Tablet Documentation

Release 1.21.02.08

Information
Computational Sciences, The James Hutton Institute

May 12, 2021

Getting started

For more information on Tablet, or to download the application, please visit the Tablet web site.

CHAPTER 1

Quickstart

Here is a brief set of instructions to get you up and running with Tablet in the shortest possible time.

Tablet is designed for the visualization and exploration of sequence assemblies, on data sets containing just a few to many millions of reads.

1.1 Opening assemblies

Tablet currently supports importing assembly data from either ACE, AFG, BAM, MAQ, SAM or SOAPAligner/soap2 file formats. There are two ways to get this data into Tablet:

- By clicking the `Open Assembly` button in the Data section of the ribbon menu. This will display the *Open Assembly* dialog.
- By dragging and dropping the file (or files) directly into Tablet.

If you are trying Tablet for the first time and don't have any assembly data readily available you can load in example datasets from the *Open Assembly* dialog.

Once the data is loaded, a list of all the contigs found within the assembly will be shown in the `Contigs Browser` down the left-hand side. This can be used to select a contig for display, or to filter the list down to a smaller size via a range of criteria.

1.2 Tablet overview

Tablet's visualizations are split into several areas. The main display provides a visualization of a single contig at a time, with reads aligned against their consensus sequence. Tablet will lay out the data in either packed (showing as many reads per line as possible without overlap) or stacked (showing one read per line) formats. Paired-end variants are available for both

The read data is supplemented with the consensus sequence and its quality scores, coverage information (per base) and up to six consensus to protein translations (3 reading frames, forward and reverse). All of this information is mapped

to a scale bar that shows the current position within the contig. The position is listed twice; giving its padded and unpadded ([x] U[x]) values, along with coverage information for the base currently under the mouse (CV[x]).

The *Overview Panel* located above the consensus can display either a scaled-to-fit summary of all the reads in a contig, or a coverage graph showing average read coverage across the contig. Toggle between the views by right-clicking on the `Overview Panel`, or by using the `Show/Hide Overview` button on the ribbon.

1.3 Browsing the data

You can move the view around the data by using the scrollbars, or by clicking on the canvas and dragging with the mouse.

An alternative method of moving around the data is to click and drag with the mouse on the `Overview Panel`. The red rectangle drawn on the `Overview` represents the area of the data that the main display currently showing. Drag it with the mouse to quickly move anywhere within the current contig.

Zoom in or out using the `Zoom slider` located on the ribbon bar. You can also double-click on main display to zoom into that area. By adjusting the `Variants slider` you can modify the intensity of read bases that differ from the consensus to highlight potential variants in the contig.

1.4 Interacting with Tablet

Notice that as you move your mouse over the display, information on the read under the mouse is displayed in a tooltip. This includes its name, padded and unpadded start and end points, as well as the padded and unpadded length of the read. The read's orientation is also displayed as a graphical arrow (green for forward/uncomplemented and blue for reverse/complemented). The tooltip also provides a scaled-to-fit graphical representation of all the read's bases.

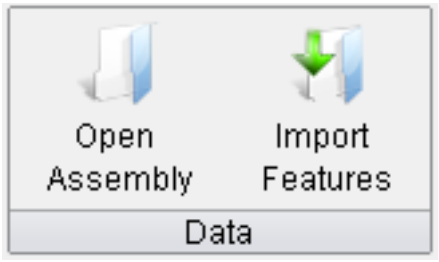
Right clicking with the mouse on many of the display components will open up additional menus showing options to change the display types, highlight regions of interest, copy data to the clipboard, jump to a read's pair if it is a paired read, etc.

CHAPTER 2

Ribbon Bar

2.1 Home: Data

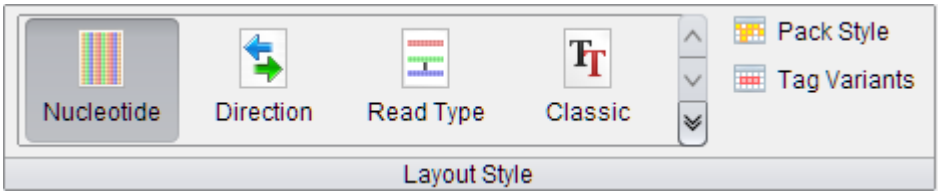
The `Data Panel` of the ribbon menu contains controls for importing data into Tablet.



- **Open Assembly** Select this option to display the *Open Assembly* dialog which allows a number of different assembly formats to be opened for viewing in Tablet.
- **Import Features** Select this option to display the *Importing Features* dialog which allows GFF3 formatted features to be imported into Tablet.

2.2 Home: Layout Style

The `Layout Style Panel` contains controls which alter the display of reads within Tablet.



- **Nucleotide** Nucleotide (formerly Enhanced) colouring displays read bases using a nucleotide colouring scheme. Any variants are shown in a lighter colour, with red (rather than black) text.

- **Direction** Direction colouring displays reads using a colour scheme that colours the read according to whether it was read on the forward or reverse strand.
- **Read Type** Read Type colouring displays reads using a colour scheme that colours the read according to whether it is a single end read, the first read in a pair, or the second read in a pair.
- **Classic** Classic colouring provides a simple black on white text scheme, but enhanced slightly to help distinguish reads from the background. Variants are shown in red.
- **Pack Style** Displays a menu with the following options:
  - **Packed** Changes the display so that reads are organised into a `packed` layout, where multiple reads are placed on the same row, if there is room to do so.
  - **Stacked** Changes the display so that reads are organised into a `stacked` layout, where every read is positioned on its own row.
  - **Packed (Paired End)** As the `packed` layout but instead of placing multiple reads on the same row, it places multiple pairs on the same row.
  - **Stacked (Paired End)** As the `stacked` layout but instead or placing each read on its own row, it places pairs of reads on their own row.
- **Tag Variants** Changes the display so that bases which do not match the reference base for their base position are coloured in red on the scaled data overview.

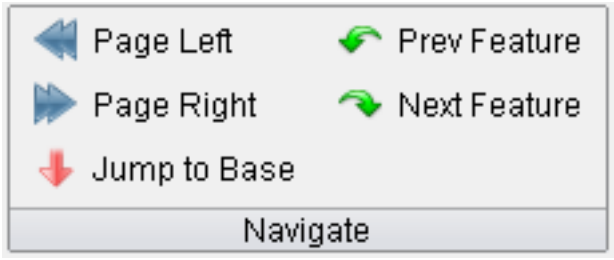
2.3 Home: Adjust

The `Adjust Panel` contains slider controls for adjusting the level of display zoom and the intensity of variant highlighting.

- **Zoom** Adjusts the zoom level of the main display
- **Variants** Controls the intensity of the highlighting of read bases that are different from the consensus at that position (variant bases).

2.4 Home: Navigate

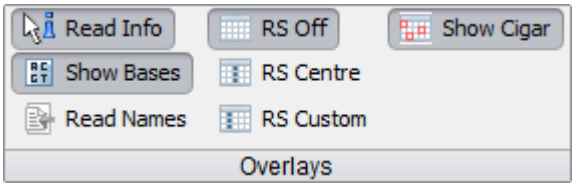
The `Navigate Panel` offers controls that aid in navigation.



- **Page Left** Jumps the view one screen's worth to the left.
- **Page Right** Jumps the view one screen's worth to the right.
- **Jump to Base** Opens the *Jump to Base* dialog, which can be used to jump the view to a specific base (padded or unpadded.
- **Prev Feature** Moves to and highlights the previous feature in the feature panel's list of features.
- **Next Feature** Moves to and highlights the next feature in the feature panel's list of features.

2.5 Home: Overlays

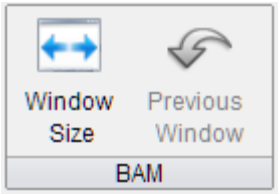
The `Overlays Panel` contains several controls for adjusting which overlays are visible when using Tablet.



- **Read Info** Toggles the display of popup information about the read currently under the mouse.
- **Show Bases** Toggles the display of text for bases in the display.
- **Read Names** Overlays the names of reads on the display. This is only available in the stacked view.
- **RS Off** Turns off the read shawowing line and read shadowing. Read shadowing highlights the reads which intersect the read shadowing line.
- **RS Centre** Enables read shadowing, with the line locked to the middle of the display.
- **RS Custom** Enables read shadowing, with the line initially following the mouse. In `RS Custom` mode it is also possible to lock the read shadowing line to a base position.
- **Show CIGAR** Overlays CIGAR insertion, deletion, skipping, and clipping events on the reads display. This feature automatically turns itself off when Tablet's zoom level is such that a base takes up less than a pixel on screen.

2.6 Advanced: BAM

The `BAM Panel` contains controls for adjusting - and controlling - the display of BAM data.



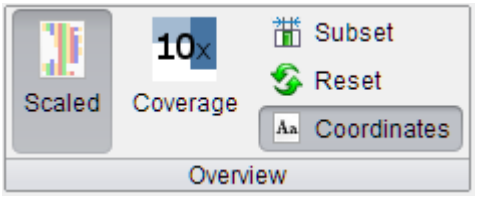
- **Window Size** Opens the *Adjust BAM Window Size* dialog, which can be used to alter the number of base positions a BAM window will display.
- **Previous Window** Moves the BAM window back to its previous location - if a previous location exists.

2.7 Advanced: Proteins

The `Proteins Panel` contains controls for adjusting the display of protein translations. Each button toggles the display of one of the protein translation tracks. The top three buttons control the forward translations and the bottom three control the reverse translations.

2.8 Advanced: Overview

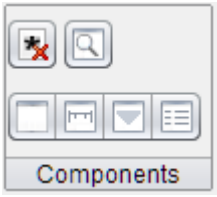
The `Overview Panel` contains controls for adjusting the display of the overview. This includes options to alter the type of overview display among others.



- **Scaled** Switches the overview display to a scaled data overview mode, which shows a scaled to fit version of the data.
- **Coverage** Switches the overview display to a coverage mode which displays the coverage across the entire dataset.
- **Subset** Opens the *Subset Overview Dialog* allowing the alteration of the start and end bases of overview drawing.
- **Reset** Resets the overview, undoing any subsetting which has been applied.
- **Coordinates** Toggle the display of the coordinates overlay. The coordinates overlay displays the start and end points of the current overview, as well as the start and end points of the current data window.

2.9 Advanced: Components

The `Components Panel` contains several controls for adjusting various miscellaneous options.



