

GeneMarker[®] HTS

Mitochondrial NGS Analysis Software

User Manual



Release Information

| | |
|--------------------------------|--------------------------------|
| <i>Document Version Number</i> | <i>GeneMarkerHTS-1.0-UG001</i> |
| <i>Software Version</i> | <i>1.0</i> |
| <i>Document Status</i> | <i>Final</i> |

Copyright

© 2017. SoftGenetics, LLC, All rights reserved.

The information contained herein is proprietary and confidential and is the exclusive property of SoftGenetics. It may not be copied, disclosed, used, distributed, modified, or reproduced, in whole or in part, without the express written permission of SoftGenetics, LLC.

Limit of Liability

SoftGenetics, LLC has used their best effort in preparing this guide. SoftGenetics makes no representations or warranties with respect to the accuracy or completeness of the contents of this guide and specifically disclaims any implied warranties of merchantability or fitness for a particular purpose. Information in this document is subject to change without notice and does not represent a commitment on the part of SoftGenetics or any of its affiliates. The accuracy and completeness of the information contained herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every user.

The software described herein is furnished under a license agreement or a non-disclosure agreement. The software may be copied or used only in accordance with the terms of the agreement. It is against the law to copy the software on any medium except as specifically allowed in the license or the non-disclosure agreement.

Trademarks

The name “*SoftGenetics*,” the SoftGenetics logo, and *GeneMarker HTS* are trademarks or registered trademarks of SoftGenetics, LLC. All other products and company names mentioned herein might be trademarks or registered trademarks of their respective owners.

Customer Support

Customer support is available to organizations that purchase *GeneMarker HTS* and that have an annual support agreement. Contact SoftGenetics at:

SoftGenetics, LLC
100 Oakwood Ave, Suite 350
State College, PA 16803
(814) 237-9340
(888) 791-1270 (US Only)
tech_support@softgenetics.com
www.softgenetics.com

Table of Contents

| | |
|--|-----------|
| Preface | 7 |
| Chapter 1: Getting Started with GeneMarker HTS | 9 |
| GeneMarker HTS System Requirements | 11 |
| Installing GeneMarker HTS..... | 12 |
| Starting GeneMarker HTS | 14 |
| The GeneMarker HTS Main Window | 15 |
| Title bar..... | 15 |
| Application Settings icon..... | 15 |
| Project Launcher options | 16 |
| About button | 16 |
| Online Help button..... | 16 |
| Specifying GeneMarker HTS Settings | 17 |
| To specify GeneMarker HTS settings..... | 17 |
| Chapter 2: GeneMarker HTS Projects | 25 |
| Overview of the New Project Dialog Box | 27 |
| Using a motif..... | 29 |
| Soft clipping based on quality scores | 33 |
| Masking personal health information (PHI) for a subject..... | 33 |
| Working with Projects in GeneMarker HTS | 36 |
| To add a new project | 36 |
| To work with an existing project..... | 40 |
| GeneMarker HTS Project Output Files | 43 |
| Chapter 3: GeneMarker HTS Viewers | 45 |
| GeneMarker HTS Viewer Layout..... | 46 |
| Title bar..... | 47 |
| Main menu | 47 |

| | |
|---|-----------|
| Global View pane | 48 |
| Pane displays | 49 |
| Variants table | 51 |
| Status Bar | 52 |
| GeneMarker HTS Viewer Navigation and Options | 53 |
| GeneMarker HTS Viewer navigation..... | 53 |
| GeneMarker HTS Viewer options | 54 |
| GeneMarker HTS Viewer Project Settings | 59 |
| To specify the Filter Settings for a project..... | 59 |
| To specify the Table Settings for a project..... | 63 |
| GeneMarker HTS Viewer Reports..... | 65 |
| To generate the Coverage Curve report for a project | 65 |
| GeneMarker HTS Comparison Viewer | 66 |
| Sample to Sample Comparison pane | 66 |
| Variant Comparison pane | 68 |
| Glossary | 69 |
| Index | 71 |

Preface

Welcome to the *GeneMarker HTS User's Manual*. The purpose of the *GeneMarker HTS User's Manual* is to answer your questions and guide you through the procedures necessary to use the GeneMarker HTS application efficiently and effectively.

Using the manual

You will find the *GeneMarker HTS User's Manual* easy to use. You can simply look up the topic that you need in the table of contents or the index. Later, in this Preface, you will find a brief discussion of each chapter to further assist you in locating the information that you need.

Special information about the manual

The *GeneMarker HTS User's Manual* has a dual purpose design. It can be distributed electronically and then printed on an as-needed basis, or it can be viewed online in its fully interactive capacity. If you print the document, for best results, it is recommended that you print it on a duplex printer; however, single-sided printing will also work. If you view the document online, a standard set of bookmarks appears in a frame on the left side of the document window for navigation through the document. For better viewing, decrease the size of the bookmark frame and use the magnification box to increase the magnification of the document to your viewing preference.

Conventions used in the manual

The *GeneMarker HTS User's Manual* uses the following conventions:

- Information that can vary in a command—variable information—is indicated by alphanumeric characters enclosed in angle brackets; for example, <Project Name>. Do not type the angle brackets when you specify the variable information.
- A new term, or term that must be emphasized for clarity of procedures, is *italicized*.
- Page numbering is “online friendly.” Pages are numbered from 1 to x, *starting with the cover* and ending on the last page of the index.



Although numbering begins on the cover page, this number is not visible on the cover page or front matter pages. Page numbers are visible beginning with the first page of the table of contents.

Preface

- This manual is intended for both print and online viewing.
 - If information appears in [blue](#), it is a hyperlink. Table of Contents and Index entries are also hyperlinks. Click the hyperlink to advance to the referenced information.

Assumptions for the manual

The *GeneMarker HTS User's Manual* assumes that:

- You are familiar with Windows-based applications and basic Windows functions and navigational elements.
- References to any third party standards or third party software functions were current as of the release of this version of GeneMarker HTS, and might have already changed.

Organization of the manual

In addition to this Preface, the *GeneMarker HTS User's Manual* contains the following chapters:

- [Chapter 1, “Getting Started with GeneMarker HTS,” on page 9](#) details the installation requirements for GeneMarker HTS, installing GeneMarker HTS, and activating your account. It also explains how to launch GeneMarker HTS and provides an overview of the major navigational elements for GeneMarker HTS. It also details how to customize the layout and display of your GeneMarker HTS instance.
- [Chapter 2, “GeneMarker HTS Projects,” on page 25](#) details the New Project dialog box, which provides all the options for adding a new project for mtDNA data analysis in GeneMarker HTS. It also explains how to open and work with existing projects in GeneMarker HTS.
- [Chapter 3, “GeneMarker HTS Viewers,” on page 45](#) details the two viewers that are available in GeneMarker HTS application for viewing and working with the results of your mtDNA alignment projects – the GeneMarker HTS Viewer and the Comparison Viewer. It also details how to generate a Coverage Curve report for the results of an mtDNA alignment.
- The [Glossary on page 69](#) provides detailed definitions of specialized terms for GeneMarker HTS.

Chapter 1

Getting Started with GeneMarker HTS

The GeneMarker HTS software application provides a streamlined workflow for mtDNA data analysis from massively parallel sequencing (MPS) methods. Developed in collaboration with leading forensic laboratories, GeneMarker HTS provides rapid analysis of multiple samples with a unique motif alignment technology. Analysis results include depth of coverage graphic, consensus sequence aligned to reference (IUPAC nomenclature), and read pile-up with depth and direction indicators. This software is ideal for the analysis of data from leading NGS sequencing platforms including Illumina and Ion Torrent. This chapter details the installation requirements for GeneMarker HTS, installing GeneMarker HTS, and activating your account. It also explains how to launch GeneMarker HTS and provides an overview of the major navigational elements for GeneMarker HTS. It also details how to customize the layout and display of your GeneMarker HTS instance

This chapter covers the following topics:

- [“GeneMarker HTS System Requirements” on page 11.](#)
- [“Installing GeneMarker HTS” on page 12.](#)
- [“Starting GeneMarker HTS” on page 14.](#)
- [“The GeneMarker HTS Main Window” on page 15.](#)
- [“Specifying GeneMarker HTS Settings” on page 17.](#)

Chapter 1
Getting Started with GeneMarker HTS

GeneMarker HTS System Requirements

GeneMarker HTS is currently available only for the Windows operating system. You must have Administrator rights for the computer on which you are installing GeneMarker HTS.

| Component | Description |
|------------------|--|
| OS | 64-bit Window 7 or newer |
| Processor | Intel Core i5 3rd generation (or equivalent) or better |
| Memory | 16 GB RAM (Minimum) |
| Hard Drive | 250 GB of free space (SSD recommended) |
| Monitor | 1366 x 768 or higher resolution monitor |

Installing GeneMarker HTS

You can install GeneMarker HTS from a disc that SoftGenetics supplies, or you can download the GeneMarker HTS executable (GeneMarkerHTS_Installer.exe) from SoftGenetics's website (<http://softgenetics.com/downloads.php>). The GeneMarker HTS installation requires the Microsoft Visual C++ 2012 Redistributable. If this redistributable is already installed on the client on which you are installing GeneMarker HTS, then no additional action is required during installation; otherwise, the GeneMarker HTS Installation wizard guides you through the steps that are necessary to install this redistributable. GeneMarker HTS is installed in the following directory, and you cannot select a different location:

C:\Program Files (x86)\SoftGenetics\GeneMarker HTS

Figure 1-1: GeneMarker HTS Installation Wizard, Page 1

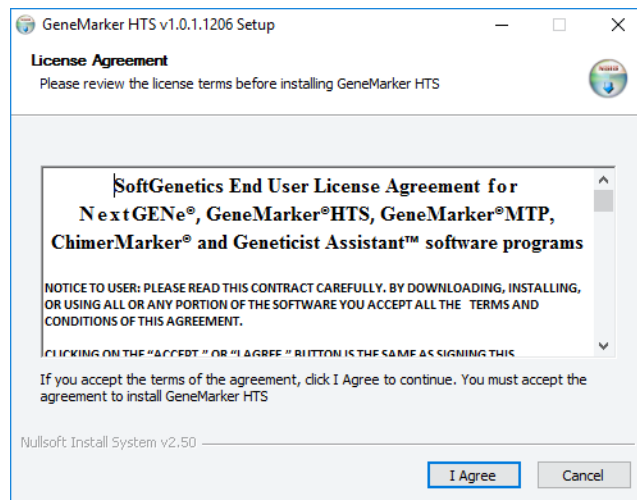


Figure 1-2: GeneMarker HTS Installation Wizard, Page 2 (Optional)

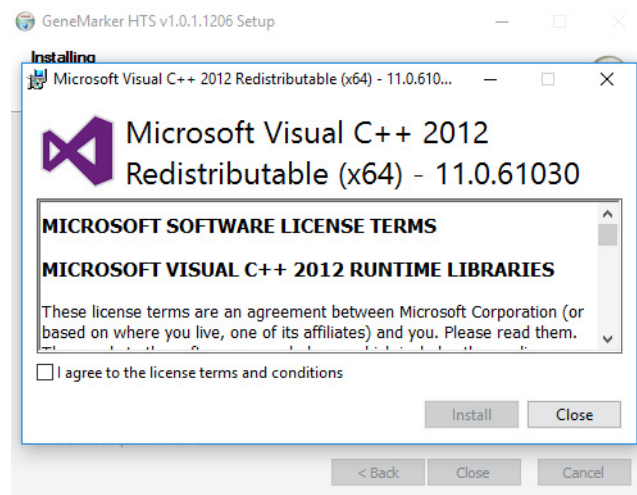


Figure 1-3: GeneMarker HTS Installation Wizard, Page 3 (Optional)

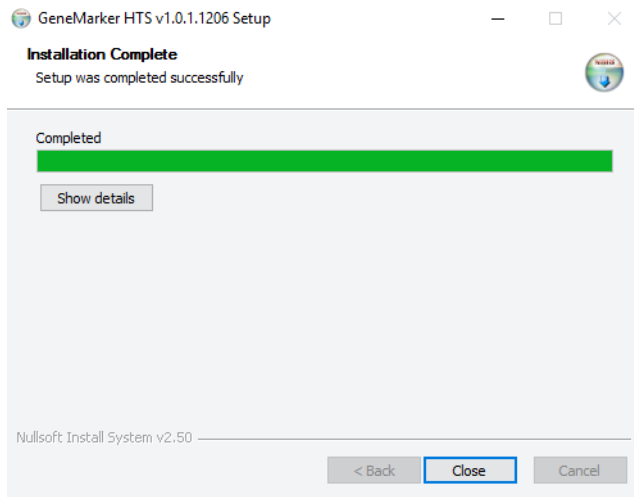
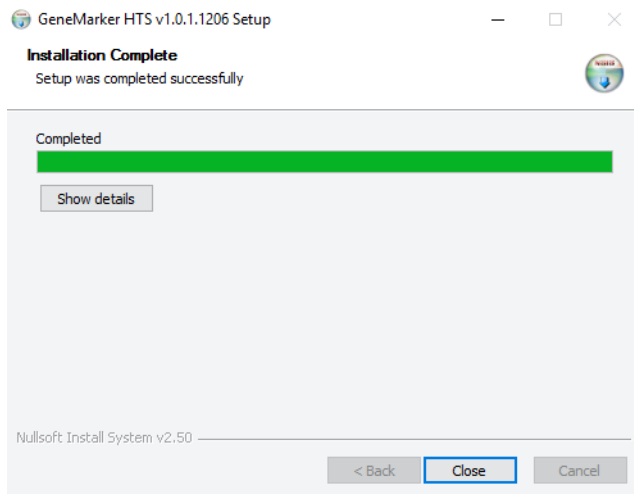


Figure 1-4: GeneMarker HTS Installation Wizard, Page 4



Starting GeneMarker HTS

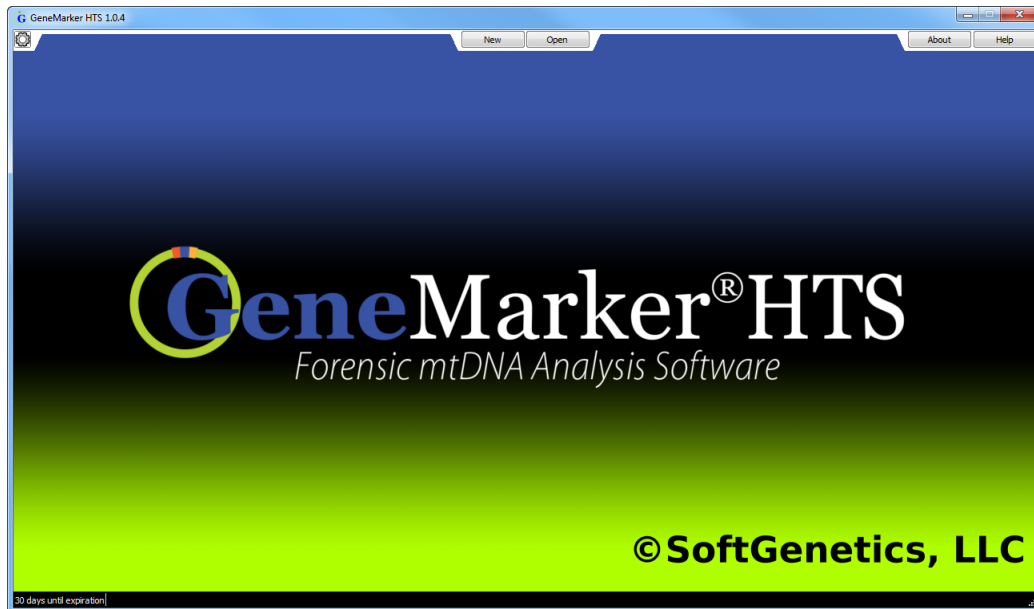
After GeneMarker HTS has been installed on your computer, a shortcut icon for the application is placed on your desktop. An option for GeneMarker HTS is also available from your Start menu. You can double-click the desktop icon to launch GeneMarker HTS, or you can select the option from your Start menu (Start > All Programs > SoftGenetics > GeneMarker HTS).

Figure 1-5: GeneMarker HTS desktop icon



After you start GeneMarker HTS, the GeneMarker HTS main window opens. See “[The GeneMarker HTS Main Window](#)” on page 15.

Figure 1-6: GeneMarker HTS main window



The GeneMarker HTS Main Window

After you start GeneMarker HTS, the GeneMarker HTS main window opens

Figure 1-7: GeneMarker HTS main window



The GeneMarker HTS main window is your starting point for the GeneMarker HTS application. The GeneMarker HTS main window has five major components—the [title bar](#), the [Application Settings](#) icon, the [Project Launcher](#) options, the [About](#) button, and the [Online Help](#) button.

Title bar


The name “GeneMarker HTS” is displayed in the title bar at the top of the GeneMarker HTS main window.

Figure 1-8: Title bar



The version of GeneMarker HTS that you are running is not displayed in the title bar. In the top right corner of the main window underneath the title bar, click [About](#) to open the About GeneMarker HTS dialog box, which displays information about your GeneMarker HTS installation, including the version.

