shows the layout of the sample assembly with arrows indicating the placement and orientation of samples.
library match	The name of one allele or the combination of two alleles (depending on the library type) that agree closely with the specimen consensus sequence.

nibbler	The algorithm that sets the clear range for each sample using the clear range settings specified in the analysis settings.
polymorphic position	A position in the library alignment that differs for some alleles in the library. See constant position.
project	A group of related sequences that share the same reference or for which there is no explicit reference.
project summary sequence	A summary of the alignment of the specimen consensi.
project template	Contains an RDG, analysis defaults, display settings, and output settings.
quality values	Measure of certainty of the basecalling and consensus-calling algorithms. Higher values correspond to lower chance of algorithm error. Sample quality values refer to the per-base quality values for a sample, and consensus quality values are per-consensus quality values.
reference	A nucleotide string that: is contiguous, not editable, is stored in the RDG, and that determines the project orientation.
reference associated data	The things that are related or assigned to a particular base or ranges of bases on a reference. There are two types of reference associated data: structural and variant.
Reference Data Group (RDG)	The data that contain the reference and the reference associated data.
Report Manager	A window that contains nine separate reports detailing the success or failure of various portions of the analysis, statistics, mutations, AA variants, and library search information.
sample data	The output of a single lane or capillary on a sequencing instrument that is input into SeqScape® Software.
Sample Manager	A window that displays sample file name, name and specimen; last used basecaller and DyeSet/Primer files; calculated basecalling results (spacing, peak 1, start and stop); and assembly status. The sample name, basecaller, and/or DyeSet/Primer file can be changed here.

sample quality values	See quality values.
sample score	The average of the per-base quality values for the bases in the clear range sequence for the sample.
sample view	A view in the SeqScape® software where you can see attributes of each AB1 file including its annotation, sequence, features, raw data, and electropherogram data.
segment	A contiguous segment of the reference sequence corresponding to a single contiguous DNA sequence.
SeqScape® Manager	The software component that manages the following settings: SeqScape® Software projects, project template, RDG, analysis defaults, and display settings.
space character	A character in an aligned sequence is either an IUB code or space, perhaps shown as a dash (-). A space indicates a deleted base in this string or, equivalently, an inserted base in one of the other aligned strings.
specimen	The container that holds all the sample data as assembled contigs from a biological source or PCR product.
specimen (consensus) quality value	See quality values.
specimen (consensus) score	The average overall of the consensus quality values in the consensus sequence.
specimen consensus sequence	The output of the consensus-calling algorithm from a biologically related group of samples.
specimen report	A concatenated list of all the reported information on a per-specimen basis.
specimen view	A view in SeqScape® software where you can see the consensus sequence and all sample files that were used to create that consensus sequence.
summary sequence	The summary consensus sequence for the entire library alignment. Pure positions in the summary sequence correspond to *constant positions*, and mixed-base positions in the summary sequence correspond to *polymorphic positions*.

Index

Numerics

A

B

C

D

E

F

G

H

Q

R

S

T

U

V

W

Headquarters
5791 Van Allen Way | Carlsbad, CA 92008 USA
Phone +1 760 603 7200 | Toll Free in USA 800 955 6288
For support visit www.lifetechnologies.com/support

www.lifetechnologies.com

life technologies™