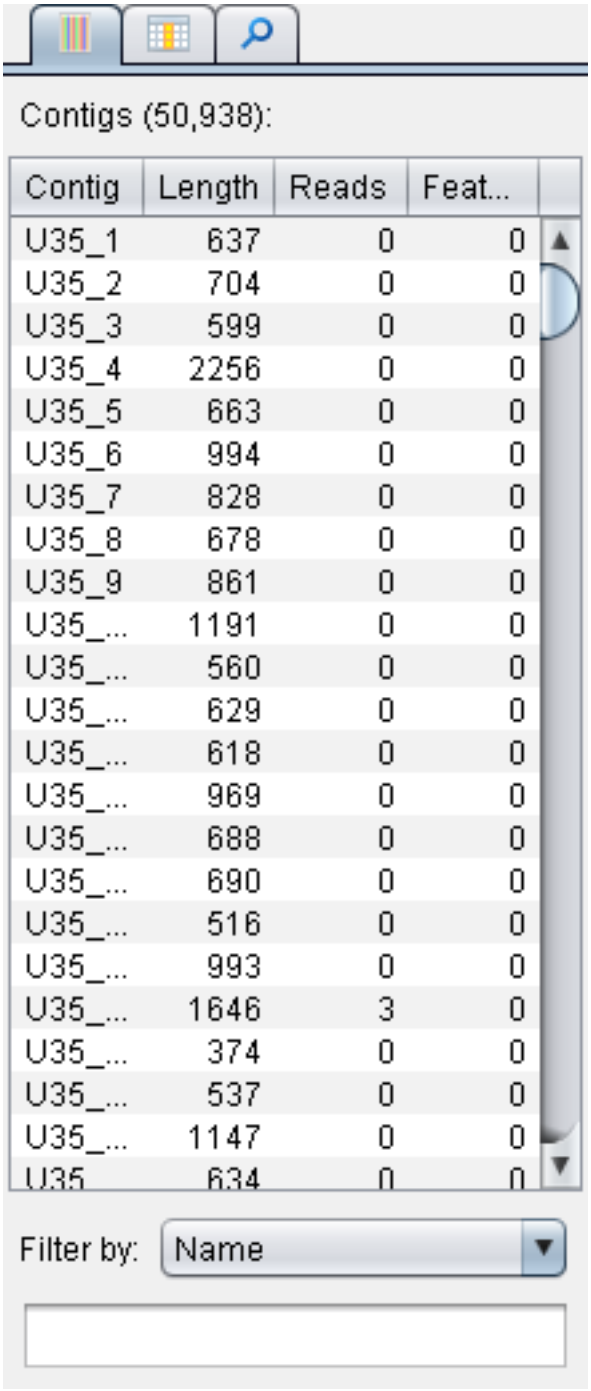
- **Hide Unpadded Values** Toggles whether or not unpadded values are shown alongside their padded counterparts in the various display components.
- **Show/Hide Overview** Opens a menu allowing you to toggle the Overview on or off, or to select which overview visualization to display.
- **Hide Consensus** Toggles on or off the Consensus display panel.
- **Hide Scale Bar** Toggles on or off the Scale Bar display panel.
- **Hide Coverage** Toggles on or off the Coverage display panel.
- **Hide Control Panel** Toggles on or off the Contigs/Features/Search Control panel.

CHAPTER 3

The Contigs Browser

The `Contigs Browser` panel displays a list of all the contigs in the current assembly. It lists each contig's name, length (of the consensus sequence), the number of reads within it, and the number of features associated with it (if any). Select any contig to visualize its data.

The browser also features a filtering mechanism, which allows the number of contigs shown in the list to be reduced. Filtering is possibly by name, minimum/maximum contig length, minimum/maximum number of reads, or minimum/maximum number of features.

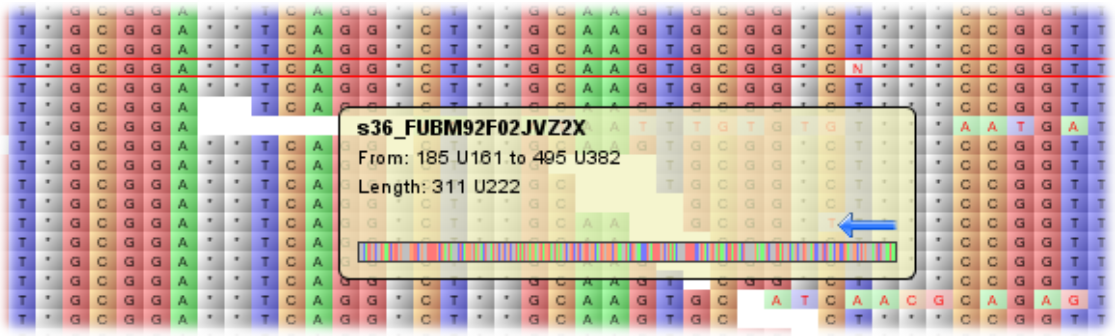

Contigs (50,938):
Contig
Length
Reads
Feat...
U35_1 637 0 0
U35_2 704 0 0
U35_3 599 0 0
U35_4 2256 0 0
U35_5 663 0 0
U35_6 994 0 0
U35_7 828 0 0
U35_8 678 0 0
U35_9 861 0 0
U35_... 1191 0 0
U35_... 560 0 0
U35_... 629 0 0
U35_... 618 0 0
U35_... 969 0 0
U35_... 688 0 0
U35_... 690 0 0
U35_... 516 0 0
U35_... 993 0 0
U35_... 1646 3 0
U35_... 374 0 0
U35_... 537 0 0
U35_... 1147 0 0
U35 634 0 0
Filter by:
Name

CHAPTER 4

Data Visualization

Tablet displays data on a per-contig basis. Select a contig using the *The Contigs Browser* to mark it for visualization.

The visualization is broken down into several areas, the most important being the display of read data.



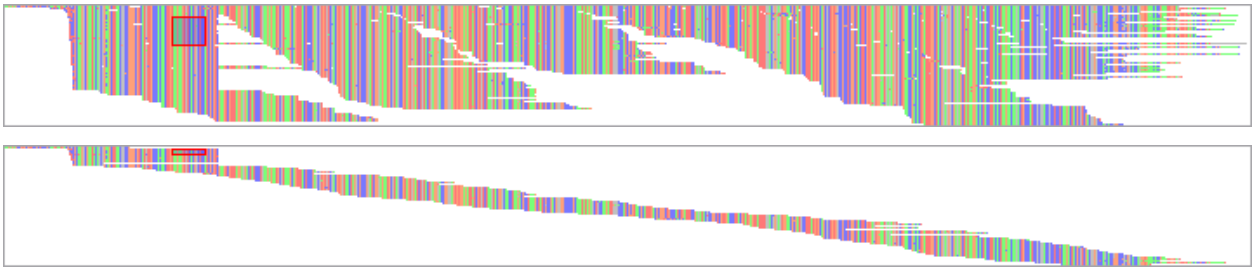
Each read is shown aligned against the consensus sequence, with its bases coloured according to the colour scheme selected. In `Nucleotide` mode, each DNA base is assigned its own colour. In contrast the `Classic` display removes colour information, and acts like more traditional text-based viewers. In `Direction` mode, each read is coloured according to whether it was read on the forward or reverse strand. In `Read Type` mode single end reads, the first read in a pair and the second read in a pair are each assigned their own colour (orange for single end, green for first in pair, blue for second in pair and red for orphaned reads). In all modes, `variant bases` - where the nucleotide in the read differs from the same base in the consensus - are displayed slightly brighter and with red rather than black text. The brightness value can be adjusted using the `Variants` slider on the *Ribbon Bar*'s `Adjust` tab.

When the mouse is over a read, a tooltip appears which displays information about that read, including its name, padded and unpadded start and end positions and the padded unpadded length of the sequence. Read orientation is displayed via an arrow and a graphic (scaled to fit the width of the tooltip) of all the sequence is also shown. The read currently under the mouse is also highlighted - in red, on the main display and in blue on the Overview (if showing the scaled-to-fit overview).

Right clicking on the display brings up a menu with the options to outline a row or column. To remove highlighting, choose clear all from the menu. There are also options to copy the name of the current read to the clipboard, or copy all of the read's data to the clipboard (its name, length, bases, etc). Finally, there are options which allow you to quickly jump the view to the start or end of a read.

4.1 Layout styles

Tablet can lay out the data in either packed (showing as many reads per line as possible without overlap) or stacked (showing one read per line) formats.



Toggle between the modes using the `Pack Style` on the `Layout Style` tab on the *Ribbon Bar*.

4.2 The scale bar

As the mouse tracks over the reads, its position is listed on the `Scale Bar`. This bar provides the column index for the left-most and right-most visible base, along with the value for the base currently under the mouse.

By default, position information is given twice - in padded and unpadded values, for example: 100 U95, meaning padded base 100, but unpadded base 95. The display of unpadded values can be toggled on or off using the `Hide Unpadded Values` toggle button located on the `Options` tab of the *Ribbon Bar*.

For the base under the mouse, coverage depth is also provided; for example C45, for a coverage value of 45.

4.3 Consensus/reference data

The `Consensus/Reference Panel` displays the consensus/reference sequence for the current contig, showing the sequence data for each base, along with a graphical representation of the base's quality information (if available). As with any of the reads, the consensus data can be copied to the clipboard by choosing the appropriate option from the popup menu that appears after right-clicking on it.

4.4 Coverage information

Per-column coverage information is provided by the `Coverage Panel`. Coverage is determined by looking for the presence or absence of read data, regardless of what the data is (or whether a read base matches the consensus base for that position). The height of the coverage bar over a particular column represents that column's coverage as a percentage of the maximum coverage for the entire contig.

CHAPTER 5

BAM Data Visualization

Tablet displays data on a per-contig basis. Select a contig using the *The Contigs Browser* to mark it for visualization.