Application Settings icon

The Application Settings icon is the icon that is displayed in the left corner of the main window underneath the title bar . Click the icon to open the Application Settings dialog box, which provides all the options for specifying the settings to customize your

Chapter 1 Getting Started with GeneMarker HTS

GeneMarker HTS instance, including Colors, Performance, Personal Health Information (PHI), Server, and User Management. See [“Specifying GeneMarker HTS Settings”](#) on page 17.

Project Launcher options

The Project Launcher options—New and Open—are displayed in the center of the main window underneath the title bar. Click New to open the New Project dialog box, and set up a new analysis project in GeneMarker HTS. Click Open to open the Open Project dialog box and browse to and select an existing GeneMarker HTS with which to work. See [Chapter 2, “GeneMarker HTS Projects,”](#) on page 25.

About button

The About button is displayed in the top right corner of the main window underneath the title bar. Click to the button open the About GeneMarker HTS dialog box, which displays information about your GeneMarker HTS installation, including the version.

Figure 1-9: About GeneMarker HTS dialog box



Online Help button

The Help button is displayed in the top right corner of the main window underneath the title bar. Click the button to open the online help for GeneMarker HTS.


Specifying GeneMarker HTS Settings

The Settings dialog box provides all the options for [specifying](#) the settings to customize the your GeneMarker HTS instance. When you make any change to any GeneMarker HTS setting, the change is persistent within the GeneMarker HTS instance and across GeneMarker HTS instances. For some settings, an option is available to reset all values to their default values in a single step.

To specify GeneMarker HTS settings



Typically, you do not adjust the default values for the Color settings and/or Performance settings until after you opened one or more projects in a viewer; however, for efficiency's sake, all five settings are discussed here.

1. On the GeneMarker HTS main window, click the Application Settings icon .

The Application Settings dialog box opens. The Colors tab is the open tab.

2. Continue to one of the following:
 - [“To specify the colors settings”](#) below.
 - [“To specify the performance settings”](#) on page 22.



The Server tab and the User Management tab are placeholders for planned future functionality and are not currently relevant when discussing GeneMarker HTS settings.

To specify the colors settings

You specify the colors that are to be used for background/highlighting and the nucleotides in the Panes display (Reference pane, Consensus Pane, and Pile-Up pane) for the GeneMarker HTS Viewer on the Colors tab of the Application Settings dialog box. You can also specify the highlighting colors that are to be used in the two panes (Sample to Sample Comparison and/or Variant Comparison) of the Comparison Viewer.



For detailed information about the viewers that are available in GeneMarker HTS, see [Chapter 3, “GeneMarker HTS Viewers,”](#) on page 45.

1. On the Application Settings dialog box, open the Colors tab.

The tab is divided in to four panes, which, from top to bottom, are the following:

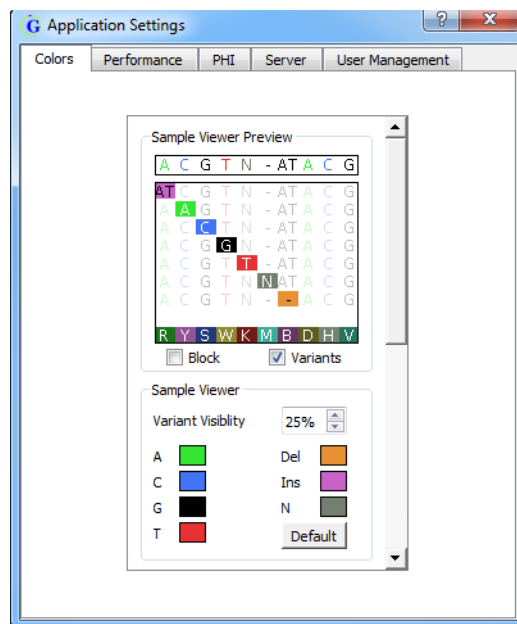
- The Preview pane shows what the display currently looks like for the panes in the GeneMarker HTS Viewer.
- The Global Viewer pane shows the currently selected colors for the Global Viewer pane display in the GeneMarker HTS viewer

- The Sample Viewer pane shows the currently selected colors for the Panes display in the GeneMarker HTS viewer.
- The Comparison Viewer pane shows the currently selected colors for the two panes (Sample to Sample Comparison and Variant Comparison) in the Comparison Viewer.



Use the scroll bar on the right side of the display to scroll through and view all the available panes.

Figure 1-10: Application Settings dialog box, Colors tab



2. Continue to either of the following as appropriate:
 - “To edit the Panes display colors” below.
 - “To edit the Comparison Viewer highlighting colors” on page 21.

To edit the Panes display colors

If you edit the Panes display colors, the Preview pane is dynamically updated with your settings. Note the following when editing the Pane display colors:

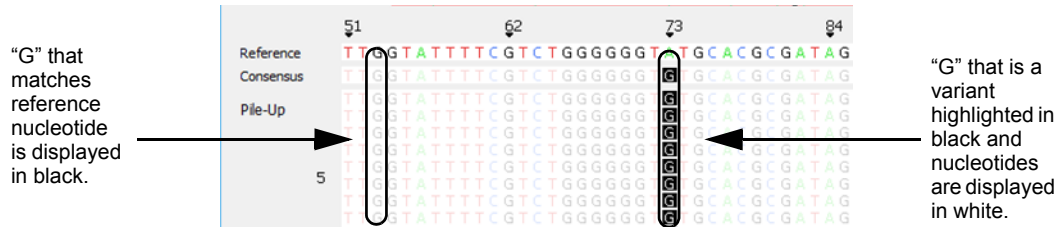
- If Blocks is selected, then all nucleotides are displayed in white on a colored background.

Figure 1-11: Blocks selected for Pane displays



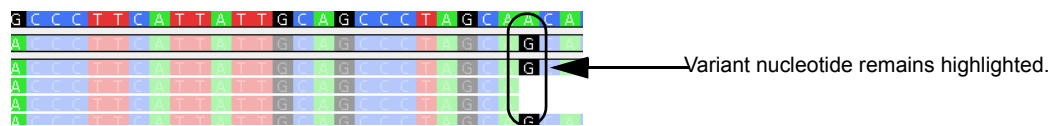
- If Variants is selected (the default value), then all nucleotides other than variant nucleotides are displayed in color on a white background. Conversely, variant nucleotides are displayed on a colored background in white (referred to as a highlighted variant).

Figure 1-12: Variants selected for Pane displays



You can select both Variants and Blocks for the display. In this case, all nucleotides are still displayed in white on a colored background; however, all nucleotides other than variant nucleotides are dimmed and all variant nucleotides are highlighted to visually differentiate them from non-variant nucleotides.

Figure 1-13: Variants and Blocks selected for pane display in the GeneMarker HTS Viewer



- *Variant Visibility* determines how dimmed non-variant nucleotides are when compared to the variant nucleotides when *both* Variants and Blocks are selected for the pane display in the Viewer. The default value is 40%. The pane also displays the various colors that are to be used in the Pane displays where the *default* values are the following:
 - For non-variant nucleotides and SNPs: If Variants is selected, then Green, Blue, Black, and Red indicate the colors for the nucleotides A, C, G, and T, respectively. If Blocks is selected, then these indicate the colors for the nucleotide backgrounds.
 - For Deletions: No nucleotide is displayed. Instead, a dash highlighted in burnt orange is displayed.
 - For Insertions: If Variants is selected, then the nucleotide is displayed in purple. If Blocks is selected, then the nucleotide background is purple.
 - If coverage areas fall below the Total Coverage value that is specified for the Filter Settings: If Variants is selected, then a gray “N” is displayed for the consensus sequence in these regions. If Blocks is selected, then the nucleotide background is displayed in gray and “N” is displayed in white.

See [Figure 1-14](#) and [Figure 1-15](#) on page 20.

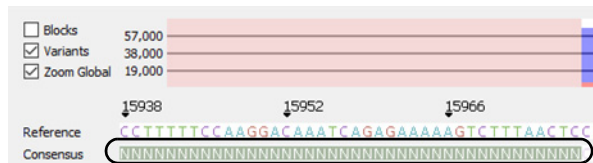


For detailed information about Filter Settings, see [“GeneMarker HTS Viewer Project Settings”](#) on page 59.

Figure 1-14: Default colors for the Pane displays in the GeneMarker HTS Viewer



Figure 1-15: “N” indicating that coverage areas fall below the Total Coverage Value that is specified in the Filter Settings



1. Do any or all of the following as needed to change colors and/or the variant visibility:

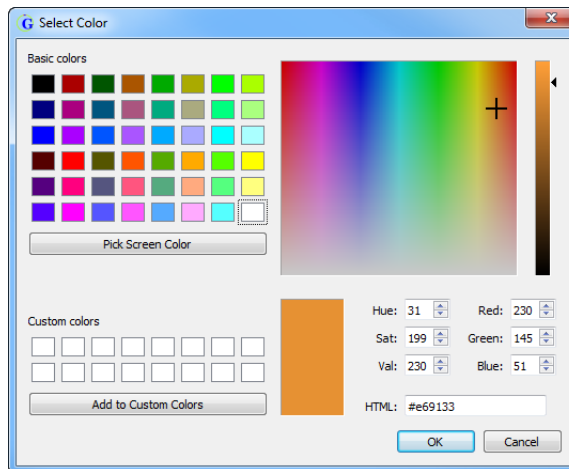


If you change colors and/or the Variant Visibility value, you might need to refresh the display to view your changes. You can use your mouse scroll wheel, or you can zoom in/zoom out to refresh the display.

- Change one or more colors for the nucleotides and/or backgrounds. To do so, double-click the color swatch that is displayed for a nucleotide or variant type to open the Select Color dialog box, select a different color, and then click OK.

The Select Color dialog box closes and you return to the Colors tab of the Application Settings dialog box. The newly selected color is displayed for the selected nucleotide or variant type.

Figure 1-16: Select Color dialog box



- Adjust the value for Variant Visibility. To make the non-variant nucleotides more visible in the pane display when both Blocks and Variants are selected, increase the value. Conversely, to make the non-variant nucleotides less visible in the pane display when both Blocks and Variants are selected, decrease the value. You can manually enter a new value, or you can use the Up/Down arrows as needed.



To reset all Color settings and/or the Variant Visibility to their default values in a single step, click Default.

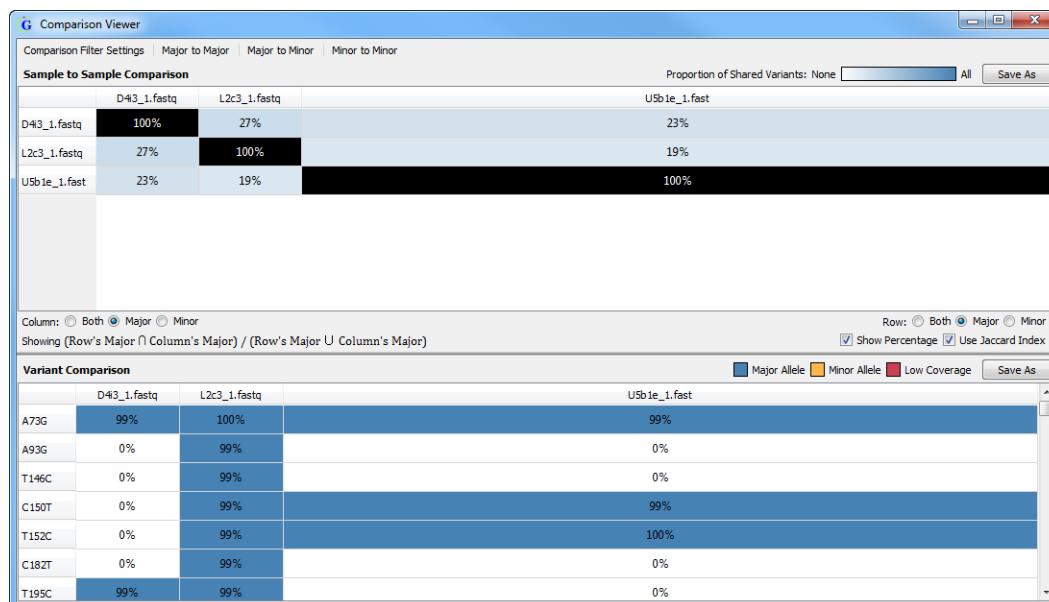
- If you are done specifying your settings, then click OK to close the Application Settings dialog box and return to the GeneMarker HTS main window; otherwise, continue to add and/or edit the information on this tab or any of the other tabs as needed.

To edit the Comparison Viewer highlighting colors

The Comparison Viewer is divided into two panes: the Sample to Sample Comparison (top) pane and the Variant Comparison (bottom) pane.

- In the Sample to Sample Comparison (top) pane, diagonal cells that compare a sample to itself are highlighted by default in black. All other cells in the correlation table compare one sample to another, and by default, are highlighted in a blue color gradient from 0 to 100%, where 0% cells are blank/white.
- In the Variant Comparison (bottom) pane, variants that can be classified as a major allele or as a minor allele are highlighted, by default, in blue or gold, respectively. If a variant was determined to be a Low Coverage variant, then by default, the variant is highlighted in red.

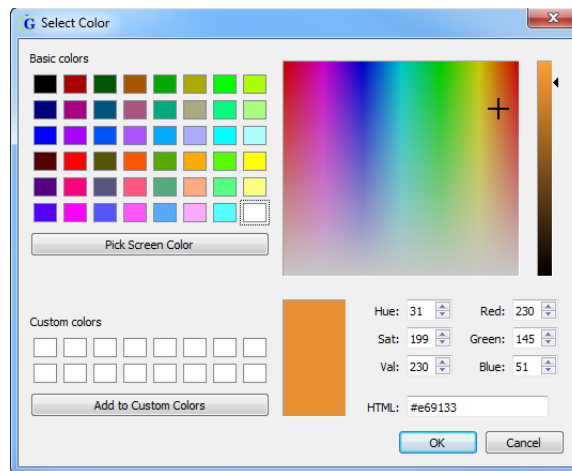
Figure 1-17: GeneMarker HTS Comparison Viewer



1. Change one or more highlighting colors as needed. To do so, double-click the color swatch that is displayed for a nucleotide or variant type to open the Select Color dialog box, select a different color, and then click OK.

The Select Color dialog box closes and you return to the Colors tab of the Application Settings dialog box. The newly selected color is displayed for the selected nucleotide or variant type.

Figure 1-18: Select Color dialog box



To reset all Color settings to their default values in a single step, click Default.

2. If you are done specifying your settings, then click OK to close the Application Settings dialog box and return to the GeneMarker HTS main window; otherwise, continue to add and/or edit the information on this tab or any of the other tabs as needed.

To specify the performance settings

Performance settings determine the speed at which a sample file opens in the GeneMarker HTS Viewer as well as the speed at which you can navigate the viewer display.

1. On the Application Settings dialog box, open the Performance tab.

The tab is divided into two panes. The RAM Settings (top) pane shows the current RAM use on a sliding scale, where:

- More RAM use allows for fewer GeneMarker HTS Viewers to be open at the same time, but navigation in these viewers is much “smoother.”
- Less RAM use allows for multiple GeneMarker HTS Viewers to be open at the same time.

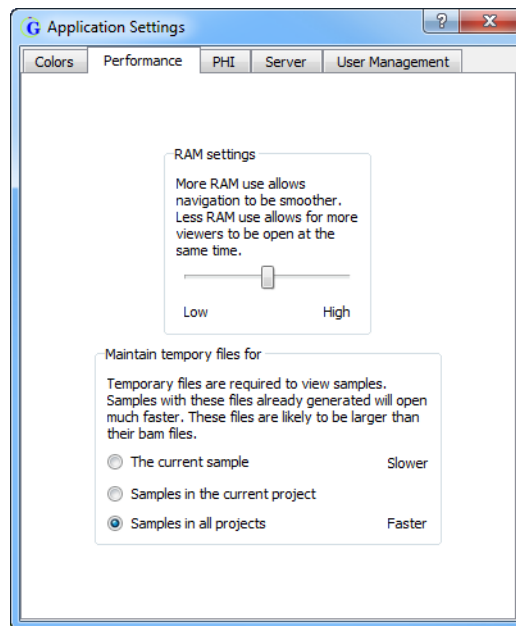
The Maintain temporary files pane shows the projects for which temporary files are maintained. Although these Temporary files might be larger than their BAM

counterparts, they are required to open sample files in a project. Sample files for which these files are present open faster in the Viewer than sample files for which they are not present. By default, Maintain temporary files is set to Samples in all projects.



Although you can manually delete temporary files, do not do so while an analysis is running.

Figure 1-19: Application Settings dialog box, Performance tab



2. Optionally, do one or both of the following:
 - Use the sliding scale to adjust the amount of RAM that your GeneMarker HTS instance uses.
 - Select a different value for the Maintain temporary files pane.
3. If you are done specifying your settings, then click OK to close the Application Settings dialog box and return to the GeneMarker HTS main window; otherwise, continue to add and/or edit the information on any of the other tabs as needed.