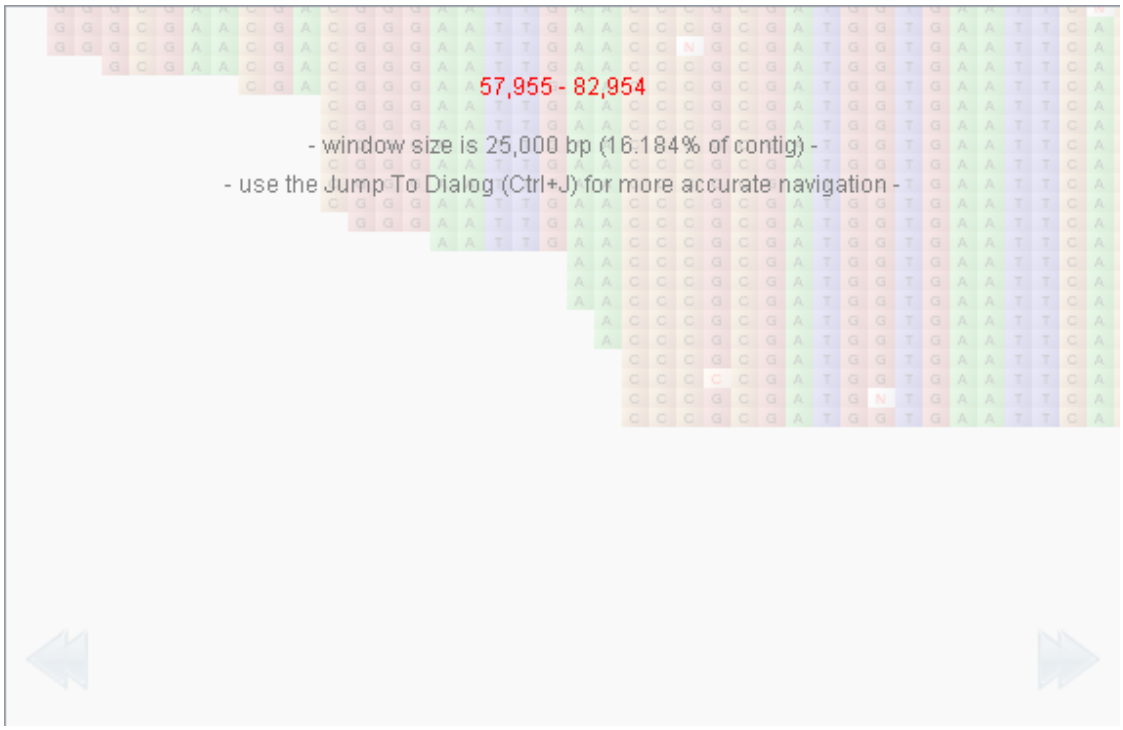
Tablet supports BAM indexed query for visualizing large alignments. Indexed query makes it possible to support visualization of alignments in the hundreds of gigabytes of data on disk. In this view, Tablet displays subsets of the data at a time; while still providing the ability to move easily through - and jump to any point within - the dataset.

The visualization of indexed BAM data is subtly different from the visualization of other assembly formats. While most of the display remains as described in *Data Visualization*, there are a few extra display components to aid the visualization - and navigation - of indexed BAM data.

5.1 The BAM bar

The BAM bar shows the position of the current data window in the overall BAM file. It also displays the size of the current data window relative to the size of the entire data set.

The BAM bar always displays the locations of the current (in white) and previous ("ghost") data windows. The white bar (representing the current data window) can be grabbed and moved to select another window of data of that size in the BAM file. Alternatively the BAM bar can be clicked on and white bar will move to this new location. While the white bar is being dragged, the reads canvas displays information to aid selection of a new data window. This includes the starting and ending bases of the new data window, the data window size and the data window size as a percentage of the entire data set.



The BAM bar also has a right click context menu. This menu contains an option for returning to the previous location in the data set and an option for adjusting the BAM window size using the *Tablet Options* dialog.

5.2 Changes to the reads canvas

The reads canvas remains largely unchanged from that described in *Data Visualization*. The main difference is that the area which can be scrolled through is defined by the BAM window size instead of the size of the entire data set. Paging through the data set is also altered. Within a BAM window, paging operates as normal, moving the display forwards or back by the number of bases which fit in the display. At the edge of a BAM window, the paging arrows change behaviour to loading the next BAM window size of data from the BAM file. When a paging arrow is going to do this its colour changes to green.

An addition to the tooltip which appears when the mouse is over to read is the inclusion of CIGAR information for that read. BAM uses an extended CIGAR format to encode information about reads. In BAM the reads are built up from the read sequence and the information in the CIGAR to display the final read.

5.3 Overviews

When viewing BAM alignments, data overviews are not of the whole data set as they are when viewing other assembly files. They are overviews of the data in the current BAM data window. Whole data set overviews are not available for BAM alignments. In all other regards the information on overviews in the *Overviews* section relates to BAM data overviews as well.

5.4 Contigs panel

The contigs panel displays the name, length, number of reads, number of features and the percentage mismatch for each contig. When viewing BAM alignments the percentage mismatch cannot be displayed until the data for a contig is loaded, and even then, it will only be the mismatch value for that particular BAM window (which may be smaller than the actual size of the contig).

5.5 Coverage printer

The coverage printer feautre is not currently available while viewing BAM files. The coverage printer prints out the coverage over each reference base. When viewing BAM files the whole data set is not available, as such the coverage printer is disabled when viewing BAM files.

CHAPTER 6

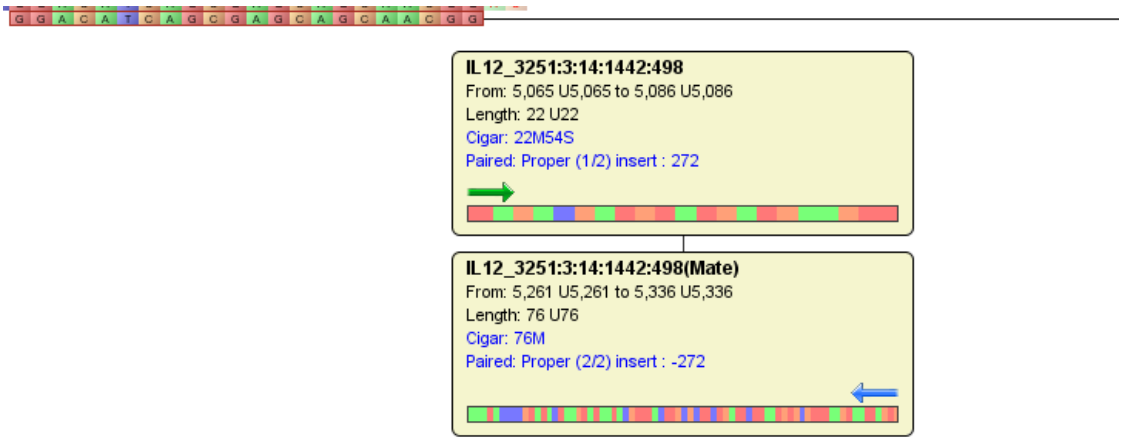
Paired-End Data Visualization

Tablet displays data on a per-contig basis. Select a contig using the *The Contigs Browser* to mark it for visualization.

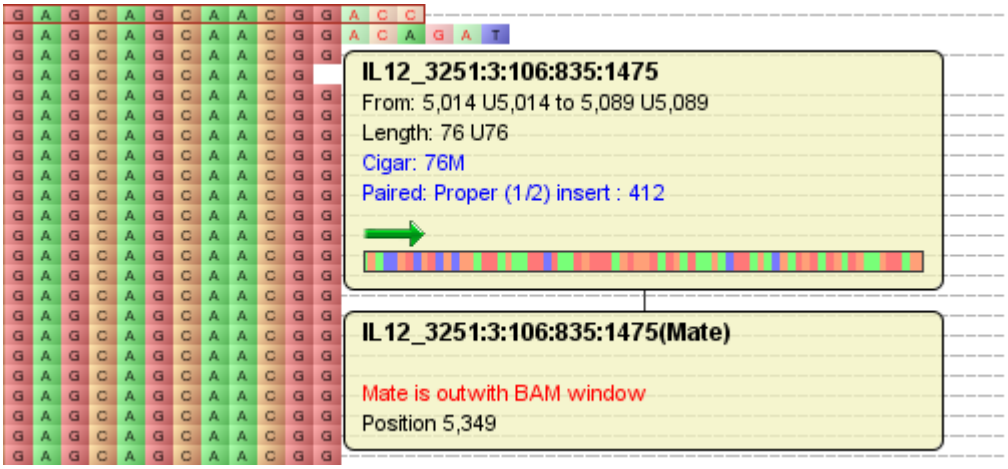
Tablet supports paired-end read data in both the BAM and SAM alignment formats. The visualization of paired-end read data is subtly different from that of single-end read data. While most of the display remains as described in *Data Visualization*, there are a few tweaks to the display components to aid the visualization of paired-end data.

When viewing single end data if the mouse is over a read, that read is highlighted by a red outline. When viewing paired-end data, the read's pair is also highlighted by a red outline. If the reads are on the same line of the display when in either of the paired-end viewing modes a dotted line is drawn between the two reads in the pair. When hovering over a read in the pair - or the line itself - the line becomes a solid black line.

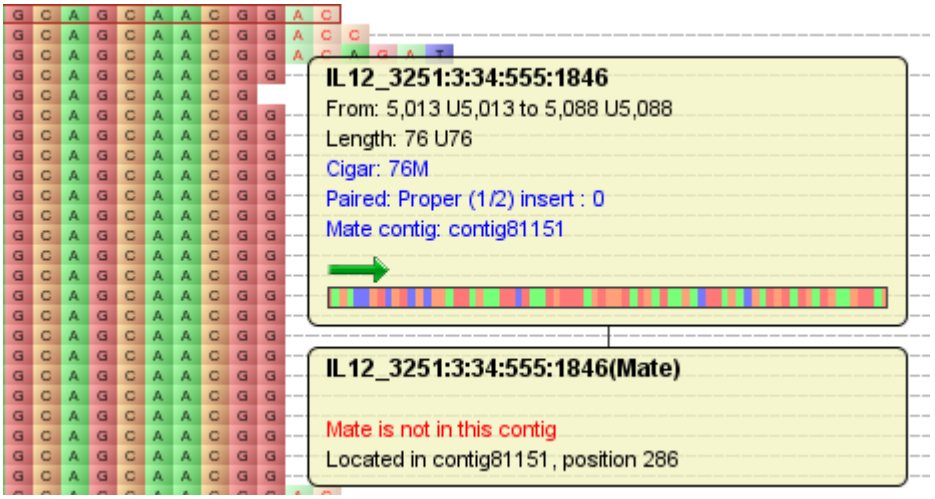
As with single-end data, when the mouse is over a read a tooltip appears which displays information about that read. Where this differs from single-end data is the information provided in the tooltip and in additional tooltips which can also be displayed. In addition to displaying a read's name, padded and unpadded start and end positions, the padded unpadded length of the sequence and the CIGAR information for the read. If the read is a paired-end read it will also display pair information including whether or not it is a proper pair, its number in the pair (1/2 or 2/2), and its mate's position. Read orientation is displayed via an arrow and a graphic (scaled to fit the width of the tooltip) of all the sequence is also shown. If the read is a member of a pair, information about its pair is displayed in an additional tooltip.



This tooltip displays the pair's equivalent of the original read's data if the pair is in the current BAM window, the following message "Mate is outwith BAM window" when the pair isn't in the current BAM window; plus the position of the pair.



Finally the following message "Mate is not in this contig" when the pair isn't in the contig; plus a message detailing which contig it is in and at which position.



6.1 Paired-End Pack Styles

In addition to the layout styles described in *Data Visualization*, two paired-end specific layout styles are available. These are Packed (Paired-End) and Stacked (Paired-End).

Tablet can lay out the data in either packed (showing as many pairs per line as possible without overlap) or stacked (showing one pair per line) formats.

Toggle between the modes using the `Pack Style` drop down menu on the `Layout Style` tab on the *Ribbon Bar*.

CHAPTER 7

Searching

To search for reads by name, open the search panel and select `Search for reads by name` from the first drop down list. The second drop down list determines whether or not the search will look through all contigs for matches, or just the currently loaded contig. Enter your search term in the combo box - this combo box stores your previous search terms, so it is easy to select a previous search - then click the search button. The table below the controls should fill in with any results. Click on a result to have that read highlighted in the main display.

7.1 Regular expressions

When searching for reads by name it is possible to specify that the search should use java regular expressions. When regular expressions aren't being used the search will only find results that match the search term exactly. When regular expressions are being used the search is attempting to match read names to the regular expression provided in the search combo box. As a simple example imagine some reads are prefixed with an identifier. To find all the reads prefixed with that identifier you would enter "prefix.*" (without the quotes) into the search box and carry out the search. Further details on the regular expression syntax used by Table can be found here.

7.2 Searching for subsequences

To search for subsequences of nucleotide data within either the reads, or the consensus / reference sequence select either `Search for subsequences` or `Search in consensus / reference` from the first drop down list. The second drop down list determines whether or not the search will look through all contigs for matches, or just the currently loaded contig. Enter your search term into the combo box and click the search button. Once the search is complete the table below will fill in with the results. Click on a result to jump to and highlight the subsequence in the read or consensus / reference.

7.3 Skipping pads

When searching for subsequences the option to `Ignore pads when searching` is available. This means that pads in the nucleotide sequences will be ignored when trying to match against the search string. Ns are also skipped when this option is selected, this means that for example the search term ACGT will match the following sequences:

A	C	G	T				
A	C	*	G	T			
A	C	*	*	*	*	G	T
A	C	N	G	T			

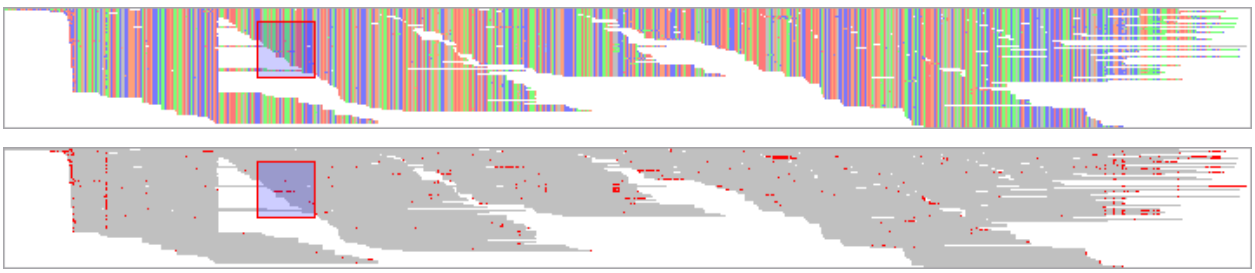
CHAPTER 8

Overviews

Tablet provides two different methods of visualizing overview information about a contig and its data - the `Scaled-to-Fit` overview and the `Coverage` overview, both selectable from the `Options` tab of the *Ribbon Bar*.

In either case, along with the overview visualization itself, the `Overview Panel` also provides a means for fast navigation within a contig, simple by clicking and dragging its red viewing rectangle. This rectangle represents the portion of the entire contig that is currently visible within the main display area.

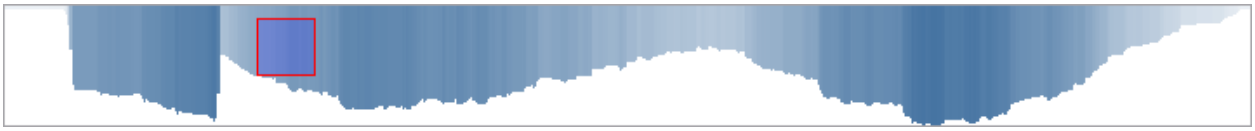
8.1 Scaled-to-fit overviews



This overview type attempts to fit all of the read data from the current contig into the available window size. This type of view may be less useful on larger contigs or contigs where read coverage is very sparse, as the proportion of empty space to read data obviously tends to favour the former.

The colour scheme in use (`Enhanced` or `Classic` will affect how this overview is rendered too.

8.2 Coverage overviews



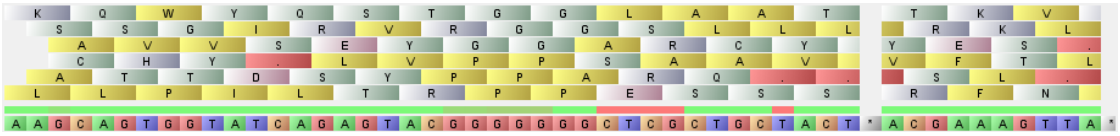
This overview displays average coverage depth over the entire contig as a histogram. The bar depth represents the average coverage for that region as a proportion of the maximum coverage for the contig. Colour intensity is used to show how the maximum depth within that averaged region relates to the overall maximum too, with darker shades representing areas with deeper coverage.

A short bar, but with very dark colouring can mean that although the average coverage for that region may be very low, it must have some columns where the coverage is very deep. These areas often warrant closer inspection.

CHAPTER 9

Protein Translations

By default, the `Protein Translation` panel displays a translation of the consensus sequence data for the first reading frame (5' to 3'). It is also capable of displaying all three forward translation frames as well as the three reverse translation frames. These can be displayed in any combination, either by right clicking on the panel itself, or by clicking the `Show/Hide Protein Translations` button on the `Options` tab on the *Ribbon Bar*.



When the mouse is over a protein, its full name will be displayed on the `Scale Bar`. Translation information can also be copied to the clipboard, again by using the panel's menu options.

CHAPTER 10

Importing Features

Select `Import Features` from the *Ribbon Bar* to import a GFF3 formatted file containing additional feature information that you want to associated with an assembly.

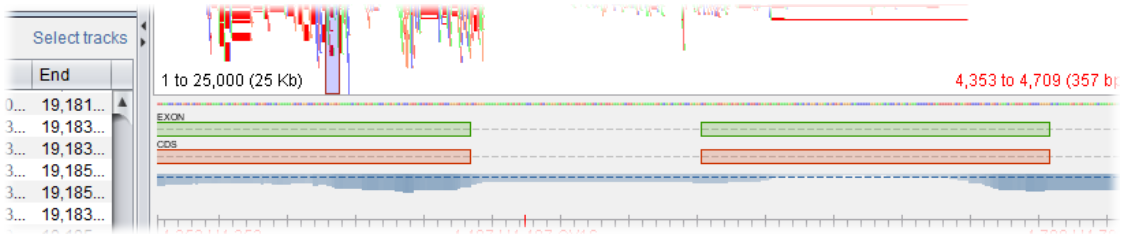
Once a feature file has been loaded, the `Features Browser` tab will become active whenever a contig is selected that contains features.

Clicking on a feature will automatically move the view to the column where the feature starts. It will also be highlighted for a few seconds. (If you wish to outline the column more permanently, right-click it and select `Outline > Outline Column` from the popup menu.)

Tablet can treat feature positions as either padded or unpadded depending on whether the `Feature values are padded` checkbox is selected or not. This setting determines where in the consensus sequence a feature will start and end, with padded values counting pad (*) characters and unpadded values ignoring them.

10.1 Visualizing features

Once features have been imported, Tablet will automatically select and display tracks for the first three feature types found.



The active tracks can be enabled, disabled, or reordered using the `Select Tracks` option located within the `Features` tab.

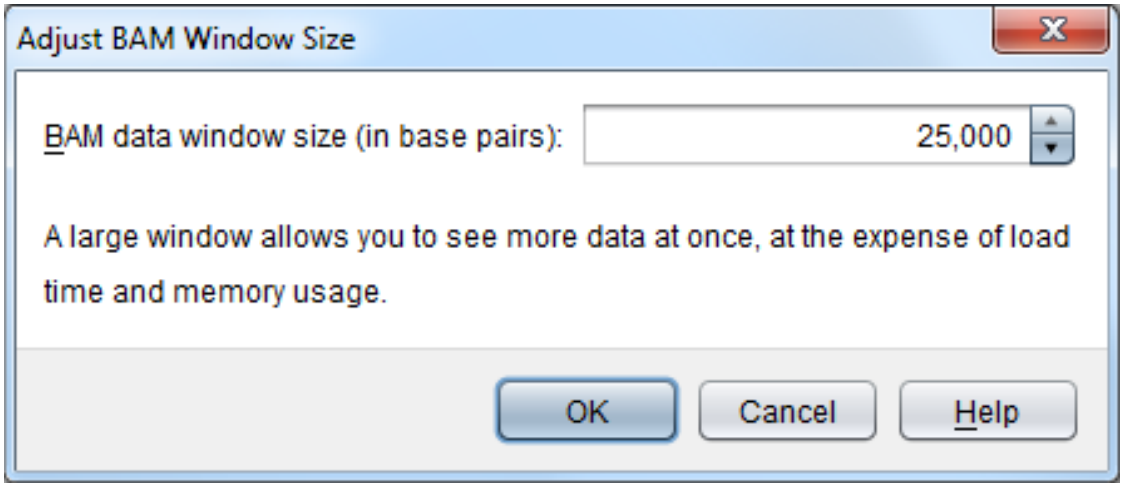
Each unique type of feature (eg, SNP, INDEL, etc) as defined in the GFF file will be rendered using a distinct colour.