Chapter 1
Getting Started with GeneMarker HTS

Chapter 2

GeneMarker HTS Projects

GeneMarker HTS is ideal for the analysis of mtDNA data from leading NGS sequencing platforms including Illumina and Ion Torrent. The New Project dialog box provides all the options for adding a new project for mtDNA data analysis. You can also open and work with existing projects in GeneMarker HTS.

This chapter covers the following topics:

- [“Overview of the New Project Dialog Box” on page 27.](#)
- [“Working with Projects in GeneMarker HTS” on page 36.](#)
- [“GeneMarker HTS Project Output Files” on page 43.](#)

Chapter 2
GeneMarker HTS Projects

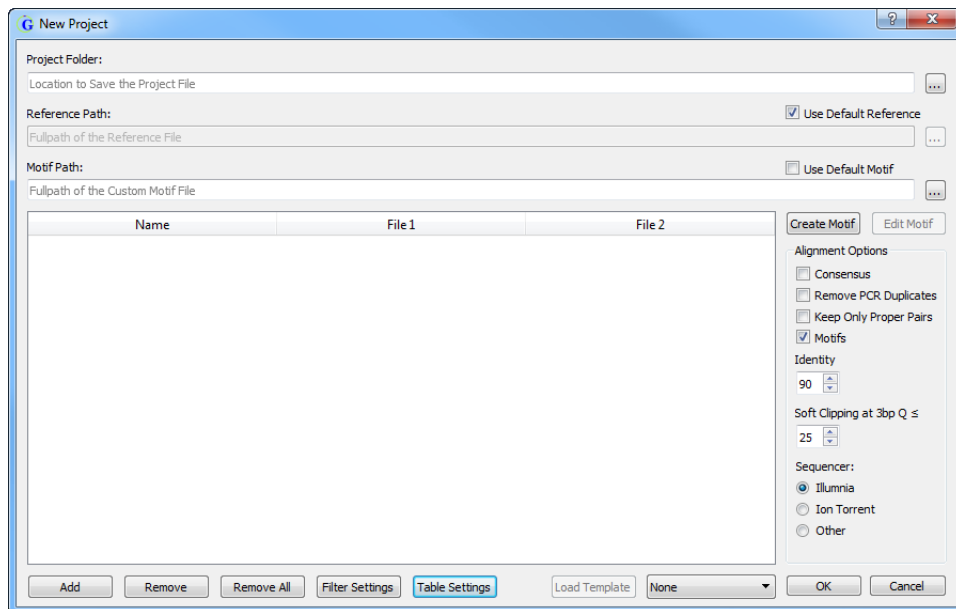
Overview of the New Project Dialog Box



For detailed information about working with an existing GeneMarker HTS project see, *“Working with Projects in GeneMarker HTS” on page 36.*

You use the New Project dialog box to set up a new project for analyzing your mtDNA sequencing data. You select the data that is to be analyzed for a project in the dialog box, and you also specify the alignment settings for the project in the dialog box. To open the New Project dialog box, at the top of the GeneMarker HTS main window, click New.

Figure 2-1: New Project dialog box



The following table describes the settings on the New Project dialog box the first time that the dialog box is opened. Any selections that you make are persistent within your GeneMarker HTS instance and across instances.

Table 2-1: New Project dialog box settings

| Setting | Description |
|--------------------------------------|--|
| Project Folder | Location in which to save the project folder. The default location is C:\Program Files\SoftGenetics\GeneMarkerHTS. |
| Use Default Reference Reference Path | Leave Use Default Reference selected to use the reference (the revised Cambridge reference sequence, or rCRS) that is supplied with GeneMarker HTS. To use a different reference, clear Use Default Reference, and then click the Browse button to open the Open Reference dialog box and browse to and select the appropriate reference. If you select a different reference, then it must in either FASTA or GBK format. |

Table 2-1: New Project dialog box settings (Continued)

| Setting | Description |
|---|---|
| <p>Note: Motif options are available only if Motifs is selected for an Alignment option. (See “Alignment Options” below.) If Motifs is not selected as an alignment option, then all Motif options are unavailable. SoftGenetics recommends the use of a motif for mtDNA analysis.</p> | |
| Use Default Motif Motif Path | You use a motif (a .txt file) to adjust the alignment to meet the expectations of your analysis. See “Using a motif” on page 29. Select Use Default Motif to the motif data that is supplied with GeneMarker HTS. To use a different motif, clear Use Default Motif and then next to the Motif Path field, click the Browse button that to open the Open Motif dialog box and browse to and select the appropriate custom motif. After you select a motif, the full directory path, including the filename, for the selected custom motif is displayed in the Motif Path field. |
| Create Motif | Opens the Motif Editor dialog box, which provides the options for creating the custom motif that is to be used for your mtDNA analysis. If you select this option, then the built-in (default) motif is used as the starting point for creating the custom motif. See “To create or edit a custom motif” on page 30. |
| Edit Motif | Available only after you select a custom motif that is to be used for your mtDNA analysis. Opens the Motif Editor dialog box, which makes the selected custom motif available for editing. See “To create or edit a custom motif” on page 30. |
| Alignment Options | |
| Consensus | Do local indel realignment based on the consensus sequence. After variant calls are made based on the initial alignment, any reads that overlap indels but do not contain the indels are adjusted to include the indels if the new alignment is at least as good as the initial alignment. |
| Remove PCR Duplicates | On or Off. Reads are identified as PCR duplicates based on their start positions. If you are using paired-end data, then each pair of reads is considered together. Select to remove PCR duplicates, or leave blank and the PCR duplicates are not removed. Note: If you are using targeted sequencing, then you typically do not select this value. If you are using whole genome sequencing, then you typically do select this value. |
| Keep Only Proper Pairs | Ignore any reads that were not marked as being part of a proper pair, where a proper pair is defined as the following: <ul style="list-style-type: none"> • Both reads from the pair must be aligned and the alignment must be on opposite strands. • The 5' end of the reverse read must be to the right of the 5' end of the forward read. |
| Motifs | If you elect to use motif alignment instead of reference alignment, then the alignments of reads that cross motif regions with expected sequences are adjusted to match the list of variants; otherwise, if you clear this option, then alignment is focused on minimizing the number of mismatches between the reference and a sample read, or a “best fit” alignment. See “Using a motif” on page 29. |
| Identify | Reads that are less similar to the reference than the indicated are not aligned. |

Table 2-1: New Project dialog box settings (Continued)

| Setting | Description |
|---|--|
| Soft Clipping at 3bp Q ≤ | When the base call quality is less than or equal to the indicated quality score, then the 3' ends of reads are trimmed accordingly. Note: If mismatched bases are found near the end of the alignment, then soft clipping is also carried out on the 3' ends of reads. See “Soft clipping based on quality scores” on page 33. |
| Note: GeneMarker HTS automatically adjusts the default values for both Identity and Soft Clipping based on the Sequencer that you select. The default settings are based on using an Illumina or Ion Torrent sequencer. If you change the value for either one or both of the settings, then GeneMarker HTS automatically selects “Other” for Sequencer. | |
| Sequencer - You can select only one value. | |
| Illumina | Selected by default. |
| Ion Torrent | |
| Other | |
| Filter Settings | Opens the Filter Settings dialog box, which contains options for customizing the display of the GeneMarker HTS Viewer. |
| PHI Settings | Opens the PHI Settings dialog box, which contains the settings for specifying which personal health information to mask for a subject. You can specify which PHI to mask for a subject only before carrying out the corresponding mtDNA alignment project. See “Masking personal health information (PHI) for a subject” on page 33. |
| Table Settings | Opens the Table Settings dialog box, which displays a list of all the data fields that are to be displayed in the Variants table in the GeneMarker HTS Viewer. |
| Note: SoftGenetics has set default values that it has determined to be the most appropriate for mtDNA analysis for both the Filter Settings and Table Settings, but you can always edit any or all of these values. See “GeneMarker HTS Viewer Project Settings” on page 59. | |

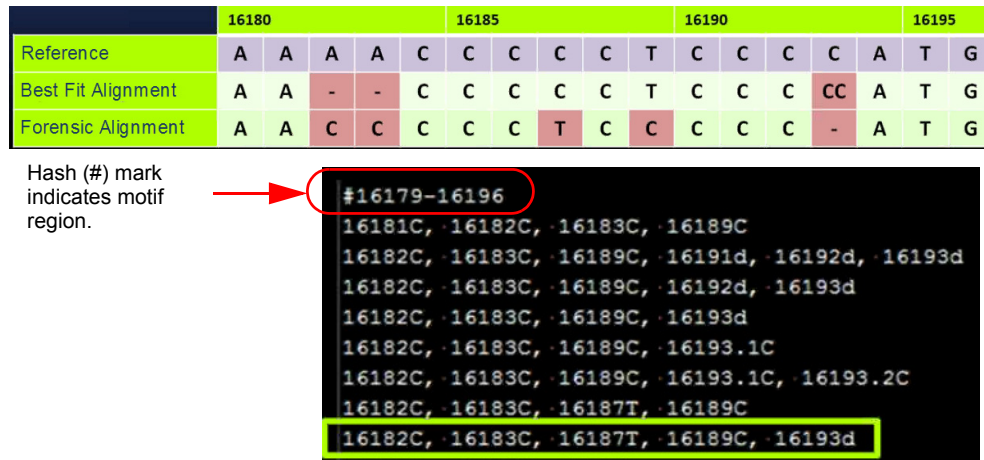
Using a motif

Traditionally, alignment is focused on minimizing the number of mismatches between the reference and a sample read, or a “best fit” alignment; however, because forensic analysis has established conventions for the position of many common variants, this traditional approach to alignment is not always preferred when analyzing mtDNA. As a result, GeneMarker HTS provides option of using a motif to better meet your analysis requirements. A *motif* is a list of variant calls within a specified region that are translated into an expected sequence. You define this list of variant calls in a *motif file*, which is a text (.txt) file that has the following format:

- The first line for each motif definition in the file is preceded with a hash tag/pound symbol (#) and this line identifies the start and end points for the region.
- The variant calls in the motif file are listed with standard forensic nomenclature, for example, Deletion: 16191d, Insertion: 16193.1C, and SNP: 1184A.

See [Figure 2-2 on page 30.](#)

Figure 2-2: Example of Best Fit alignment compared to Forensic alignment



When you are creating a motif file, note the following:

- Spacing of the variants is not absolute, but you must use commas to separate the variant calls.
- Regions cannot overlap.
- No two motifs in the same region can resolve to the same sequence.



You can use the Check For Errors option that is available as part of the built-in Motif Editor to confirm that no two motifs in the same region resolve to the same sequence.

If you elect to use motif alignment instead of reference alignment, then the alignments of reads that cross motif regions with expected sequences are adjusted to match the list of variants. Reads that start or end in motif regions are soft-clipped. You can use the built-in motif that is supplied with GeneMarker HTS, or you can the Motif Editor to [create and/or edit](#) your own custom motifs.

To create or edit a custom motif



Although you can manually edit a motif file, SoftGenetics strongly recommends that you use the built-in Motif Editor instead.

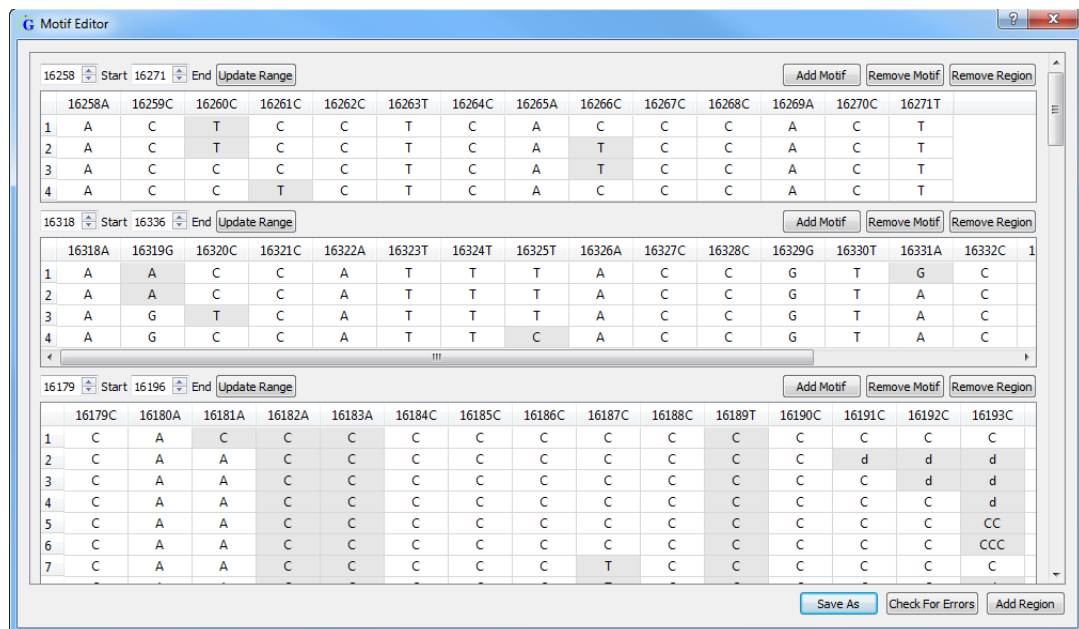
1. On the New Project dialog box, do one of the following:
 - To create a custom motif, under Alignment Options, select Motifs, and then click Create Motif.

The Motif Editor dialog box opens. The built-in (default) motif is loaded in the dialog box. Expected variant calls are highlighted in gray. See [Figure 2-3 on page 31](#).

- To edit an existing custom motif, do the following:
 - i Next to the Motif Path field, click the Browse button to open the Open Motif dialog box, and then browse to and select the appropriate custom motif.
 - ii. Click Edit Motif.

The Motif Editor dialog box opens. The selected custom motif is loaded in the dialog box. Expected variant calls are highlighted in gray.

Figure 2-3: Motif Editor dialog box



2. For *each* motif region, do any or all of the following as needed:

| Option | Description |
|-------------------------|--|
| To edit the motif range | Adjust the Start and/or End values for a motif region (you can manually edit the values, or you can use the Up Arrow/Down Arrow buttons to adjust the value one base at a time), and then click Update Range. Note: Because motifs require exact matches, shorter motifs are better for alignment. |
| Add Motif | Adds a new motif as the last in the list of displayed motifs for the region. The default sequence for a new motif is always set to the reference sequence. |
| Remove Motif | Select a motif, and then click to remove the selected motif from the motif region. |
| Remove Region | Click to remove a region entirely from the motif. |

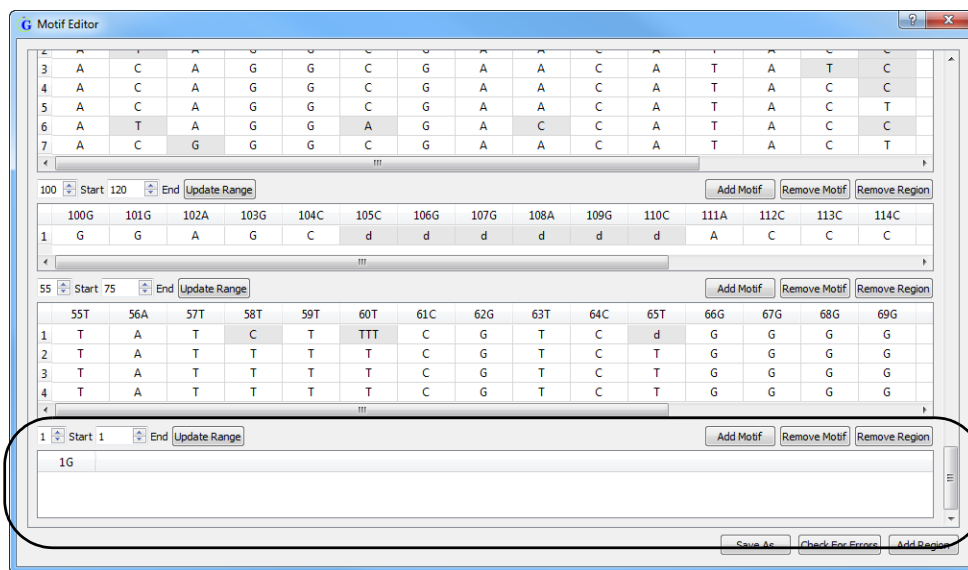
3. If you done with customizing the motif, then continue to [Step 4](#); otherwise, optionally, to add one or more regions to the motif, do the following:
 - a. Click Add Region.

A blank new region placeholder is displayed at the bottom of the Motif Editor dialog box.



Use the scroll bar on the right side of the dialog box to scroll to the bottom of the dialog box and view the placeholder region.

Figure 2-4: Motif Editor dialog box with a new region placeholder



- b. Specify the start and end positions for the new region, and then click Update Range.
 - c. Create the needed motifs for the new range. (See [Step 2](#).)
4. Optionally, to check for errors before saving the customized motif, click Check For Errors.

If errors exist in the customized motif, then a Motif File Errors dialog box that details the errors in the motif file opens; otherwise, the Motif File Errors dialog box displays the message “No errors.”

5. Click OK to close the Motif File Errors dialog box, and if the file did not contain any errors continue, to [Step 6](#); otherwise, iteratively correct the file, and then click Check For Errors until you receive the message “No Errors,” and then continue to [Step 6](#).
6. To save the new motif file, click Save As.

The Save Motif dialog box opens. The file format is set to .txt and you cannot change this.

7. Specify the name for the motif file, and then browse to and select the location in which to save the file, and then click Save.

The Save Motif dialog box closes. The Motif Editor dialog box remains open.

8. Click “X” to close the Motif Editor dialog box.

You return to the New Project dialog box. The full path to the customized motif file, including the file name, is displayed in the Motif Path field. The motif can now be used for mtDNA analysis.

Soft clipping based on quality scores

When the base call quality is less than or equal to the quality score that is specified in the Alignment Settings for a project, then the 3' ends of reads are soft-clipped accordingly. If mismatched bases are found near the end of the alignment, then soft clipping is also carried out on the 3' ends of reads according to the following:

1. Starting at the end of the alignment sequence, move towards the middle of the alignment sequence, with a quality score for each nucleotide in the sequence calculated as follows.
 - Add 1 for a matched base; otherwise, subtract 3 for a mismatched base. A quality score < -6 is not allowed.
2. Continue calculating a quality score for each base in the alignment sequence until a nucleotide with a quality score of 6 is found.
3. Move back from this position towards the end of the read until a mismatch is found.
4. Soft clip from this mismatch through the end of the read.

For example:

- The alignment results in a CIGAR string of $100=2X10=1X3=$.
- A score is calculated going back from the end: 0, 1, 2, 3, 0, 1, 2, 3, 4, 5, 6.
- Moving back to the end, soft clipping is started at the first mismatch through the end, resulting in four bases total being soft-clipped: $100=2X10=4S$.

Masking personal health information (PHI) for a subject

When carrying out a forensic analysis, a lab technician must remain aware of the potential of exposing personal health information (PHI) about a subject that is not relevant to the investigation. For example, the alignment results could reveal that a subject is a carrier for an inherited disease such as Leigh's syndrome. To prevent this, you can hide regions from the alignment that could reveal PHI. You can elect to load default hidden regions, which hides the PHI regions that MITOMAP has identified, and/or hide the PHI regions that you explicitly define in a BED file.



For detailed information about the format for a BED file, see “BED file” on page 69.



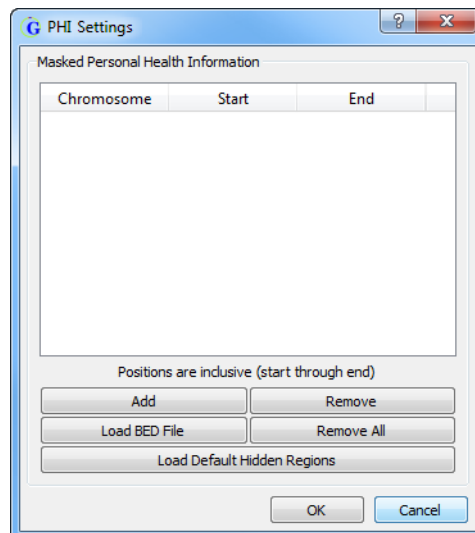
You can specify which personal health information to mask for a subject only before carrying out the corresponding mtDNA alignment project.

For a default or custom hidden region, the following is applicable:

- Nucleotides in the region are not displayed in the Pile-Up pane in the GeneMarker HTS Viewer.
 - The CIGAR string is hidden in the tooltip for a read in the Pile-Up pane.
 - You cannot copy sequences for reads that cross the region.
 - Allele counts are hidden from all reports.
 - No variants are called within the region.
1. On the New Project dialog box, click PHI Settings.

The PHI Settings dialog box opens.

Figure 2-5: PHI Settings dialog box



2. Do one or both of the following:

- To hide the PHI regions that MITOMAP has identified, click Load Default Hidden Regions.



For a list of the MITOMAP locations that are the built in to PHI, see the following:

- <http://www.mitomap.org/foswiki/bin/view/MITOMAP/MutationsRNA>
 - <http://www.mitomap.org/foswiki/bin/view/MITOMAP/MutationsCodingControl>
- To hide the PHI regions that you have identified, click Load BED File to open the Open BED File dialog box, and then browse to and select the appropriate BED file.

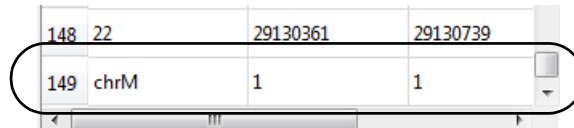
3. Optionally, after you have loaded the default and/or custom regions that are to be hidden, do any or all of the following as needed:
 - To manually add a region, click Add.



You can manually add as many regions as needed.

A new region is added as the last region in the list of displayed regions. By default, the chromosome is set to M, and the Start and End positions are both set to “1,” but you can edit any or all of these values as needed.

Figure 2-6: Manually added PHI region



| | | | |
|-----|------|----------|----------|
| 148 | 22 | 29130361 | 29130739 |
| 149 | chrM | 1 | 1 |

- To remove a region, select the region, and then click Remove.



You can remove only a single region at a time. No confirmation message is displayed before you remove a region, so be cautious when doing so.

- To remove all regions in a single step, click Remove All.
4. Click OK.

The PHI Settings dialog box closes. The selected regions will be hidden from alignment.

Working with Projects in GeneMarker HTS

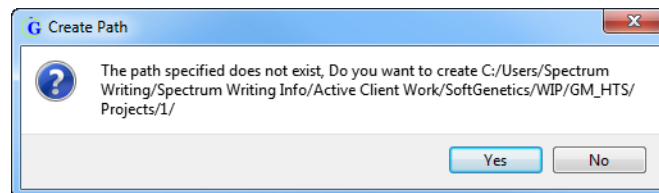
The New Project dialog box provides all the options for [adding](#) a new project for mtDNA data analysis in GeneMarker HTS. You can also open and [work](#) with existing projects in GeneMarker HTS.

To add a new project

Before you add a new project in GeneMarker HTS, it is important to note the following:

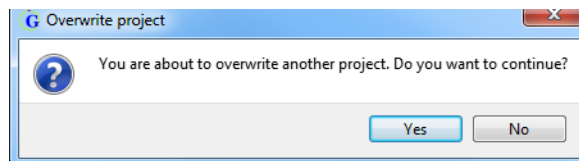
- If the folder that you specify for saving the project does not exist, then GeneMarker HTS prompts you to create the folder before the project is added.

Figure 2-7: Create Path prompt



- Every GeneMarker HTS project is named project. pjf. As a result, if you specify a folder in the Project Folder field in which a project has already been saved, then GeneMarker HTS warns you that you are about to overwrite another project and asks you if you want to continue. *To prevent one GeneMarker HTS project from overwriting another, you should always save a project in its own unique folder.*

Figure 2-8: Overwrite project prompt



- The default reference that is supplied in GeneMarker HTS is the revised Cambridge Reference Sequence (rCRS). If you select a different reference, then it must be in either FASTA or GBK format.
- When you add a new project in GeneMarker HTS, the sample files can be in one of two formats – compressed FASTQ files (fastq.gz) or uncompressed FASTQ files (.fastq).
- The sample file names for a project are automatically generated from the names of the loaded files, but you can always edit a sample file name.
- GeneMarker HTS automatically groups paired reads in a sample, but you can split a group of paired reads into individual samples.
- GeneMarker HTS automatically adjusts the default values for both Identity and Soft Clipping based on the Sequencer that you select. If you change the value for either one or

both settings, then GeneMarker HTS automatically select “Other” for Sequencer, and you cannot select a different value.

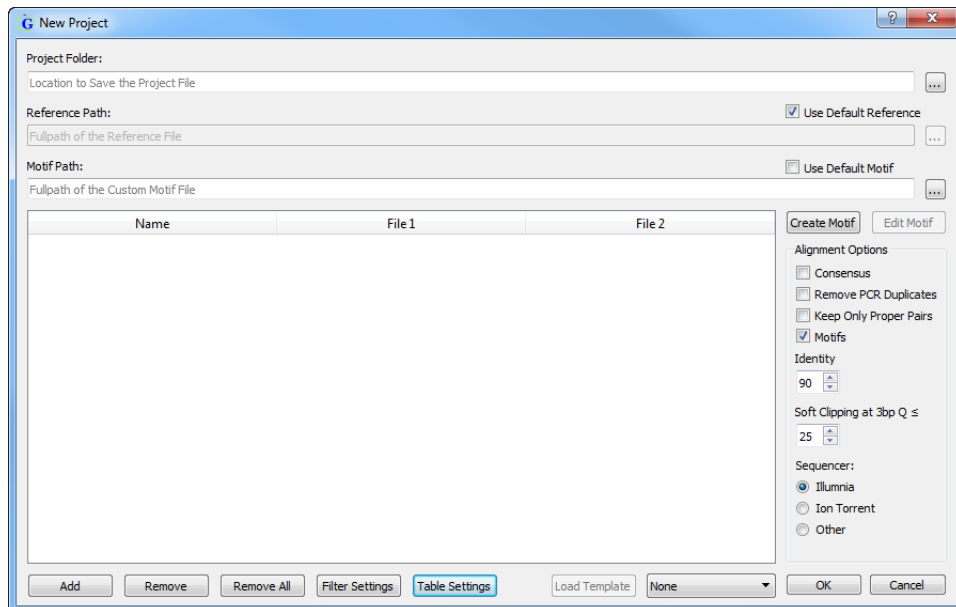
- Filter settings and Table Settings control the criteria for variant calling and for displaying project information in the Variants table in the GeneMarker HTS Viewer. You can adjust Filter Settings and Table Settings before or after alignment of the samples in a project.
- PHI Settings control which personal health information to mask for a subject. You can adjust PHI Settings only *before* the alignment of the samples in a project.

To add a new project in GeneMarker HTS, do the following:

1. On the GeneMarker HTS main window, click New.

The New Project dialog box opens.

Figure 2-9: New Project dialog box



2. Do one of the following to specify the location in which to save the project:
 - Next to the Project Folder field, click the Browse button to open the Project Save Location dialog box, and then browse to and select where the project (*.pjt) file is to be stored.
 - In the Project Folder field, manually enter the full path to the folder in which the project is to be saved.



If the specified folder does not exist, then GeneMarker HTS will prompt you to create the path before the project is created. See [Step 10](#).

3. If the project is to use the default reference that is supplied with GeneMarker HTS, then leave Use Default Reference selected; otherwise, next to the Reference Filename field, click the Browse button to open the Open Reference dialog box, and then browse to and select the reference that the project is to use.



The default reference is the revised Cambridge reference sequence (rCRS). If you select a different reference, then it must be in either FASTA or GBK format.

4. If the project is to use the built-in motif that is supplied with GeneMarker HTS, then select Use Default Motif; otherwise, do one of the following:
 - Select Create Motif, and then create a custom motif.
 - Next to the Motif Path field, click the Browse button to open the Open Motif dialog box and browse to and select an existing motif.



If needed, you can edit the existing motif before using it in the project. For information about creating and editing a motif, see [“To create or edit a custom motif”](#) on page 30.

5. Specify the alignment options for the project. See [“Alignment Options”](#) on page 28.
6. Select the appropriate sequence – Illumina, Ion Torrent, or Other.



GeneMarker HTS automatically adjusts the default values for both Identity and Soft Clipping based on the Sequencer that you select. The default values are based on using an Illumina or Ion Torrent sequencer. If you change the value for either one or both settings, then GeneMarker HTS automatically select “Other” for Sequencer, and you cannot select a different value.

7. To add the samples for the project, click Add to open the Add Samples dialog box, and then select the sample files for the project.



The sample files can be in one of two formats – compressed FASTQ files (.fastq.gz) or uncompressed FASTQ files (.fastq).

The sample files are added to the project and are displayed in the Sample Files pane of the New Project dialog box. The sample file names are automatically generated from the names of the loaded files.

8. Optionally, you can do any or all of the following for loaded sample files as needed:
 - To split paired read sample files into separate sample entries, right-click any one of the paired read sample files, and on the context menu that opens, click Split.



To combine split paired read files, right-click any of the split files, and on the context menu that opens, click Combine.

- To change the name of a sample file, double-click the name of the file to select it, and then change the name of the file as needed.
 - To remove one or more sample files from a project, select the sample file (CTRL-click to select multiple files), and then click Remove.
 - To remove all sample files in a single step, click Remove All.
9. Optionally, do any or all of the following as needed:
- To adjust the settings that affect variant calling (and therefore, the variants that are displayed in the Variants table of the GeneMarker HTS Viewer), click Filter Settings to open the Filter Settings dialog box, and then edit any or all of the values as needed.
 - To adjust the settings that affect the information that is displayed in the Variants table of the GeneMarker HTS Viewer (and therefore, project report generation), click Table Settings to open the Table Settings dialog box, and then indicate which information is to be displayed in the Variants table and which is not.



For detailed information about Filter Settings and Table Settings, see “GeneMarker HTS Viewer Project Settings” on page 59.

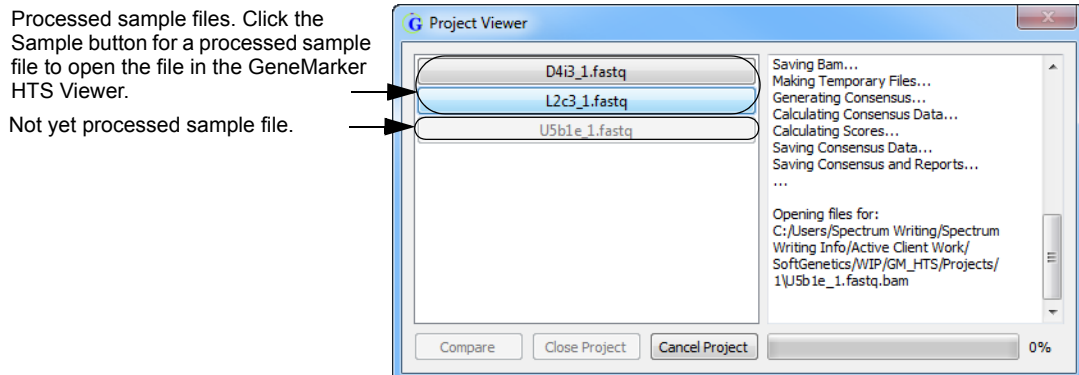
- To adjust the settings that mask the personal health information (PHI) for the subject, click PHI Settings to open the PHI Settings dialog box, and then add and/or remove default and/or custom hidden regions as appropriate. See “Masking personal health information (PHI) for a subject” on page 33.
10. Click OK.

The Project Viewer dialog box opens. The samples that are being processed for the project are displayed on the left side of the dialog box in the form of “Sample buttons.” The progress of processing a sample file is displayed on the right side of the dialog box. After a sample file is successfully processed for a project, the Sample button for the project is enabled. You can click the Sample button for any completed sample file to view its results (alignment, variant calls, and so on) in the GeneMarker HTS Viewer, even if other sample files are being processed or are waiting to be processed. See [Figure 2-10 on page 40](#).



Depending on your project setup, other messages might open before the Project Viewer dialog box opens, such as the message about creating a non-existent project folder. Answer all the prompts as needed to resolve the issues to which the prompts are referring and open the Project Viewer dialog box.

Figure 2-10: Project Viewer dialog box with processed and unprocessed Sample files



For a detailed list of the output files for a GeneMarker HTS project, see “GeneMarker HTS Project Output Files” on page 43.

11. Continue to [Chapter 3, “GeneMarker HTS Viewers,”](#) on page 45.

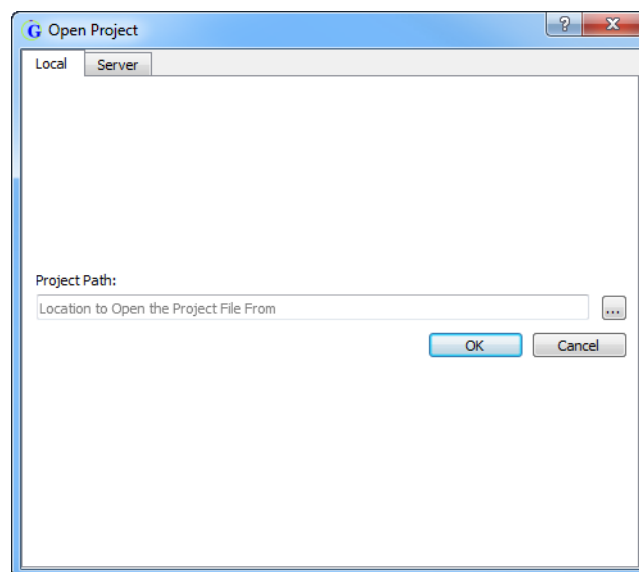
To work with an existing project

You can load a project that has been saved locally to the GeneMarker HTS client, or you can open a project that has been saved to a network server.

1. On the GeneMarker HTS main window, click Open.

The Open Project dialog box opens. The Local tab is the open tab.

Figure 2-11: Open Project dialog box, Local tab



2. Click the Browse button to open a secondary Open Project dialog box, and then browse to and select the project (*.pjt) file.