Please note that this functionality is still experimental, and is subject to change. Tablet currently makes no attempt to link features, or to separate features with overlapping positions onto new tracks.

CHAPTER 11

Adjust BAM Window Size

The `Adjust BAM Window Size` dialog is used to adjust the total amount of data (known as the BAM window size) that Tablet will load/hold at once when viewing a BAM assembly.



The default size is 25,0000 bases. This means that for any contig viewed, you will only see up to 25,000 bases at once, regardless of the actual size of the contig. See *BAM Data Visualization* for more details.

If you change the size to be either larger or smaller than the current value, then Tablet will reload its view of the selected contig, either growing or shrinking the view to fit. A larger window obviously allows you to see more data at once, but at the expense of the time it takes to load that data from the BAM file, and the additional memory required to hold any reads found.

Ideally, you should set a BAM Window Size that matches what you know about your data. With a low coverage data set, you will be able to set a much larger window size (and maintain fast loading, with low memory usage) than with a deep coverage data set.

CHAPTER 12

Application Menu

The `Application Menu` is accessed by clicking the Tablet icon (the red globe) located on the left-hand side of the *Ribbon Bar*. The menu contains the following options.

12.1 Open

Displays the *Open Assembly* dialog, which prompts you for the files required to open your assembly data. Hovering the mouse over the option also displays a list of recently accessed assemblies.

12.2 Save

Save is a placeholder option for possible future functionality.

12.3 Save As

Save As is a placeholder option for possible future functionality.

12.4 Export Image

Allows you to export the current view (reads visualization) as a PNG image saved to disk.

12.5 Export Coverage Summary

Allows you to export a summary of the coverage across all contigs of the currently loaded assembly. This exports a list of contigs with the coverage value for each base of the consensus. There is an additional config file (called printPads)

which allows you to choose if padded bases have their coverage output. This defaults to true.

12.6 Close

Closes any currently open assembly and returns to the Tablet front page.

12.7 Additional options

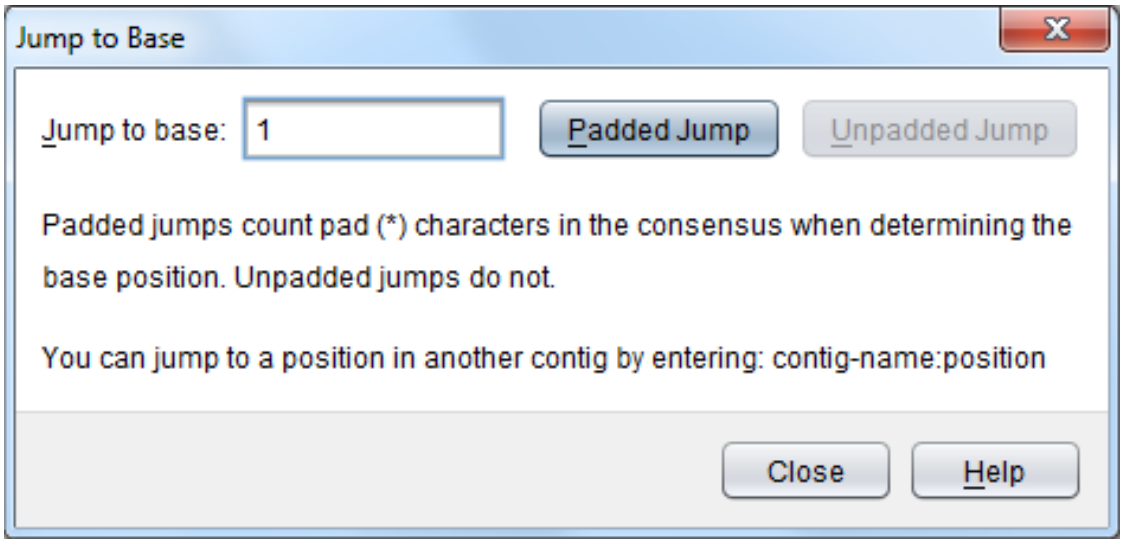
The following additional options are also provided by the Application Menu (although only `Exit Tablet` is provided when running under OS X - `About Tablet` and `Tablet Options` can both be found on the Tablet system menu).

- **About Tablet** Displays an information dialog providing Tablet version, copyright and author details.
- **Tablet Options** Displays the *Tablet Options* dialog that allows you to adjust various Tablet settings.
- **Exit Tablet** Closes any current assembly and exits the application.

CHAPTER 13

Jump to Base

The `Jump to Base` dialog is used to quickly jump to a specific base in the assembly.



To jump to a base, enter its number, then perform either a padded, or unpadded jump. Padded jumps count pad characters in the consensus sequence, whereas unpadded jumps ignore any padding. Note that sequences displayed in Tablet always count their first base as position "1" (rather than "0").

The following table provides a short example of mapping between padded and unpadded values.

Base	A	C	*	G	T
Padded position	1	2	3	4	5
Unpadded position	1	2	-	3	4

13.1 Jumping to a position in another contig

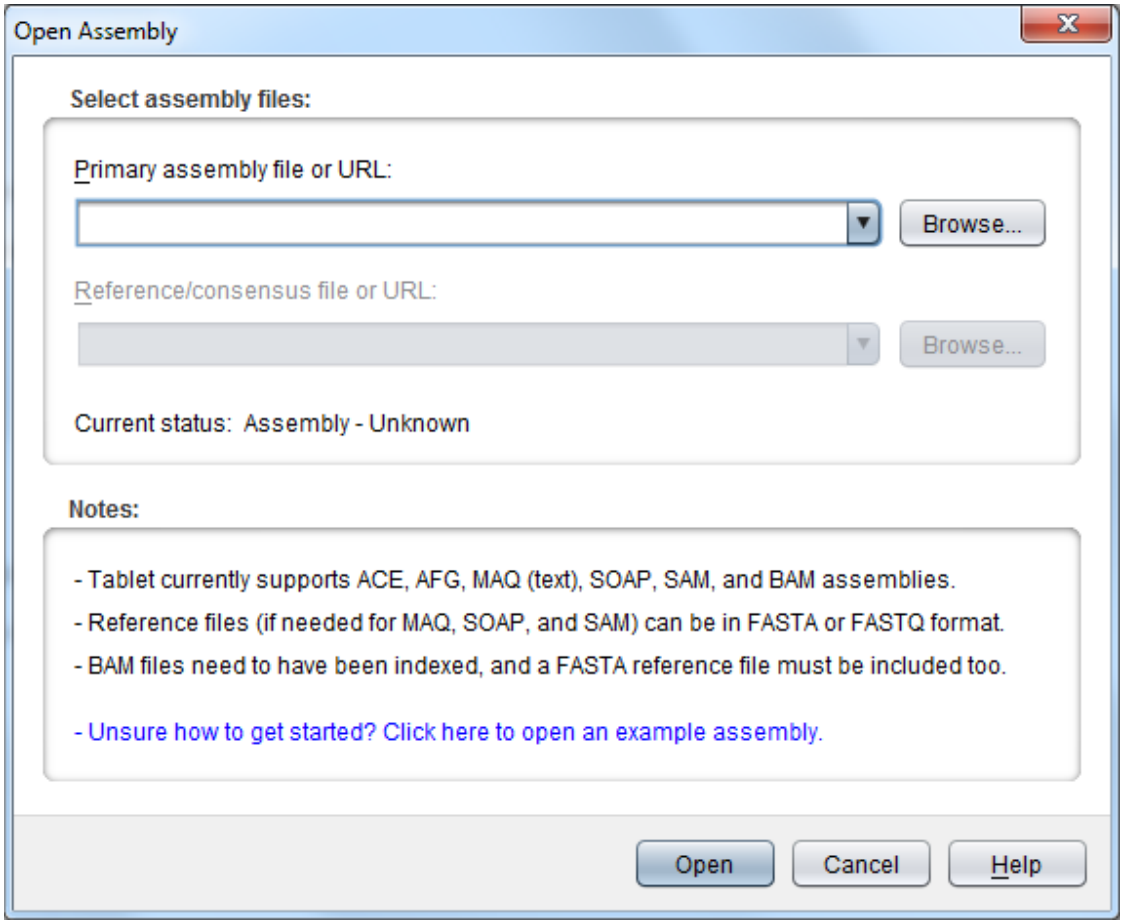
It is also possible to jump to a base position in a contig other than the contig you are currently viewing. To jump to a base in another contig instead of simply entering the base position, enter the contig name and the base position separated by a colon e.g. `Contig1:200`.

CHAPTER 14

Open Assembly

The `Open Assembly` dialog is used to load assemblies into Tablet for viewing.

Access the dialog by selecting either `Open Assembly` from the `Data` tab of the *Ribbon Bar* or by selecting `Open` from the *Application Menu*.



Tablet can currently view assemblies/alignments which are stored in the ACE, AFG, MAQ, SOAP, SAM or BAM file formats, with accompanying reference/consensus data (if needed) being read from a FASTA file.

The `Primary assembly file` refers to the main file containing your assembly or alignment data. The `Reference/consensus file` refers to any additional reference file that contains reference/consensus data and is needed with assembly formats that do not include this data in the primary file (such as SAM, BAM, MAQ, and SOAP). This additional data is not needed by Tablet, but it is advisable to include it if you have it, otherwise Tablet cannot provide a visualization of the reference/consensus sequence within each contig.

Tablet can load files locally from disk, or remotely from a web server. The files can either be uncompressed or compressed with gzip. (BAM files are already compressed, and should be provided as is).

14.1 Importing ACE files

An ACE formatted file includes information on each contig, its consensus sequence, and the reads that are aligned against it. A single ACE file provides all the information that Tablet requires.

14.2 Importing AFG files

An AFG formatted file includes information on each contig, its consensus sequence, and the reads that are aligned against it. A single AFG file provides all the information that Tablet requires.

14.3 Importing MAQ files

The MAQ assembler ultimately generates a binary-formatted map file (with a .map extension). To be readable by Tablet, this file must be converted into a text-based file containing the read information.

Using the command line MAQ assembler tools, run: `maq mapview input.map > output.txt`. This file must be provided to Tablet.