The secondary Open Project dialog box closes. You return to the primary Open Project dialog box. The full path to the folder in which the project was saved, including the project name (project.pjt) is displayed in the Project Path field.

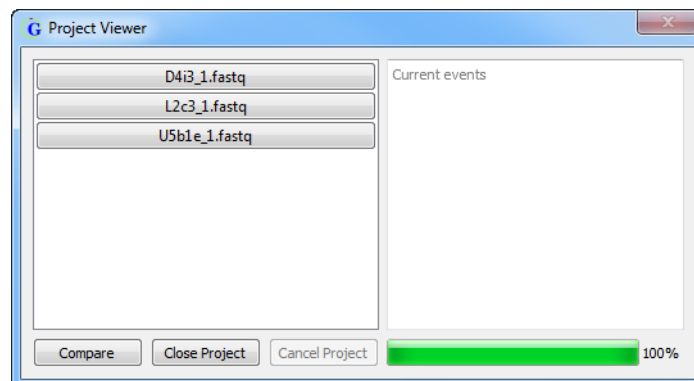


The last project directory path that you select here is persistent for a GeneMarker HTS instance and across GeneMarker HTS instances.

3. Click OK.

The primary Open Project dialog box closes and the Project Viewer dialog box opens. The left side of the dialog box displays all the samples for the project. The project status is shown as 100% completed.

Figure 2-12: Project Viewer dialog box for a completed project



4. Do either of the following:
 - To view the alignment results of a single sample in the project in a single GeneMarker HTS Viewer window, click the Sample button for the project. You can open multiple individual sample files at the same time. See [Figure 2-13 on page 42](#).
 - To compare the results for all the samples in the project in a single Comparison Viewer window, click Compare. See [Figure 2-14 on page 42](#).



When you open a sample file for a project in the GeneMarker HTS Viewer, the Filter Settings and Table Settings that were in place at the time that the project was saved determine the variants that are called and the information that is displayed in the Variants table in the viewer. You can always edit these settings as needed to better suit your working needs. For detailed information about the GeneMarker HTS Viewer, see [Chapter 3, “GeneMarker HTS Viewers,” on page 45](#).

Chapter 2
GeneMarker HTS Projects

Figure 2-13: GeneMarker HTS Viewer

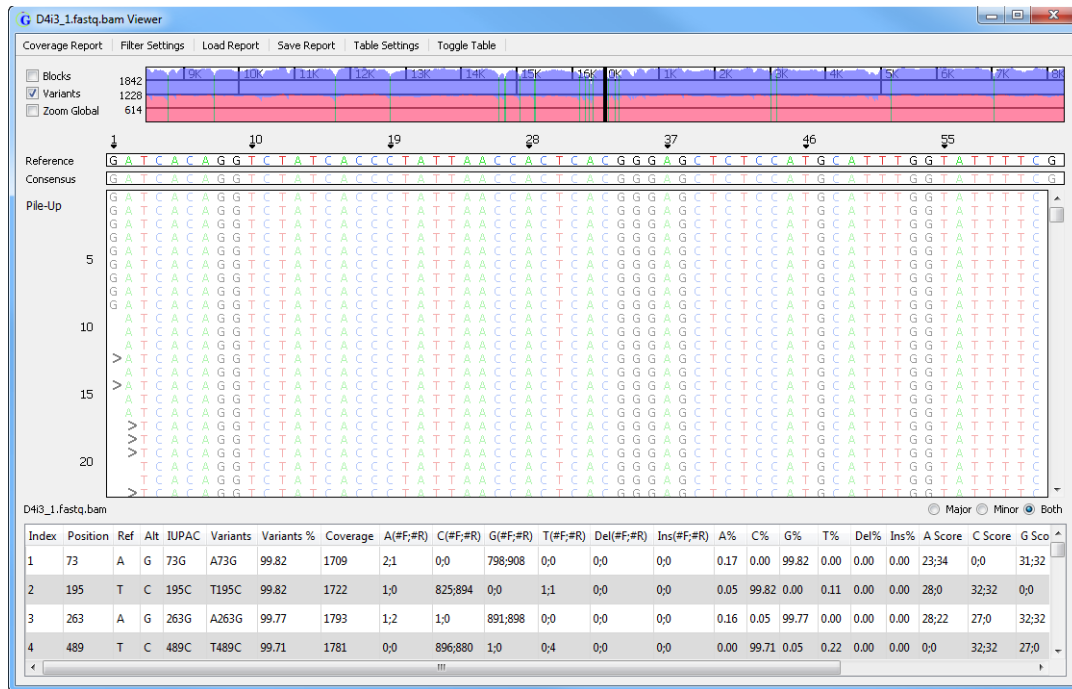
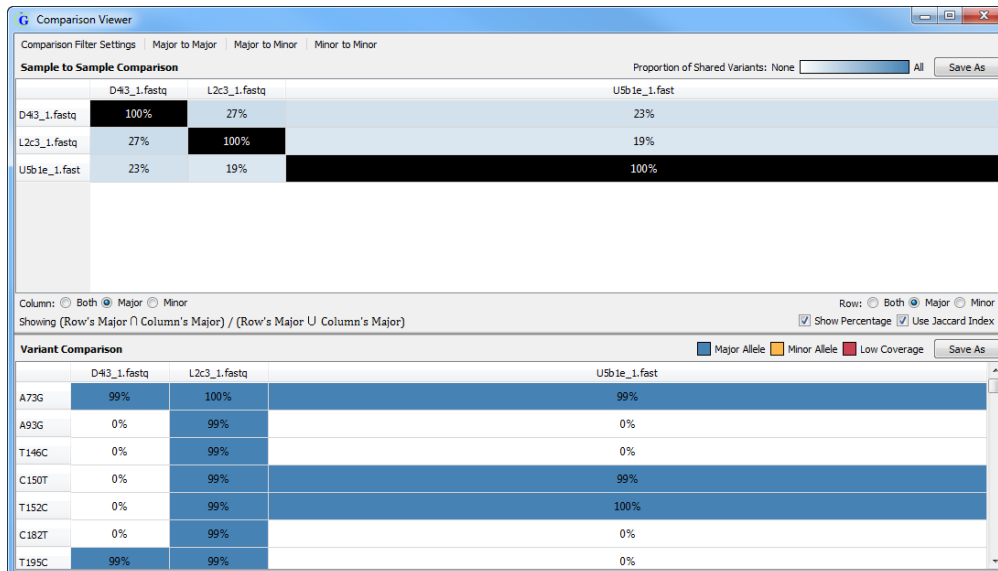


Figure 2-14: GeneMarker HTS Comparison Viewer



GeneMarker HTS Project Output Files

The table below lists the files and reports that are generated for each sample in an mtDNA alignment project.

| File | Description |
|-------------------------|--|
| BAM/BAI | Contains the fully detailed alignment results. |
| Alignment Statistic Log | Contains a summary of the alignment results. |
| Consensus Sequence | Details the consensus sequence with the major alleles in FASTA format. Note: A major allele is the allele with the highest frequency at a given position. All major alleles are reported unless the allele is the same as the reference. For detailed information about viewing and reporting on major and minor alleles for a sample alignment, see Chapter 3, "GeneMarker HTS Viewers," on page 45 . |
| Consensus Statistics | Lists the number of reads with each allele at each position in the sample. |
| Major Allele Report | Variant report for all major alleles. |
| Minor Allele Report | Variant report for all minor alleles. Note: A minor allele is any other allele with a frequency that is less than the allele with the highest frequency at a given position. For detailed information about viewing and reporting on major and minor alleles for a sample alignment, see Chapter 3, "GeneMarker HTS Viewers," on page 45 . |
| Project Settings | The settings (alignment, motif selection, and so on) for the project. Note: You can import a project settings file from one project to use in another. See "To specify the Filter Settings for a project" on page 59 . |

Chapter 2
GeneMarker HTS Projects

Chapter 3

GeneMarker HTS Viewers

Two viewers are available in GeneMarker HTS application for viewing the results of your mtDNA alignment projects: the GeneMarker HTS Viewer and the Comparison Viewer. The GeneMarker HTS Viewer is a viewing and editing tool that you can use to view and edit the results of your mtDNA alignment projects and produce a variety of interactive reports that summarize the alignment and other analysis information. The Comparison Viewer is a viewing tool that you can use to view a sample to sample comparison as well as a variant comparison of all the samples in the project at the same time.

This chapter covers the following topics:

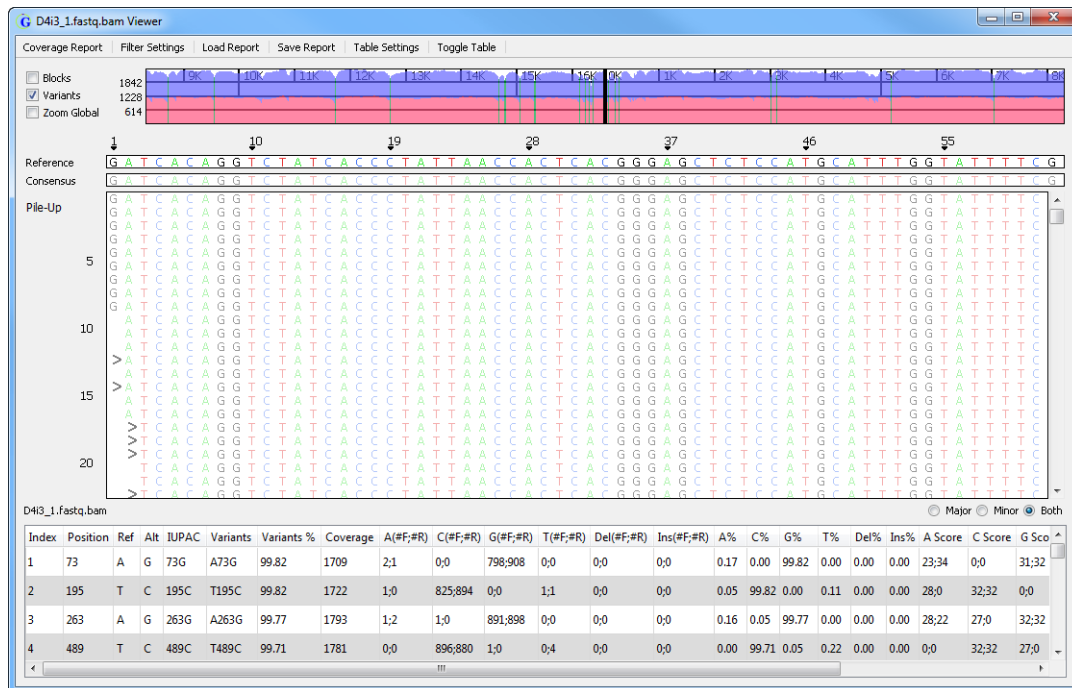
- [“GeneMarker HTS Viewer Layout” on page 46.](#)
- [“GeneMarker HTS Viewer Navigation and Options” on page 53.](#)
- [“GeneMarker HTS Viewer Project Settings” on page 59.](#)
- [“GeneMarker HTS Viewer Reports” on page 65.](#)

GeneMarker HTS Viewer Layout

You use the GeneMarker HTS Viewer to view the results of your mtDNA alignment projects. You can also edit the display settings for the viewer as well as the results of an alignment project. You can load one or more sample files in their own GeneMarker HTS Viewer windows while a project is being run, or you can load one or more sample files for a saved project. (See “Working with Projects in GeneMarker HTS” on page 36.) When you open a sample file for any project in the GeneMarker HTS Viewer, the Filter Settings and Table Settings that were in place at the time that the project was saved determine the variants that were called and the information that is displayed in the Variants table in the viewer, but you can always edit these settings as needed to better suit your working needs. (See “GeneMarker HTS Viewer Project Settings” on page 59.)

- If you load multiple sample files from *different* projects, then each project opens in its own separate GeneMarker HTS Viewer window, and the viewer windows are *not* linked. You can independently navigate and edit the display settings for each sample file in its own viewer window.
- If you load multiple sample files from the *same* projects, then each project opens in its own separate GeneMarker HTS Viewer window, but the viewer windows *are* linked. All the viewer windows zoom and scroll *horizontally* in unison. (The windows do not scroll vertically in unison.) Any changes that you make to the Filter Settings for one viewer window are applied to all opened viewer windows for the project’s sample files.

Figure 3-1: GeneMarker HTS Viewer



From top to bottom, the GeneMarker HTS Viewer has six major components:

- The title bar. See [“Title bar”](#) below.
- The main menu. See [“Main menu”](#) below.
- The Global View pane. See [“Global View pane”](#) on page 48.
- The Pane displays. See [“Pane displays”](#) on page 49.
- The Variants table. See [“Variants table”](#) on page 51.
- The Status bar. See [“Status Bar”](#) on page 52.

Title bar

The GeneMarker HTS Viewer title bar displays the name for the alignment project file that is being analyzed.

Figure 3-2: GeneMarker HTS Viewer Title bar



Main menu

The GeneMarker HTS Viewer main menu is set up in a standard Windows menu format with menu commands grouped into menus (Coverage Report, Filter Settings, Load Report, Save Report, Table Settings, and Toggle Table) across the menu bar.

| Menu Option | Description |
|-----------------|--|
| Coverage Report | Generate the Coverage Curve report for the sample. See “To generate the Coverage Curve report for a project” on page 65. |
| Filter Settings | Opens the Filter Settings dialog box, which contains options for changing the viewer display. See “GeneMarker HTS Viewer Project Settings” on page 59. |
| Load Report | Opens the Load Report dialog box, in which you can select a sample report for display in the viewer. See “To save/load reports for a project” on page 57. |
| Save Report | Opens the Save Report dialog box, in which you specify a folder for saving the information that is currently displayed in the Variants table as a report. See “To save/load reports for a project” on page 57. |
| Table Settings | Opens the Table Settings dialog box, which displays a list of all the columns that currently displayed in the Variants table of the GeneMarker HTS Viewer. By default, all available columns are initially displayed in the Variants table. To remove a column from the Variants table display, clear its selection. To add a column to the display, select it. See “GeneMarker HTS Viewer Project Settings” on page 59. |
| Toggle Table | Toggles the display (On or Off) of the Variants table in the GeneMarker HTS Viewer. See “Variants table” on page 51. |

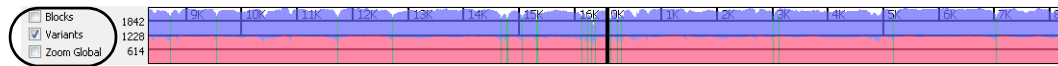
Global View pane

The Global View pane shows the depth of coverage for the sample. By default, the forward read coverage is displayed in blue and the reverse read coverage in red. A solid black line/hollow black rectangle indicates the range of the pile-up. By default, green lines indicate the locations of major variants, and orange lines indicate minor variants.



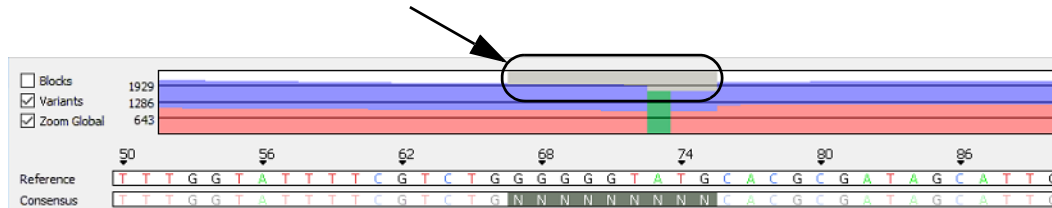
The zoom level for the Pile-Up pane determines the appearance of a line/rectangle indicator in the Global View pane. You can change all the default display colors for the pane on the Colors tab of the Application Settings dialog box. See “Specifying GeneMarker HTS Settings” on page 17.

Figure 3-3: Global View pane



By default, gray shading indicates areas of coverage that fall below the Total Coverage threshold that was specified for the Filter Settings for the project.

Figure 3-4: Shading alerts for low coverage regions in the Global View pane



For detailed information about the Total Coverage threshold, see “To specify the Filter Settings for a project” on page 59. “N” is displayed for the consensus sequence in these areas of low coverage in the Consensus Sequence pane. See “Pane displays” on page 49.

Select Zoom Global to zoom the display of the Global View pane to only that for the pile-up region.



For information about the Blocks option and the Variants option, see “Pane displays” below.

Figure 3-5: Zoom Global selected



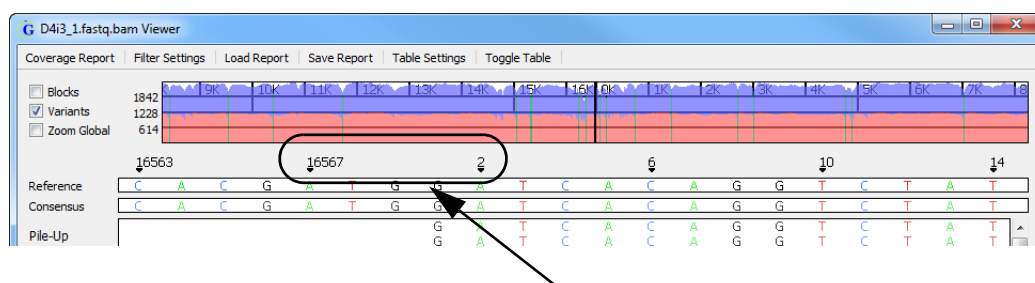
Pane displays

The Pane displays consist of three separate panes. Reference Position numbers are displayed at the top of the Panes displays. By definition, an mtDNA reference is circular, and this is reflected in the reference position numbers. As you scroll horizontally through a project display, the position numbers increase up to the end position of the reference, and then the reference position numbering starts over again.



For information about horizontally scrolling the display in the GeneMarker HTS Viewer, see “GeneMarker HTS Viewer Navigation and Options” on page 53.

Figure 3-6: Circular reference reflected in reference position numbering in the Panes display



From top to bottom, the three panes in the Panes display are the following:

- The Reference pane – The Reference pane displays the sequence of the reference to which the samples were aligned. To maintain position numbering in an mtDNA alignment project, an “N” is used instead of a nucleotide read in the Reference sequence. Variant calls are not reported for reference positions with an “N.”

Figure 3-7: “N” for a reference position to maintain position numbering

| 3102 | 3106 | 3110 | | | | | | | | |
|------|------|------|---|---|---|---|---|---|---|---|
| T | C | T | A | C | N | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |
| T | C | T | A | C | - | T | T | C | A | A |

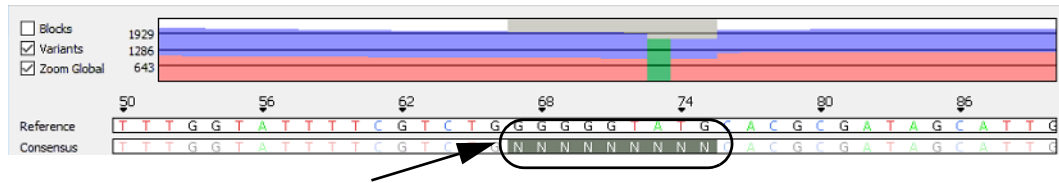
Position 3107 in the rCRS has an “N” to maintain position numbering. Bases that are aligned at this position are ignored.

- Consensus Sequence pane – The Consensus Sequence pane displays the consensus sequence for the major alleles, where a major allele is defined as the allele with the highest frequency at a given position. If any areas of coverage fall below the Total Coverage threshold that was specified for the Filter Settings for the project, then the consensus sequence is set to “N” for these regions in the Consensus Sequence pane.



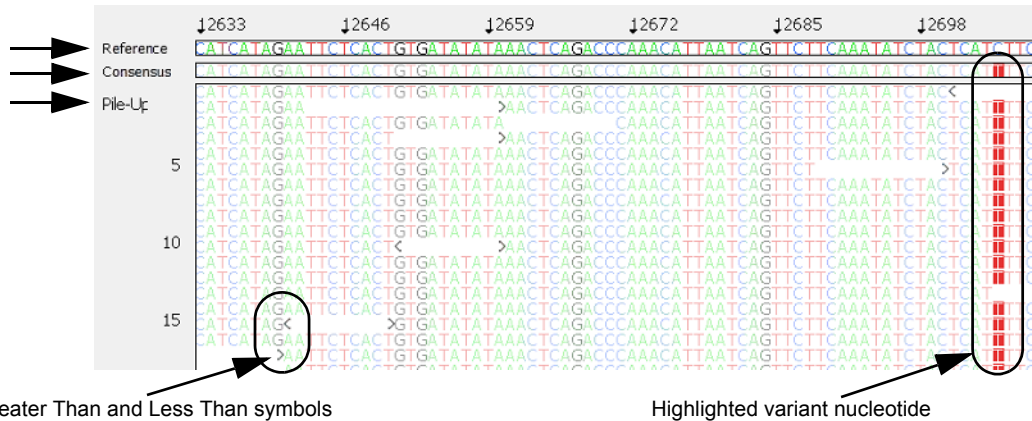
For detailed information about the Total Coverage threshold, see “To specify the Filter Settings for a project” on page 59. The corresponding areas of low coverage are highlighted in gray in the Global View pane. See “Global View pane” on page 48.

Figure 3-8: Shading alerts for low coverage regions in the Global View pane



- The Pile-Up pane – The Pile-Up pane displays a view of all the reads for the sample as they align to the reference sequence. Greater Than and Less Than symbols point to the 5' end of the read, where the Greater Than (>) symbol is used for forward reads and the Less Than (<) symbol is used for reverse reads.

Figure 3-9: Variants selected for Pane displays



Two options displayed next to the Global View pane are available for changing the Pane displays:

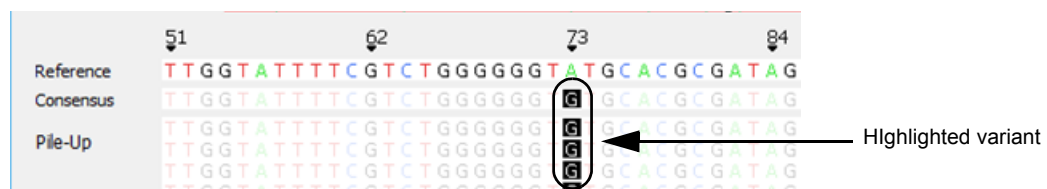
- Blocks – If Blocks is selected, then the background for each nucleotide is a color and the nucleotide (A, T, C, and G) is white.


Figure 3-10: Blocks selected for Pane displays



- Variants – Selected by default. If Variants is selected, then for all non-variant nucleotides, the background is white and the nucleotide (A, T, C, and G) is displayed in color. For variant nucleotides, the background is colored, and the nucleotide is displayed in white, which is referred to as a *highlighted* variant.

Figure 3-11: Variants selected for Pane displays



 To change the background and nucleotide colors for the Pane displays, see [“Specifying GeneMarker HTS Settings” on page 17.](#)


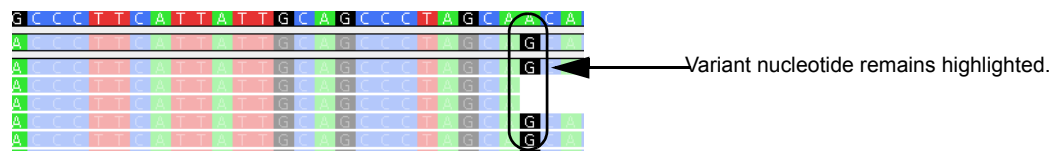
 You can select both Blocks and Variants for the display. In this case, all nucleotides are still displayed in white on a colored background; however, all nucleotides other than variant nucleotides are dimmed and all variant nucleotides are highlighted to visually differentiate them from non-variant nucleotides. See [Figure 3-12](#) below. You can adjust the visibility of the dimmed variants. See [“Specifying GeneMarker HTS Settings” on page 17.](#)

Figure 3-12: Blocks and Variants selected for pane display in the GeneMarker HTS Viewer



Variants table

By default, the Variants table lists each variant call in order of their sequence position, where:

- A *major* allele is defined as the allele with the highest frequency at a given position.
- All major alleles are reported unless the allele is the same as the reference.
- *Minor* alleles are any other allele at a given position with a frequency that is greater than the Minimum Threshold.
- Insertions are reported as major alleles if they occur at a frequency that is greater than or equal to 50% and as minor alleles if they occur at a frequency that is less than 50%.
- Because reference alleles can represent a secondary mutation of a major allele, reference alleles are reported as minor alleles if their frequencies occur at a rate that is greater than the minimum threshold for Allele Frequency.

See [Figure 3-13 on page 52.](#)

Figure 3-13: Variants table

Indicate which type of alleles to display in the Variants table.

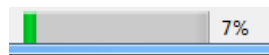
| Index | Position | Ref | Alt | IUPAC | Variants | Variants % | Coverage | A(#F;#R) | C(#F;#R) | G(#F;#R) | T(#F;#R) | Del(#F;#R) | Ins(#F;#R) | A% | C% | G% | T% | Del% | Ins% | A Score | C Score | G Score | T |
|-------|----------|-----|-----|-------|----------|------------|----------|----------|----------|----------|----------|------------|------------|-------|-------|-------|------|------|------|---------|---------|---------|----|
| 1 | 73 | A | G | 73G | A73G | 99.82 | 1709 | 2,1 | 0,0 | 798,908 | 0,0 | 0,0 | 0,0 | 0.17 | 0.00 | 99.82 | 0.00 | 0.00 | 0.00 | 23,34 | 0,0 | 31,32 | 0, |
| 2 | 195 | T | C | 195C | T195C | 99.82 | 1722 | 1,0 | 825,894 | 0,0 | 1,1 | 0,0 | 0,0 | 0.05 | 99.82 | 0.00 | 0.11 | 0.00 | 0.00 | 28,0 | 32,32 | 0,0 | 2, |
| 3 | 263 | A | G | 263G | A263G | 99.77 | 1793 | 1,2 | 1,0 | 891,898 | 0,0 | 0,0 | 0,0 | 0.16 | 0.05 | 99.77 | 0.00 | 0.00 | 0.00 | 28,22 | 27,0 | 32,32 | 0, |
| 4 | 489 | T | C | 489C | T489C | 99.71 | 1781 | 0,0 | 896,880 | 1,0 | 0,4 | 0,0 | 0,0 | 0.00 | 99.71 | 0.05 | 0.22 | 0.00 | 0.00 | 0,0 | 32,32 | 27,0 | 0, |
| 5 | 750 | A | G | 750G | A750G | 99.56 | 1855 | 5,3 | 0,0 | 940,907 | 0,0 | 0,0 | 0,0 | 0.43 | 0.00 | 99.56 | 0.00 | 0.00 | 0.00 | 26,26 | 0,0 | 32,32 | 0, |
| 6 | 1438 | A | G | 1438G | A1438G | 99.78 | 1875 | 2,0 | 0,0 | 904,967 | 0,0 | 1,1 | 0,0 | 0.10 | 0.00 | 99.78 | 0.00 | 0.10 | 0.00 | 21,0 | 0,0 | 32,32 | 0, |
| 7 | 2706 | A | G | 2706G | A2706G | 99.83 | 1859 | 1,2 | 0,0 | 928,928 | 0,0 | 0,0 | 0,0 | 0.16 | 0.00 | 99.83 | 0.00 | 0.00 | 0.00 | 16,28 | 0,0 | 32,32 | 0, |
| 8 | 3010 | G | A | 3010A | G3010A | 99.88 | 1818 | 871,945 | 0,0 | 1,0 | 0,1 | 0,0 | 0,0 | 99.88 | 0.00 | 0.05 | 0.05 | 0.00 | 0.00 | 32,32 | 0,0 | 27,0 | 0, |

By default, the Variants table shows both (major and minor) allele types, but you can choose to show only major alleles or only minor alleles. The specific alleles and the information that is displayed for the alleles in the Variants table are determined by the Filter Settings and Table Settings that were in place at the time that the project was saved. See “GeneMarker HTS Viewer Project Settings” on page 59. You can click a column header to change the sort order of the data. An icon in the column header indicates the sort direction. An Up arrow indicates an ascending sort and a Down arrow indicates a descending sort.

Status Bar

After the GeneMarker HTS Viewer opens, a temporary status bar is displayed in the lower left corner of the viewer below the Variants table. The status bar displays the progress of loading the project in the viewer. After the project is fully loaded, the status bar briefly displays 100%, and then it is closed and is no longer displayed in the viewer.

Figure 3-14: Status bar



Wait until after the status for loading a project is 100% and the status bar closes before beginning to work in the GeneMarker HTS Viewer.

GeneMarker HTS Viewer Navigation and Options

You can easily [navigate](#) the GeneMarker HTS Viewer using your mouse as well as some keyboard hotkeys. In various areas of the viewer, your mouse as well as some keyboard hotkeys to [work](#) with and [edit](#) the variant information in the viewer and in some areas, you can also right-click to open a context menu that contains a list of options for working with and editing the variant information.

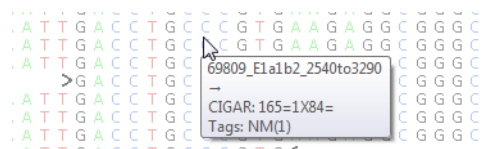
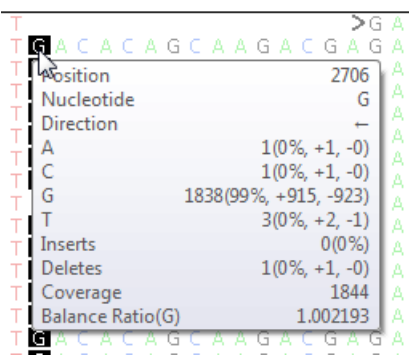
GeneMarker HTS Viewer navigation

You can easily navigate the GeneMarker HTS Viewer using your mouse as well as some keyboard hotkeys in the Global View pane, the Pane displays, and the Variants table.

| Option | Description |
|---|--|
| Global View pane | |
| Double-click any region in the Global View pane to move the Pane displays to the same area. | |
| Pane displays | |
| Zoom In | Hold down the left mouse button and draw a box from the upper left hand corner of the pane towards the lower right hand corner. A box is formed around the area that being reduced for viewing. |
| Zoom Out | Hold down the left mouse button and draw a box from the lower right hand corner of the pane towards the upper left hand corner. Note: The magnification for zooming out is always 100%. |
| Horizontal Scroll | Hold down the right mouse button and then drag the cursor across the display from left to right, or right to left. |
| Vertical Scroll | Use the Vertical Scroll bar to the right of the Panes display, or use the Scroll wheel on your mouse. |
| Moving to each variant position in the Pile-Up pane | When the GeneMarker HTS Viewer first opens, the starting position for the Pile-Up pane is the first reference nucleotide position (Pos #1). You can use the following keyboard shortcuts to move the display to each <i>variant</i> position: <ul style="list-style-type: none"> • CTRL + f moves the display forward one variant position at a time. • CTRL +b moves the display backward one variant position at a time. |
| Move to a closest variant position in the Variants table | Right-click any position in the Pile-Up pane and on the context menu that opens, click Go to closest position in table. The variant that is closest to the selected position is highlighted in the Results table. |
| Variants table | |
| Double-click any entry in the Variants table to move the Pane displays to the corresponding variant position. | |

GeneMarker HTS Viewer options

In various areas of the viewer, you can use your mouse as well as some keyboard hotkeys to work with and edit the variant information in the viewer. In some areas, you can also right-click to open a context menu that contains a list of options for viewing and editing the variant information

| Option | Description |
|---|---|
| Pile-Up pane | |
| <p>Open a tooltip with the following information for a specific read (variant or non-variant):</p> <ul style="list-style-type: none"> • Read name • Read direction • Read CIGAR string • Read Tags | <p>Do one of the following:</p> <ul style="list-style-type: none"> • SHIFT + mouse scroll wheel click • CTRL +SHIFT and then left mouse click. <p><i>Figure 3-15: Read tooltip</i></p>  |
| <p>Open a tooltip with the following information for a specific variant:</p> <ul style="list-style-type: none"> • The position number for the allele • The major allele at the position • The counts for the different alleles at the position • The read direction for the allele • The total coverage for the position • The Allele Balance Score | <p>Do one of the following:</p> <ul style="list-style-type: none"> • Click the mouse scroll wheel • CTRL + left mouse click <p><i>Figure 3-16: Variant tooltip</i></p>  |
| <p>Go to closest position in table</p> | <p>Right-click any position, and on the context menu that opens, select this option to highlight the entry in the Variants table for the variant that is closest to this selected position.</p> |
| <p>Copy read's sequence</p> | <p>To copy the sequence from any read to your clipboard, right-click any read, and on the context menu that opens, click Copy read's sequence. You can then use standard Paste commands to paste the copied information into a third-party product such as Microsoft Word or Notepad.</p> |
| <p>Adding a variant</p> | <p>See "To add a variant at a given position" on page 55.</p> |

| Option | Description |
|------------------------|--|
| Variants table | |
| Copy selected | Click and hold the left mouse button and then draw a box (square or rectangle) around the cells that are to be copied. Right-click any of the copied cells, and on the context menu that opens, select Copy selected to copy the selected table cells to your clipboard. Use standard keyboard commands or menu commands to paste the copied cells into a third-party application such as Microsoft Word or Excel. |
| Deleting a variant | See “To delete a variant at a given position” on page 56. |
| Saving/Loading reports | See “To save/load reports for a project” on page 57. |

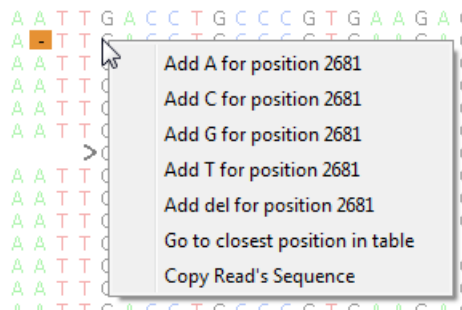
To add a variant at a given position

You can manually add a variant (insertion or deletion) at a position of your choice in the Pile-Up pane.