14.4 Importing SOAPAligner output

The output from the SOAPAligner mapping tool is a text-based alignment file that includes the read data only. A separate FASTA formatted file containing the reference sequence(s) can be provided separately but this is optional. SOAPDenovo output is currently not supported.

14.5 Importing SAM files

Tablet will attempt to load a text-based .sam file, with or without SAM headers. A SAM file does not include reference/consensus information, so if it is to be included, it must be provided in a separate fasta/fastq file.

When processing the CIGAR information from a SAM file, Tablet will create a list of features (that will be shown in the Features Table) for each CIGAR insertion, deletion, skip, and clip event that is found. Each feature gives the position of the insertion and the number of reads that have an insertion at that position.

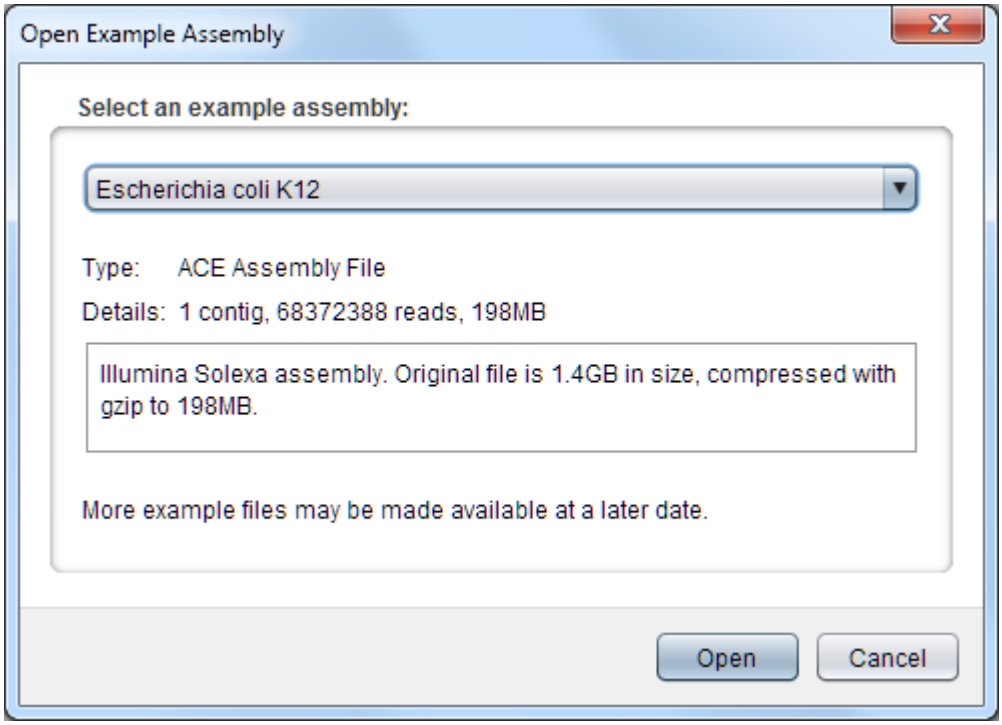
14.6 Importing BAM files

Tablet supports BAM in its native (indexed) format. It is important to note that the BAM file must be sorted and indexed, with an associated .bai file located in the same directory as the .bam file (named either <assembly_name>.bam.bai or <assembly_name>.bai).

When processing the CIGAR information from a BAM file, Tablet will create a list of features (that will be shown in the Features Table) for each CIGAR insertion, deletion, skip, and clip event that is found. Each feature gives the position of the insertion and the number of reads that have an insertion at that position.

14.7 Importing example data

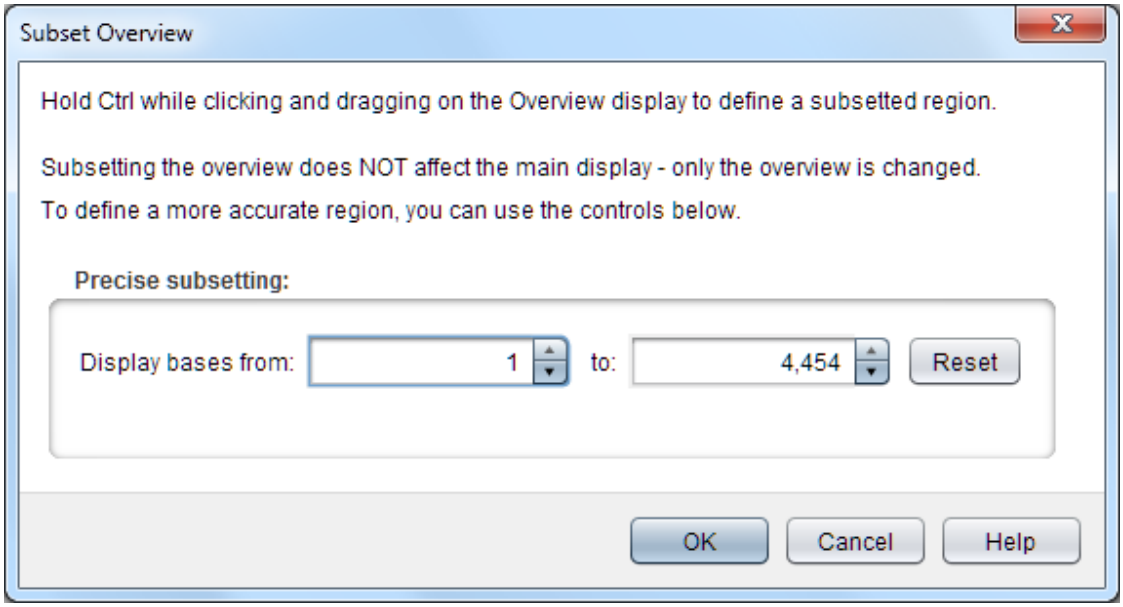
The `Unsure how to get started? Click here to open an assembly.` link brings up the dialog to load example datasets. This contains a drop down menu which allows for selection between the various example datasets available. Below the drop down menu is a description of the currently selected example dataset. To open a dataset select it from the drop down menu and click the `Open` button.


Open Example Assembly
Select an example assembly:
Escherichia coli K12
Type: ACE Assembly File
Details: 1 contig, 68372388 reads, 198MB
Illumina Solexa assembly. Original file is 1.4GB in size, compressed with gzip to 198MB.
More example files may be made available at a later date.
Open
Cancel

CHAPTER 15

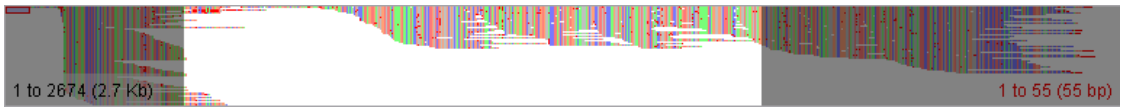
Subset Overview

The `Subset Overview` dialog is used to adjust the start and end points of the overviews that Tablet displays. This allows you to specify a subset of the data set which you would like to see represented in the overview.



The default values for the *display bases from* and *to* are the start and end points of the currently displayed contig. If you alter the subset values, you should ensure that the from value should always be lower than the to value. If it is not, the overview will be reset to its original state.

The `Subset Overview` dialog allows for very precise subsetting of the overview, however it is also possible to quickly subset the overview from within the main tablet window. To do this you click and drag on the overview canvas with the correct keyboard modifier for your platform (CTRL for Windows/Linux, CMD for OSX). The point of the first click is the start point for the new subset and the point of release of the mouse is the end point.



When subsetting the overview in this manner the area that will be visible in the new overview remains the same colour, whereas areas which won't be visible in the new overview are greyed out.

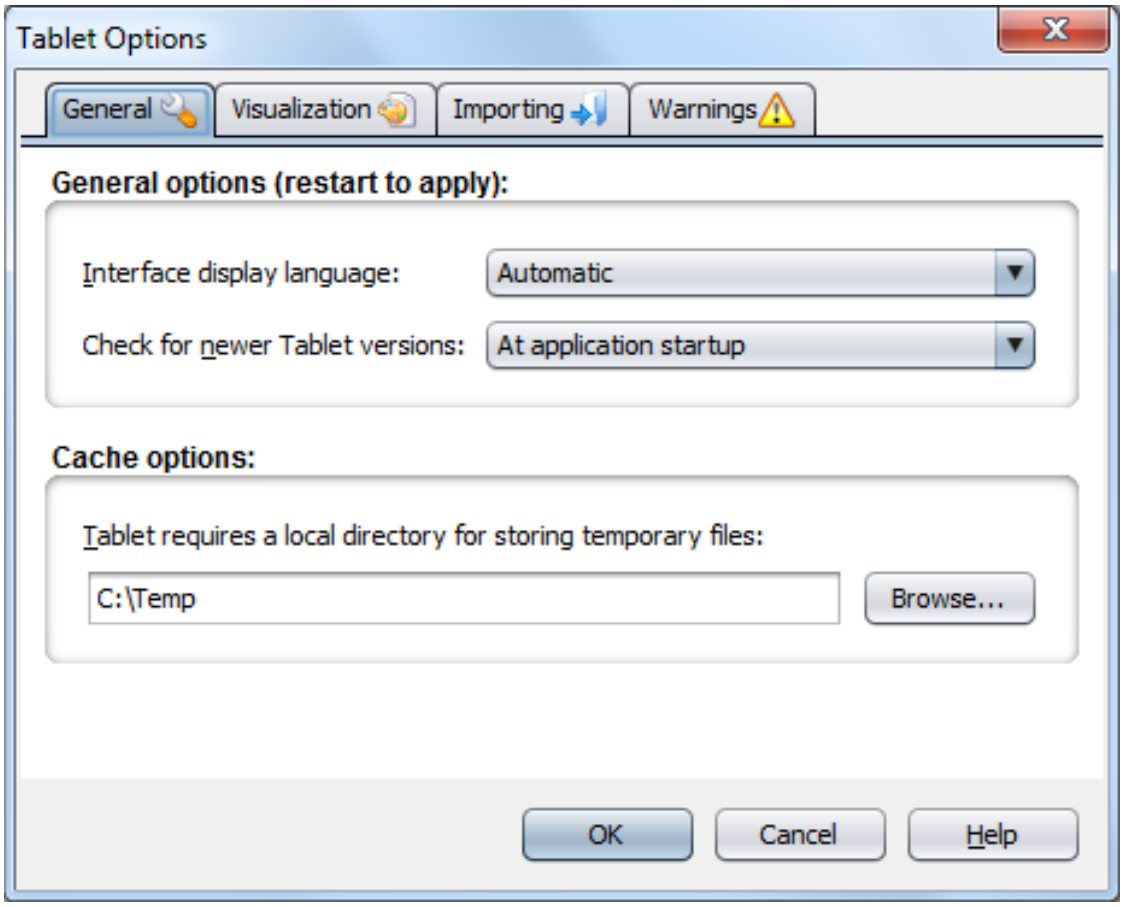
CHAPTER 16

Tablet Options

The Preferences dialog can be used to modify various settings that affect the way Tablet is used and is displayed.

16.1 General

The `General` tab includes options which are specific to Tablet's general operation.



General options (restart to apply):

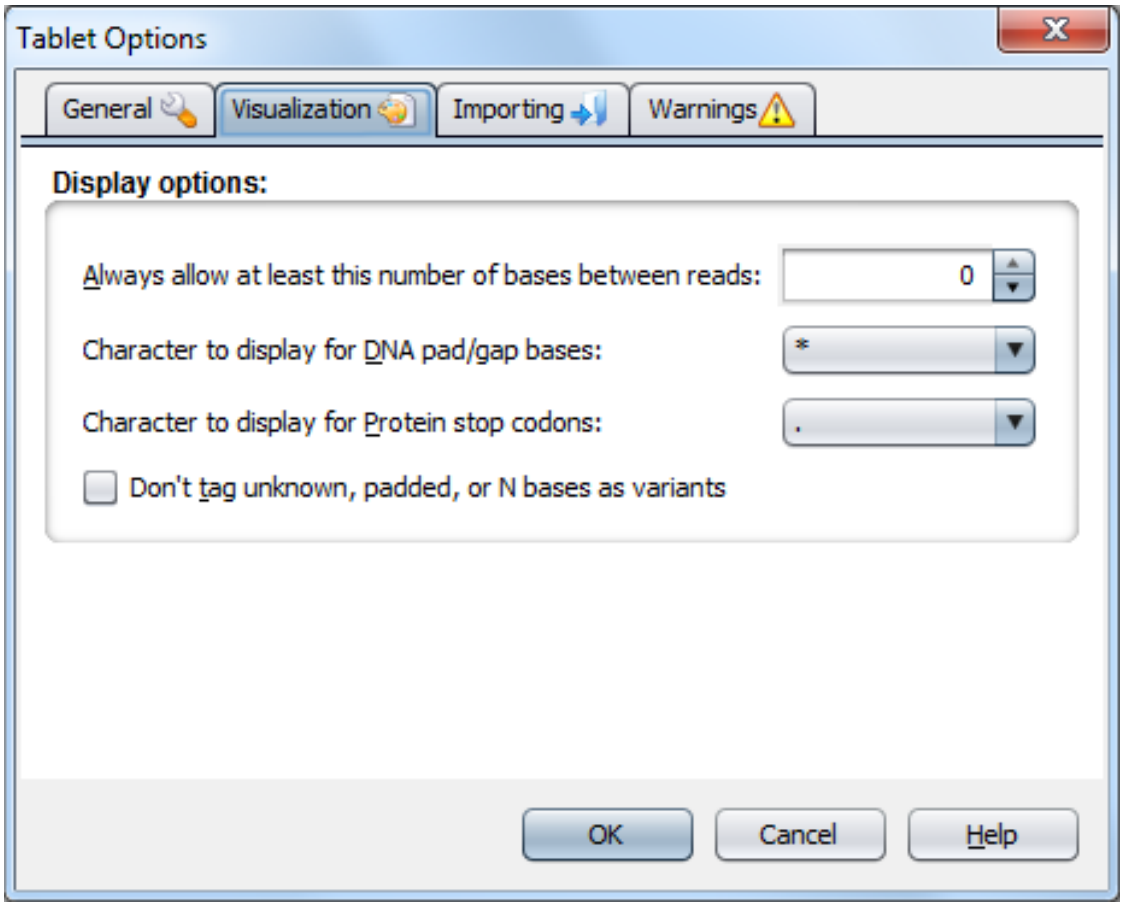
- **Interface Display Language** This setting determines what language Tablet will display its user interface in.
- **Check for newer versions of Tablet** This setting determines how often Tablet will attempt to connect back to its download server to see if a newer version is available. The options available are Never, At application startup, Once a day, Once a week, and Once a month.

Cache options:

- **Tablet requires a local directory for storing temporary files** Specifies the directory where Tablet will keep its disk cache (of the assembly data). The size of the cache depends on the size of the assembly, so it is advisable to pick a location with plenty of free space.

16.2 Visualization

The `Visualization` tab allows you to customise elements of Tablet's visual display.

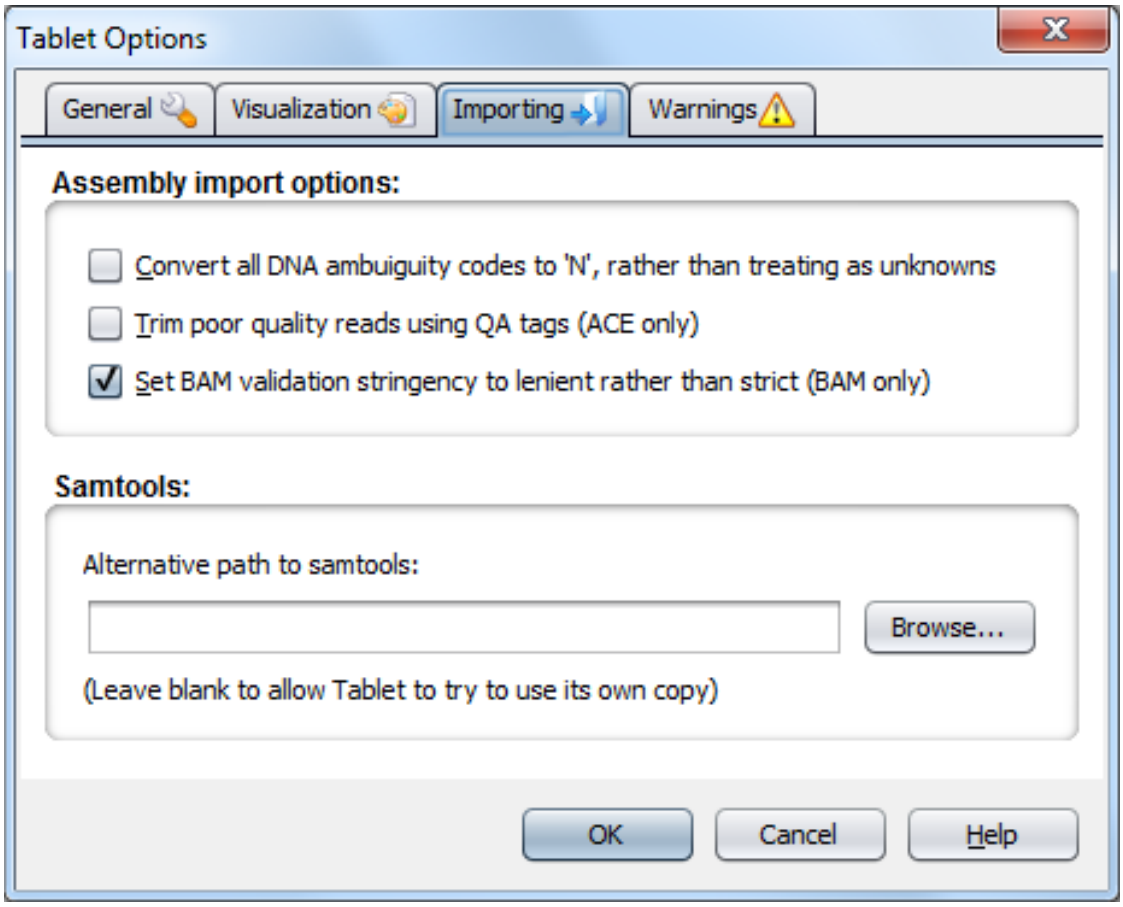


Display options:

- **Always allow at least this number of bases between reads** Forces Tablet to ensure that adjoining reads are always separated by at least the number of bases specified. As this affects how the data are packed, reopen the current contig to see any changes.
- **Character to display for DNA pad/gap bases** Determines what character will be shown whenever a padded base is displayed.
- **Character to display for Protein stop codons** Determines what character will be shown whenever DNA translates to a stop codon.
- **Don't tag unknown, padded, or N bases as variants** If selected, Tablet will no longer render these bases as variants, meaning they will not stand out from the rest of the data (or be highlighted in red).

16.3 Importing

The `Importing` tab allows you to modify how Tablet reads from data files.



Assembly import options:

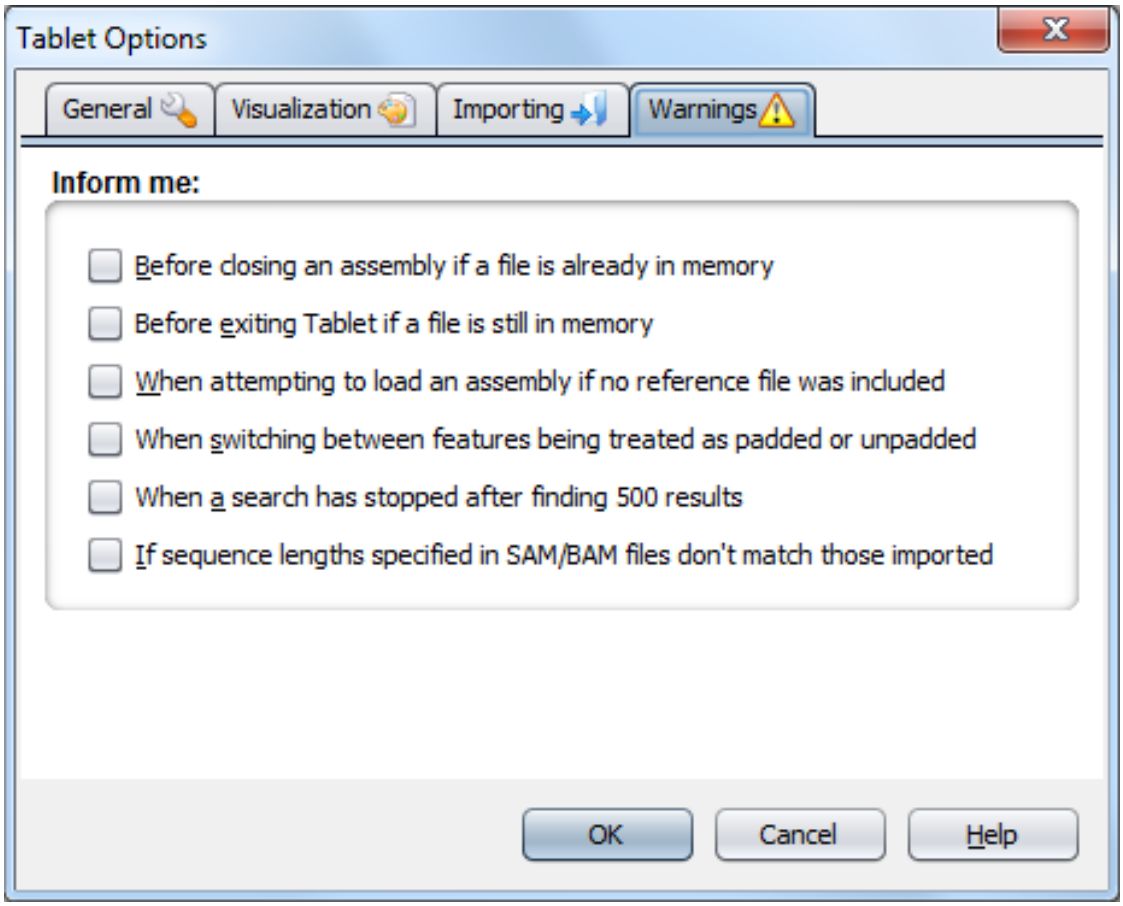
- **Convert all DNA ambuiguity codes to 'N', rather than treating as unknowns** If selected, any nucleotides encoded using ambuiguity codes will be read as N rather than being treated as unknown (Tablet has no support for ambuiguity codes). This will affect how the protein translations are displayed.
- **Trim poor quality tags using QA tags (ACE only)** Affecting the import of ACE files only, if selected, this option will cause Tablet to parse the QA tags found within ACE files. These tags cause reads to be trimmed for quality. The resultant read once loaded into Tablet may be shorter than if the option is left unselected and regions of low quality will no longer be displayed.
- **Set BAM validation stringency to lenient rather than strict (BAM only)** Affecting the import of BAM files only, if selected, Tablet will ignore any errors often associated with malformed BAM assemblies and will attempt to display the file as if it was OK. Any errors that are found will be written to the output.log file in Tablet's install directory. See samtools/picard documentation for more details on strict/lenient parsing options.
- **Always cache read data to disk while importing** Selecting this option will ensure Tablet keeps the majority of read data for an assembly in a disk-based cache, rather than in memory. This marginally reduces performance, but has a significant saving on memory.
- **Bypass read disk caching for BAM files** Because BAM files are already indexed and allow for quick access, there is no need for Tablet to (re)cache them for its own use. The only time you may wish to disable this option is if you are running Tablet with a very large BAM window size and wish to further reduce memory usage.

Samtools:

- **Alternative path to samtools** When opening BAM files, Tablet attempts to use its own bundled copy of samtools to gather read count statistics from the index file (using the "samtools idxstats" command). Tablet's distributable versions may not work on every operating system, so you can use this option to provide the path to a known working version of samtools (0.1.8 or higher).

16.4 Warnings

The `Warnings` tab allows you to switch on or off the various warnings that Tablet can pop up from time to time. If an item is checked, Tablet will display a relevant warning to you whenever its associated action occurs. Unchecking an item means that Tablet will proceed with the action without warning you.



- **Before closing an assembly if a file is already in memory** This warning occurs if you attempt to close an open assembly file. Some files can be large, and take time to load, so Tablet will check that you really '"do"' want to close the file before doing so.
- **Before exiting Tablet if a file is still in memory** Similar to the above warning; this time the warning happens if you attempt to close Tablet while a file is still loaded.
- **When attempting to load an assembly if no reference file was included** Not all assembly file formats include reference/consensus information in the primary file and it must be provided in a separate file. Tablet does not need this information, but can warn you if it is not included.
- **When switching between features being treated as padded or unpadded** GFF3 feature files include position information that Tablet can treat as either a padded or unpadded position. Depending on the data,

the difference between a padded or unpadded position can be significant, and you may wish Tablet to remind you of this.

- **When a search has stopped after finding 500 results** Continuing to search after 500 matching hits have been found has a negative impact on memory consumption. Disable this option to stop Tablet from reminding you of this fact each time it searches.
- **If sequence lengths specified in SAM/BAM files don't match those imported** Assemblies in SAM or BAM formats - although they do not store reference data - do store the name and length of the each reference sequence. Check this option to have Tablet warn whenever these specified lengths do not match the actual lengths of any reference sequences that you import.

CHAPTER 17

Command Line Options

Tablet supports command line options for automatically loading and navigating to locations in an assembly.

17.1 Loading assembly files

Provide your assembly file - and reference file if one is required - as an argument after tablet.exe on the command line. Tablet automatically works out which is the assembly file and which is the reference file, so you can either run:

```
tablet.exe assembly_file reference_file
```

or

```
tablet.exe reference_file assembly_file
```

17.2 Automatically open contig

Tablet can automatically open a contig in a provided assembly from the command line. To do this simply specify the name of a contig from your assembly file along with the assembly file when running Tablet from the command line prefixed with *view*:

```
tablet.exe assembly_file reference_file view:contig_name
```

It is also possible to move to a location in the given contig when loading from the command line by specifying the base position after a colon, e.g.

```
tablet.exe assembly_file reference_file view:contig_name:200
```

CHAPTER 18

Tablet XML

Tablet supports Tablet XML files, or .tablet files, for specifying sets of data which can be loaded together. If you have Tablet installed you can double click a .tablet file to launch Tablet with the data specified in that file. Alternatively you can drag and drop a .tablet file onto an already open instance of Tablet to load the associated data, or load data using a .tablet file as the argument to Tablet on the command line.

Note: There is no way to open a Tablet XML (.tablet) file from within Tablet, other than by dragging and dropping the file onto Tablet.

18.1 Tablet XML Elements

Tablet XML files allow you to specify the following elements, an assembly file, a reference file, an annotation file, a contig and a position. The only required element is an assembly file. The position element only works when paired with a contig element. Below are examples of how to specify each element:

- `Assembly file` (required)

  This can be in any of the assembly file formats that Tablet *supports*.

```
<assembly>https://bioinf.hutton.ac.uk/tablet/sample-data/book/example4/
↪Example4.bam</assembly>
```

- `Reference file` (optional)

  A fasta formatted reference file.

```
<reference>https://bioinf.hutton.ac.uk/tablet/sample-data/book/example4/
↪Example4.fasta</reference>
```

- `Annotation file` (optional)

  A gff3 formatted annotation file.

```
<annotation>https://bioinf.hutton.ac.uk/tablet/sample-data/book/example4/
↪Example4.gff</annotation>
```

- `Contig` (optional)

  This should be the name of a contig in the assembly file.

```
<contig>contig_53395</contig>
```

- `Position` (optional)

  A position within the specified contig.

```
<position>14000</position>
```

18.2 Example

```
<tablet>
        <assembly>https://bioinf.hutton.ac.uk/tablet/sample-data/book/example4/
↪Example4.bam</assembly>
        <reference>https://bioinf.hutton.ac.uk/tablet/sample-data/book/example4/
↪Example4.fasta</reference>
        <annotation>https://bioinf.hutton.ac.uk/tablet/sample-data/book/example4/
↪Example4.gff</annotation>
        <contig>contig_53395</contig>
        <position>14000</position>
</tablet>
```

This example file can be downloaded by clicking here.

CHAPTER 19

Allocating Memory

Tablet is written in Java, and due to way the Java Runtime works the amount of memory available for use by it must be defined before the application is started. The default value set at install time is one gigabyte (1024MB). If you need to allocate more (or less) memory than this, the setting can be adjusted by following the relevant instructions below.

Note though, that for any 32-bit system, the maximum amount of memory Tablet will be able to use will be somewhere between 1.5GB and 2GB, regardless of the total amount of memory installed. If you have data sets requiring more memory than this then you must run a 64-bit copy of Tablet on a 64-bit operating system.

19.1 Windows & Linux

Navigate to the directory in which Tablet is installed and locate the file **tablet.vmoptions** and open it with a text editor. You will see a line containing **-Xmx1024m** or **-Xmx4096m** - replace '1024' or '4096' with a memory allocation value (in MB) of your choice.

19.2 macOS

Navigate to Tablet's application icon (usually located in /Applications) and CTRL/right-click the icon, selecting **Show Package Contents** from the popup menu. Open /Contents/vmoptions.txt and replace the **4096** part of the line containing **-Xmx4096m** with a value (in MB) of your choice. The default is 4GB of memory.

CHAPTER 20

Tips & Shortcuts

Hints and tips on how to use Tablet are shown periodically in its status bar running along the bottom edge of its main window. For easy reference, the following is a list of all the tips that may appear:

- Many of Tablet's menu options are also accessible by right-clicking on the display canvas
- Navigate around a view quickly by clicking and dragging on one of the Overview displays
- Tablet will periodically check for new versions at startup
- The red rectangle on an Overview display shows the region of the data set currently being viewed on the main canvas
- Mouse over any read or base position to view further information on it
- You can move the canvas's viewpoint around by simply clicking on it and dragging with the mouse
- Right click on a visible read, consensus or protein translation to see options for copying its data to the clipboard
- Position data is often supplemented with U (unpadded position) and CV (read coverage at that position) values providing the unpadded value and CV providing a value for coverage
- Navigate around an alignment by clicking and dragging on either the overview display area or the main display area
- GFF3 data can be displayed on the features track after importing
- Right click on the features track to access the option to select which tracks are visible
- Use the Search tab to search for subsequences within the reads, or within the consensus / reference
- Mousing over CIGAR "I" features on the features track highlights the reads - and locations - the insertion relates to
- The visible reads tab contains a table of information on the reads currently on screen
- CTRL / CMD drag the mouse on the overview display to subset the overview
- Right click on a read to access the jump to navigation options for jumping to a read's start, end or pair
- Right click on the line connecting paired reads to access jump to options for jumping to the left and right read of the pair

- When the read shadower is in custom mode, right click on any base position to lock the highlighting
- Mouse over a feature on the feature track to see information including its name and start and end positions