1. In the Pile-Up pane, right-click the position at which you are adding the variant.

A context menu opens with various options, including options for adding a variant.

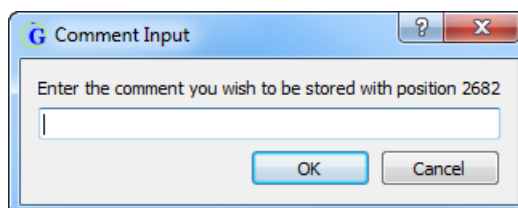
Figure 3-17: Pile-Up pane context menu



2. Select the appropriate variant addition option.

The context menu closes and Comment Input dialog box opens.

Figure 3-18: Comment Input dialog box



3. Optionally, enter the comment that is to be stored for the indicated position, and then click OK.

The Comment Input dialog box closes. A new entry that is highlighted in green is displayed for the added variant in the Variants table. See [Figure 3-19 on page 56.](#)

Figure 3-19: New variant entry in Variants table

| Index | Position | Ref | Alt | IUPAC | Variants | Variants % | Coverage | A(#F,#R) | C(#F,#R) | G(#F,#R) | T(#F,#R) | Del(#F,#R) | Ins(#F,#R) | A% | C% | G% | T% | Del% | Ins% |
|-------|----------|-----|-----|--------|----------|------------|----------|----------|----------|----------|----------|------------|------------|------|-------|------|------|------|------|
| 9 | 2706 | A | C | Z706G | A2706C | 0.05 | | | | | | | | | | | | | |
| 10 | 2706 | A | G | Z706G | A2706G | 99.67 | | | | | | | | | | | | | |
| 11 | 2710 | C | A | Z710C | C2710A | 0.00 | 2164 | 0,0 | 960,1201 | 0,1 | 1,0 | 0,1 | 0,0 | 0.00 | 99.86 | 0.04 | 0.04 | 0.04 | 0.00 |
| 12 | 3027 | T | C | Z3027C | T3027C | 99.88 | 1800 | 0,0 | 900,898 | 0,0 | 2,0 | 0,0 | 0,0 | 0.00 | 99.88 | 0.00 | 0.11 | 0.00 | 0.00 |

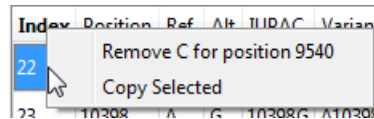
To delete a variant at a given position

You can manually delete a variant (insertion or deletion) from a project in the Variants table.

1. In the Variants table, right-click the variant that you are deleting.

A context menu opens with two options: Remove the variant for the selected position and Copy Selected.

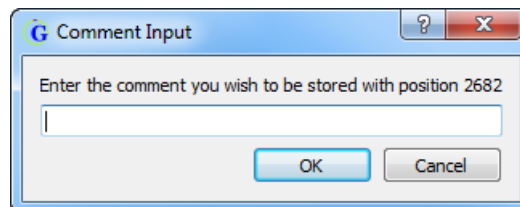
Figure 3-20: Variants table context menu



2. Select Remove <variant> for position <>.

The context menu closes and Comment Input dialog box opens.

Figure 3-21: Comment Input dialog box



3. Optionally, enter the comment that is to be stored for the indicated position, and then click OK.

The Comment Input dialog box closes. A new entry that is highlighted in red is displayed for the deleted variant in the Variants table.

Figure 3-22: New variant entry in Variants table

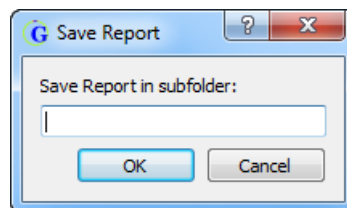
| Index | Position | Ref | Alt | IUPAC | Variants | Variants % | Coverage | A(#F,#R) | C(#F,#R) | G(#F,#R) | T(#F,#R) | Del(#F,#R) | Ins(#F,#R) | A% | C% | G% | T% | Del% | Ins% |
|-------|----------|-----|-----|--------|----------|------------|----------|----------|----------|----------|----------|------------|------------|------|-------|------|------|------|------|
| 9 | 2706 | A | C | Z706G | A2706C | 0.05 | | | | | | | | | | | | | |
| 10 | 2706 | A | G | Z706G | A2706G | 99.67 | | | | | | | | | | | | | |
| 11 | 2710 | C | A | Z710C | C2710A | 0.00 | 2164 | 0,0 | 960,1201 | 0,1 | 1,0 | 0,1 | 0,0 | 0.00 | 99.86 | 0.04 | 0.04 | 0.04 | 0.00 |
| 12 | 3027 | T | C | Z3027C | T3027C | 99.88 | 1800 | 0,0 | 900,898 | 0,0 | 2,0 | 0,0 | 0,0 | 0.00 | 99.88 | 0.00 | 0.11 | 0.00 | 0.00 |

To save/load reports for a project

After you generate a project for the first time, the project settings (Filter Settings and Table Settings) are saved as Initial_Report at the Project Folder level. After you make any changes to the project settings and/or any manual edits to the project (for example, adding or deleting a variant), you can save these changes as another project report. Each report that you save is saved in a separate sub-folder under the primary project folder with a Change Log file that details the changes made to the project settings and project results. With a project opened in the GeneMarker HTS Viewer, you can load any report that has been saved for the project to dynamically update the information that is displayed for the project in the viewer.

1. Do any or all of the following as needed:
 - On the GeneMarker HTS Viewer main menu, click Filter Settings to open the Filter Settings dialog box, make any needed edits to the settings, and then click OK to save the edited settings. See [“To specify the Filter Settings for a project” on page 59.](#)
 - On the GeneMarker HTS Viewer main menu, click Table Settings to open the Table Settings dialog box, make any needed edits to the settings, and then click OK to save the edited settings. See [“To specify the Table Settings for a project” on page 63.](#)
 - In the Pile-Up pane, add any variants as needed. See [“To add a variant at a given position” on page 55.](#)
 - In the Variants table, delete any variants as needed. See [“To delete a variant at a given position” on page 56.](#)
2. On the GeneMarker HTS Viewer main menu, click Save Report.
The Save Report dialog box opens.

Figure 3-23: Save Report dialog box



3. Enter a name for the subfolder in which the report is to be saved, and then click OK.

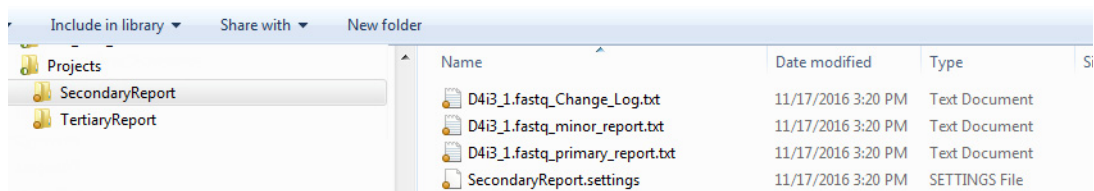


The subfolder name cannot contain any spaces or special characters. Underscores () are allowed.

The Save Report dialog box closes. The report is saved in a separate sub-folder under the primary project folder with a Change Log file that details the changes made to the project results. You can now load this report for a project to dynamically update the information that is displayed for the project in the viewer. See [Figure 3-24 on page 58.](#)

Chapter 3
GeneMarker HTS Viewers

Figure 3-24: Saved Project report and folder structure



GeneMarker HTS Viewer Project Settings

Two types of settings determine the information that is displayed for a project in the GeneMarker HTS Viewer: Filter Settings and Table Settings. When you open a sample for any project in the GeneMarker HTS Viewer, the [Filter Settings](#) and [Table Settings](#) that were in place at the time that the project was saved determine the information that is displayed in the viewer, but you can always edit these settings as needed to better suit your working needs.

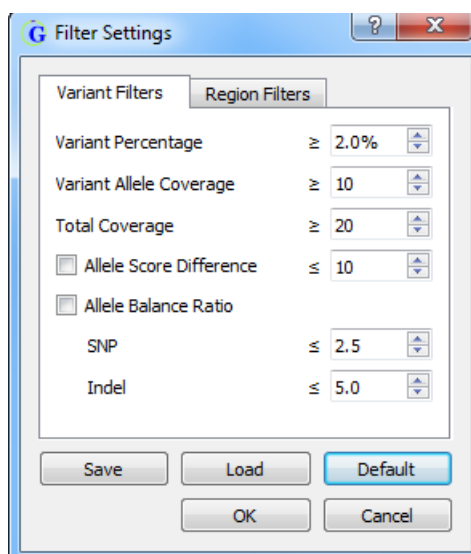
To specify the Filter Settings for a project

The Filter Settings dialog box provides all the options for adjusting the variant calling settings for a project, which determine the variants that are displayed in the Variants table in the GeneMarker HTS Viewer. You can edit the default values for the Filter Settings, or you can load a project.settings file for a previously run project, and then apply the settings as-is, or optionally, you can first edit any of the values as needed, and then apply the settings.

1. On the GeneMarker HTS Viewer main menu, click Filter Settings.

The Filter Settings dialog box opens. The Variant Filters tab is the open tab.

Figure 3-25: Filter Settings dialog box, Variant Filters tab





For Step2, note the following:

- If you edit the values for any Filter Settings, then to return all settings to their default values (the factory values set by SoftGenetics) at the same time, click Default.
- If you edit any of the Filter Settings, then before you click OK to apply the settings, you can click Save to save the Filter Settings to a Settings(.settings) file.

2. Do one of the following:

- Edit the values for any of the Filter Settings values as needed.
- Click Load to open the Load Settings dialog box, and then browse to and select the appropriate project settings (.settings) file.

The Filter Settings dialog box is populated with the values for the loaded project settings file. You can leave the values as-is, or you can edit any of the values as needed.

Table 3-1: Filter Settings dialog box, Variant Filters tab

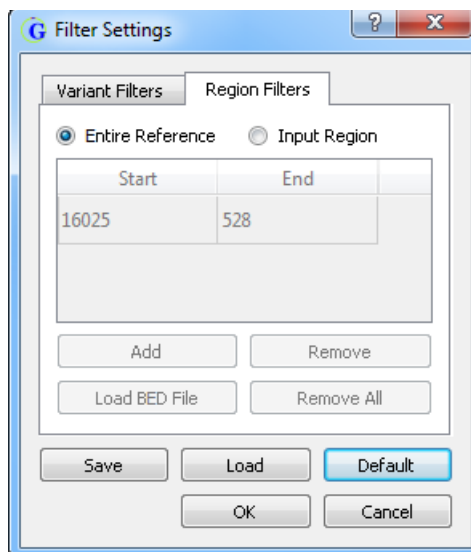
| Option | Description |
|-------------------------|---|
| Variant Percentage | The percent of reads for which the variant was called at the indicated position. |
| Variant Allele Coverage | The number of reads that contained the variant at the indicated position. |
| Total Coverage | The total number of reads for the indicated position. |
| Allele Score Difference | A value that is calculated for each minor allele to determine if the allele is a true variant and therefore, is to be included in the analysis results, or if the allele is a false positive and therefore, is to be filtered from the results. See “Allele Score Difference” on page 62, |
| Allele Balance Ratio | A value that is calculated for each allele to determine if the allele is a true variant and therefore, is to be included in the analysis results, or if the allele is a false positive and therefore, is to be filtered from the results. See “Allele Balance Ratio” below. |
| SNP | Default value is ≤ 2.5 . |
| Indel | Default value is ≤ 5.0 . |

3. If you need to change the reference region for the analysis, then go to [Step 4](#); otherwise, click OK to close the Filter Settings dialog box and immediately apply the settings.

The GeneMarker HTS display is dynamically updated based on the settings.

4. Open the Region Filters tab.

Figure 3-26: Filter Settings dialog box, Region Filters tab



5. Do one of the following:
 - To call variants in any region, select Entire Reference. (The default value.)
 - To limit variant calls to a single specific region, select Input Region, and then adjust the Start and End values as needed.



By default, the Start and End values are set to define the region (16025 and 528, respectively) that is the most commonly specified for targeted sequencing.

- To limit variant calls to multiple specific regions, select Input Region, select Load BED File, and then in the Load BED Settings dialog box, browse to and select the BED file that defines these multiple regions.



For detailed information about the format for a BED file, see “BED file” on page 69. After you load a BED file, you can use the Add and Remove options to add and remove regions for the file. If you add a region, you must specify the Start and End positions for it.

6. Click OK to close the Filter Settings dialog box and immediately apply the settings.
The GeneMarker HTS display is dynamically updated based on the settings.

Allele Score Difference

In GeneMarker HTS, a quality score is calculated for every SNP or insertion allele as the median of the base quality score from the original data. Because base quality scores are not possible for a deletion, the quality scores for the bases that surround a deletion are used instead. The Allele Score Difference filter, which is based on the base quality score for an allele, is applied individually *for each direction* as follows:

- The total number of reads for the *minor* allele is recorded.
- The base quality score for each minor allele is subtracted from the base quality score for the major allele to calculate the Allele Score Difference.

If the Allele Score Difference is greater than its set threshold and the total number of reads for the minor allele is greater than or equal to five, then the variant is assumed to be a false positive and the minor allele is filtered from the alignment results; otherwise, if any of these criteria are not met, then the allele is assumed to be a true allele and is not filtered from the results.

Allele Balance Ratio

The Allele Balance Ratio is another filter that you can use to determine if an allele is a true allele or a false positive. The Allele Balance ratio is defined as the maximum value for *one* of the following:

- Allele Forward %/Total Forward % *or* Total Forward %/Allele Forward %
- Allele Reverse %/Total Reverse % *or* Total Reverse %/Allele Reverse %

with different default thresholds specified for SNPs (2.5) and indels (5.0).

Consider the following scenario, where:

- Two forward reads and 79 reverse reads with the allele (a SNP), for 81 total *allele* reads.
- 1758 total forward reads and 6055 total reverse reads, for 7812 total reads.

This results in the following:

| Value | Calculation |
|--|-------------------------------------|
| Total Forward % | $1758/7813 = 0.22501$ |
| Total Reverse % | $6055/7813 = 0.77499$ |
| Allele Forward % | $2/81 = 0.02469$ |
| Allele Reverse % | $79/81 = 0.97531$ |
| Allele Forward %/Total Forward % | $0.02469/0.22501 = 0.9113$ |
| Total Forward %/Allele Forward % | $0.22501/0.02469 = \mathbf{9.1134}$ |
| Allele Reverse %/Total Reverse % | $0.97531/0.77499 = 1.2585$ |
| Total Reverse %/Allele Reverse % | $0.77499/0.97531 = 0.7946$ |
| Because 9.1134 is > 2.5, this allele is filtered from the results. | |

Although SoftGenetics has supplied the default values for what they have determined to be the most appropriate for the calculation of the Allele Balance Ratio, you can always edit one or both of these thresholds as needed. If you edit a value, you must try to maintain a balance of sensitivity versus specificity. For example, if you are concerned that variants are being called that are actually false positives and the variants also appear to be unbalanced, then you might tighten the restrictions on variant calling and decrease the appropriate threshold value. Conversely, if some expected variants are not being called, then you might loosen the restrictions on variant calling and increase the appropriate threshold value.



If an allele is determined to be one directional, then an Allele Balance Ratio is not calculated. Instead, the allele is not filtered from the results only if it is a major allele.

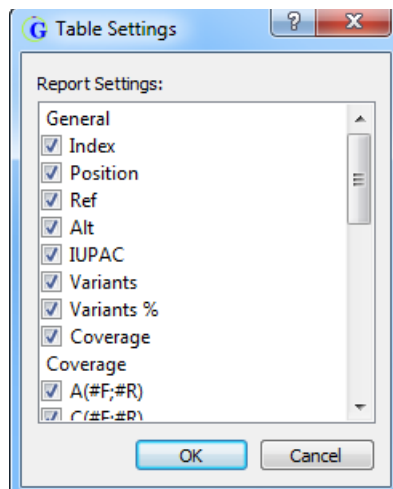
To specify the Table Settings for a project

The Table Settings determine the data columns (information) that are displayed in the Variants table for each called allele in a project. By default, all available data columns are displayed for an allele in the Variants table, but you can always specify which columns are to be displayed, and which are not. The information that is displayed for an allele is categorized as one of the following: General, Coverage, Coverage Percent, and Score.

1. On the GeneMarker HTS Viewer main menu, click Table Settings.

The Table Settings dialog box opens. By default, all columns are selected for inclusion in the Variants table.

Figure 3-27: Table Settings dialog box



2. Leave the columns that are to be included in the Variants table selected; otherwise, to hide a column in the Variants table, clear its selection.

| Column | Description |
|---|---|
| General | |
| Index | Unique system-generated identifier for the variant. |
| Position | The nucleotide position in the chromosome where the variant occurs. |
| Ref | The reference nucleotide at the variant position. |
| Alternate Allele | The alternate (variant) nucleotide at the variant position. |
| IUPAC | The IUPAC nucleotide code for the allele. If multiple alleles are called at a position, then the appropriate IUPAC ambiguity code is reported. Note: Mixtures of a deletion and other alleles use a lowercase letter. |
| Variants | The standard variant call. |
| Variant % | The variant frequency at the indicated position. |
| Coverage | The total number of reads (total coverage) at the variant position. |
| Coverage | |
| A #(F,R), C #(F,R), G #(F,R), T #(F,R) | The number of reads that show the indicated nucleotide at the variant location in the forward direction and the number of reads that show the indicated nucleotide at the variant location in the reverse direction. |
| Del #(F,R) | The number of reads that show a deletion at the variant location in the forward direction and the number of reads that show a deletion at the variant location in the reverse direction. |
| Ins #(F,R) | The number of reads that show an insertion at the variant location in the forward direction and the number of reads that show an insertion in the reverse direction at the variant location. |
| Coverage Percent | |
| A(%), C(%), G(%), T(%) | The percentage of reads that show the indicated nucleotide at the variant location. |
| Del(%) | The percentage of reads that show a deletion at the variant location. |
| Ins(%) | The percentage of reads that show an insertion at the variant location. |
| Score | |
| A Score, C Score, G Score, T Score | The median nucleotide call quality for the indicated allele (SNP) in both the forward and reverse directions. |
| Del Score | The median nucleotide call quality for deletion alleles. |
| Ins Score | The median nucleotide call quality for insertion alleles. |

3. Click OK to close the Table Settings dialog box and immediately update the display of the Variants table.

GeneMarker HTS Viewer Reports

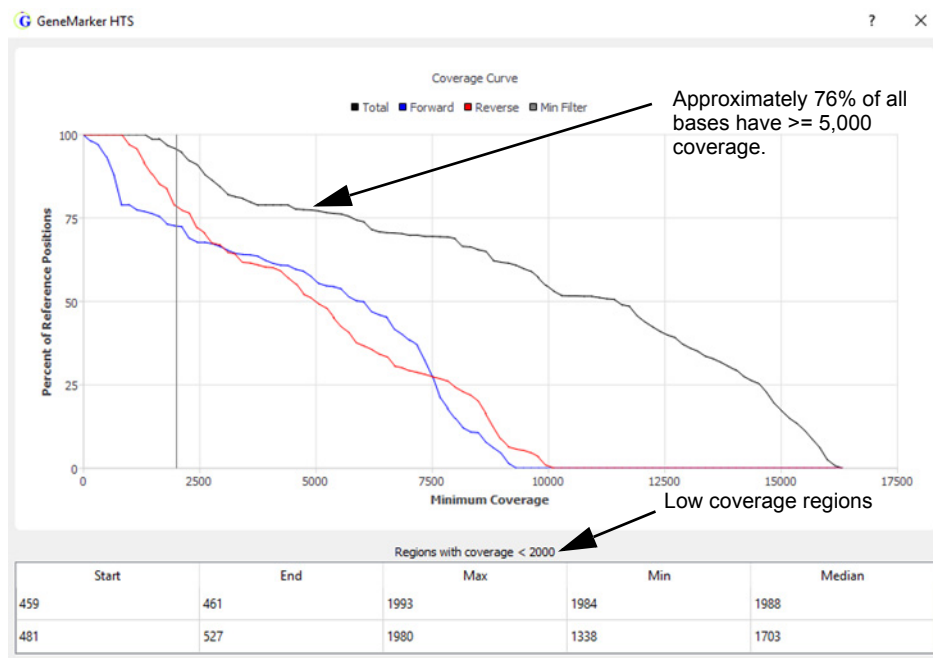
Currently, a single report, the Coverage Curve report, is available for [generation](#) in GeneMarker HTS. The Coverage Curve provides a high-level overview of the fraction of bases (the entire reference range, or the range of bases that was specified for the Filter Settings for the project) that have some level of coverage that is greater than or equal to the Total Coverage threshold that was specified for the project. The report displays the coverage distribution of sample reads without directional information and individually for both the forward and reverse directions and reports low coverage regions in a separate table below the report. The report is useful for identifying regions that were not adequately sequenced because of low coverage.

To generate the Coverage Curve report for a project

1. On the GeneMarker HTS Viewer main menu, click Coverage Report.

The Coverage Curve report is automatically generated and opens in a separate window.

Figure 3-28: Coverage Curve report



2. Optionally, to save the Low Coverage Regions table as a text file, click Save As.

You must provide a name for the report. By default, the Low Coverage Regions report is saved in the top level project folder, but you can always select a different location.

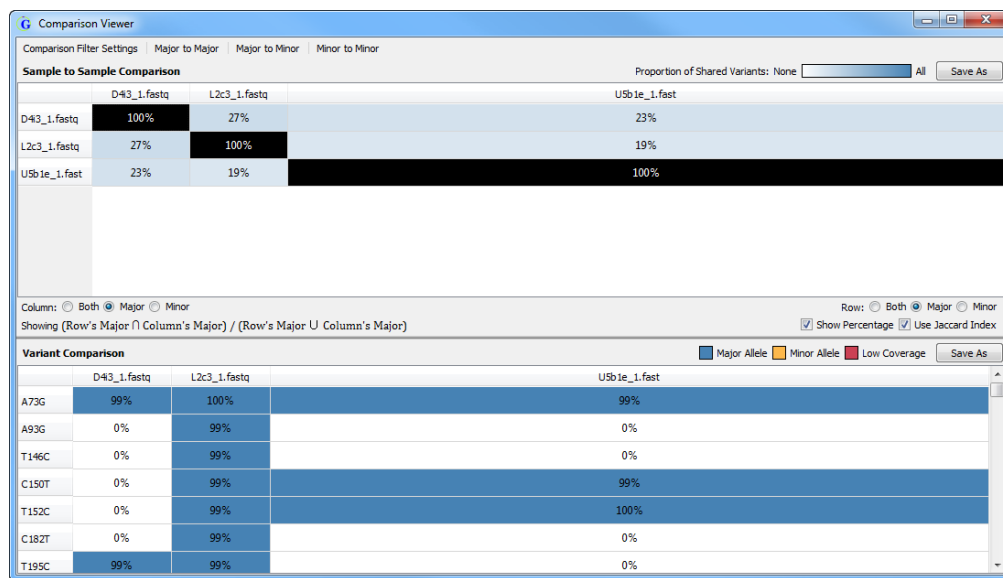
GeneMarker HTS Comparison Viewer

You can compare the results for all the samples in the project in a single Comparison Viewer window. To do so, open the appropriate project, and then in the Project Viewer dialog box, click Compare to open the Comparison Viewer window.



For detailed information about opening an existing project, see “To work with an existing project” on page 40.

Figure 3-29: GeneMarker HTS Comparison Viewer



The Comparison Viewer is divided into two panes: the **Sample to Sample Comparison** (top) pane and the **Variant Comparison** (bottom) pane. Each pane displays unique information about a project. You can use the Save As option to save the information that is displayed in both panes as text (.txt) file with a filename of your choosing. By default, the text file is saved in the same directory as the project, but you can always select a different location.

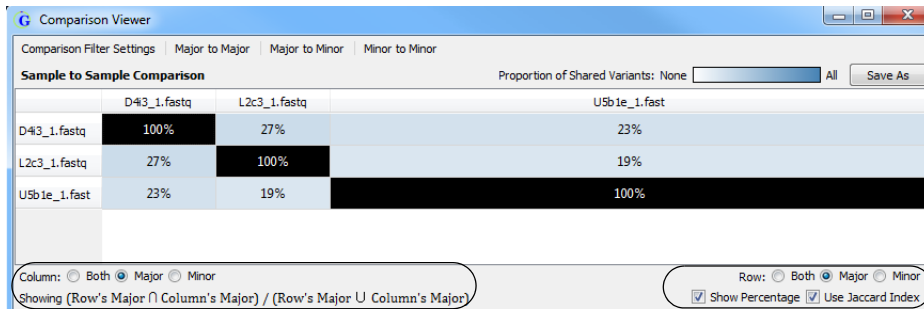
Sample to Sample Comparison pane

The Sample to Sample Comparison pane for the Comparison Viewer is a correlation table, that, by default, displays the percentage of major alleles that are shared among all the project samples, with the Jaccard Index measurement used to calculate the comparison. Diagonal cells in the correlation table compare a sample to itself and by default, are highlighted in black. All other cells in the correlation table compare one sample to another, and by default, are highlighted in a blue color gradient from 0 to 100%, where 0% cells are blank/white. See [Figure 3-30 on page 67](#).



You can change the default highlighting colors that used in the Sample to Sample Comparison pane. See “Specifying GeneMarker HTS Settings” on page 17.

Figure 3-30: Comparison Viewer, default display



Options are available at the bottom of the Sample to Sample Comparison pane for the following:

- Changing the variant types that are used for calculating the correlation (Major, Minor, or Both). You can select the same or different variant types for the column and the row.

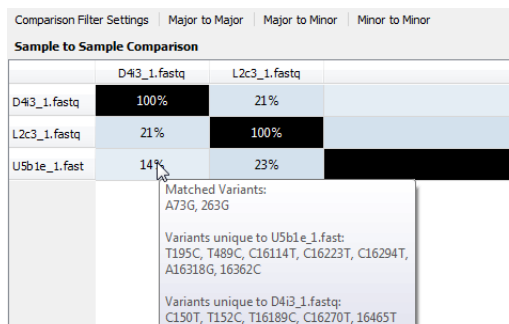


You can also use the options that are available on the Comparison Viewer main menu (Major to Major, Major to Minor, and Minor to Minor) to quickly reset the variant types that are used for the calculation of the correlation.

- Changing the calculation that is used for the correlation. By default, the Jaccard Index measurement is used for the calculation, which results in the variant values in either the row or the column being used in the divisor. If you clear this option, then a Simple measurement is used for the calculation, which results in only the variant value in the column being used in the divisor.
- Showing the percentage of shared variants, or the absolute number of shared variants. By default, Show Percentage is selected, but you can clear this option to show absolute number.

As you change any of the options for the correlation calculation, both the correlation table display and the correlation calculation are updated accordingly. You can hold your mouse pointer over any entry in the correlation table to open a tooltip that displays the percentage or absolute number of matched variants, and if applicable, all the variants that are unique to each appropriate sample.

Figure 3-31: Correlation Table tooltip



Variant Comparison pane

The Variant Comparison pane lists all the variants that were found in at least one sample for the project, where each column is a different sample and the variant percentage for the sample is displayed. The display provides a quick way of visually determining which variants were shared among samples and which were not. Variants that can be classified as a major allele or as a minor allele are highlighted, by default, in blue or gold, respectively. If a variant was determined to be a Low Coverage variant, (the Total Coverage for the variant falls below the Total Coverage threshold that was set for the project), then by default, the variant is highlighted in red.



You can change the default highlighting colors that used in the Variant Comparison pane. See [“Specifying GeneMarker HTS Settings” on page 17](#).

Figure 3-32: Variant Comparison pane

| Variant Comparison | SED01_1.f.a | SED02_1.f.a | SED03_1.f.a | SED04_1.f.a | SED05_1.f.a |
|--------------------|-------------|-------------|-------------|-------------|-------------|
| A73A | 0% | 0% | 0% | 0% | 99% |
| A73G | 99% | 99% | 99% | 99% | 0% |
| A83G | 0% | 4% | 0% | 0% | 0% |
| T14C | 0% | 4% | 0% | 0% | 0% |
| C20F | 0% | 16% | 0% | 0% | 0% |
| T15C | 0% | 16% | 0% | 0% | 0% |
| C30T | 0% | 5% | 0% | 0% | 0% |
| G15A | 0% | 0% | 0% | 13% | 0% |
| T15C | 99% | 92% | 0% | 4% | 99% |
| T15T | 0% | 17% | 99% | 99% | 0% |
| C20F | 0% | 5% | 0% | 0% | 0% |
| G20A | 0% | 0% | 0% | 14% | 0% |
| A20G | 99% | 99% | 99% | 99% | 99% |
| C20T | 0% | 0% | 0% | 14% | 0% |

The specific variants that are initially displayed in the Variant Comparison pane are determined by the Filter Settings, including the Total Coverage threshold, that were in place at the time that the project was saved. To dynamically change the display, on the Comparison Viewer main menu, click Comparison Filter Settings to open the Filter Settings dialog box, and make any needed modifications. See [“To specify the Filter Settings for a project” on page 59](#).

Glossary

B

BED file

Also known as Region of Interest (*.bed file). A BED file is a tab-delimited text file. You can upload a BED file only if the reference sequence contains chromosome information. Each row in the file contains a region of the reference that is to be used for the analysis, and at a minimum, the file must contain the following information:

- Field #1 - Chromosome number or name for the region (for example, chrM)
- Field #2 - Chromosome start position (for example, 300)
- Field #3 - Chromosome end position (for example, 305)
- Field #4 - Optional description column (for example, Region 102)

The positions are 0-based and open-ended. For example, a start of 10 and end of 15 would include reference positions 11, 12, 13, 14, and 15.

C

Consensus alignment

An optional alignment step. Because each read is considered on its own, indels that are near the end of reads are sometimes not correctly aligned. If this alignment option is selected, then an extra step is added after the initial alignment. In this extra step, indel positions are re-examined to possibly improve the alignment of these reads based on the alignment of other reads that have the indel.

M

Motif alignment

An optional alignment step. When equivalent sequences can be aligned in multiple ways (motifs), an analyst might not prefer the alignment method that the software selects. This step ensures that the selected motif is the motif that is defined by the motif file.

P

PCR duplicates

PCR duplicates are a set of pairs (paired-end data) or reads (non-paired-end data) that have been generated from the same original fragment. The “Remove PCR Duplicates” optional

alignment step attempt to identify these reads and ignores all but one pair (or read) in each set.

Personal Health Information (PHI)

Personal health information can be inferred from some regions of whole mtDNA sequence. To maintain privacy during whole mtDNA analysis, a BED file can be used to specify the positions that must be hidden from the analyst.

Proper Pairs

A pair of aligned reads is referred to as a “proper pair” if the reads are aligned on opposite strands in the correct orientation. Non-proper pairs can result from misalignment, sequencing error, or off-target amplification.

Index

A

- About button on the GeneMarker HTS main window..... 16
- Application Settings icon on the GeneMarker HTS main window... 15
- Assumptions for the manual 8

C

- context menu for the GeneMarker HTS Viewer 54
- Conventions used in the manual ... 7

G

- GeneMarker HTS
 - installing..... 12
 - main window 15
 - About button 16
 - Application Settings icon 15
 - Online Help button..... 16
 - Project Launcher options..... 16
 - title bar..... 15
 - specifying the customization settings for 17
 - starting 14
- GeneMarker HTS Comparison Viewer
 - Sample to Sample Comparison pane..... 66
 - Variant Comparison pane 68
- GeneMarker HTS Viewer
 - adding a variant in the Pile-Up pane for..... 54
 - context menu for 54
 - deleting a variant for a project in the Variants table for..... 54

- layout of 46
 - Global View pane..... 48
 - main menu 47
 - Pane displays 49
 - status bar 52
 - title bar 47
 - Variants table..... 51
- navigation of..... 53

- Global View pane
 - GeneMarker HTS Viewer..... 48

L

- layout of the GeneMarker HTS Viewer..... 46
 - Global View pane..... 48
 - main menu 47
 - Pane displays..... 49
 - status bar 52
 - title bar 47
 - Variants table..... 51

M

- main menu
 - GeneMarker HTS Viewer..... 47
- main window for GeneMarker HTS..... 15
- motif
 - creating or modifying..... 30
 - defined 29
- motif file
 - format for..... 29

N

- navigating the GeneMarker HTS Viewer..... 53
- New Project dialog box
 - overview of..... 27

O

- Online Help button on the GeneMarker HTS main window ... 16
- Organization of the manual..... 8
- output files for a GeneMarker HTS project 43

P

- Pane displays
 - GeneMarker HTS Viewer..... 49
- personal health information
 - defined 33
 - masking..... 33
- PHI
 - see personal health information
- project
 - adding new in GeneMarker HTS..... 36
 - output files for..... 43
 - working with an existing one in GeneMarker HTS..... 40

- Project Launcher options on the GeneMarker HTS main window ... 16

S

- Sample to Sample Comparison pane in the GeneMarker HTS Comparison Viewer..... 66

| | |
|-------------------------------|----|
| settings | |
| specifying for GeneMarker | |
| HTS..... | 17 |
| soft-clipping | |
| defined | 33 |
| Special information about the | |
| manual | 7 |
| status bar | |
| GeneMarker HTS Viewer..... | 52 |

T

| | |
|----------------------------|----|
| title bar | |
| GeneMarker HTS main | |
| window | 15 |
| GeneMarker HTS Viewer..... | 47 |

U

| | |
|-----------------------|---|
| Using the manual..... | 7 |
|-----------------------|---|

V

| | |
|-------------------------------------|----|
| variant | |
| adding in the Pile-Up pane for the | |
| GeneMarker HTS Viewer..... | 54 |
| deleting from the Variants table in | |
| the GeneMarker HTS Viewer..... | 54 |
| Variant Comparison pane in the | |
| GeneMarker HTS Comparison | |
| Viewer..... | 68 |
| Variants table | |
| GeneMarker HTS Viewer..... | 51 |