Version

1

NEOMORPHIC Software, Inc.

Integrated Analysis and Annotation

Annotation Station User’s Guide

Integrated Analysis and Annotation

Annotation Station User’s Guide

© Neomorphic Software, Inc.

2612 8th Street • Berkeley, CA 94710

Phone 510.704.1030 • Fax 510.704.1013

Table of Contents

|  |  |  |
| --- | --- | --- |
| Foreword | | v |
| Preface | | vi |
| Acknowledgements | | viii |
| CHAPTER 1: ANNOTATION STATION COMPONENTS | | 1 |
| THE MAIN WINDOW | | 2 |
|  | The Login Dialogue Panel | 3 |
|  | The Genome Viewer and Chromosome Viewer Panels | 4 |
|  | The Curation-Analysis Tree and Annotated Bio-Sequence Panels | 6 |
|  | The File Menu | 8 |
|  | The Edit Menu | 10 |
|  | The View Menu | 11 |
|  | The Help Menu | 12 |
| THE GENE EDITOR WINDOW | | 13 |
|  | The Annotation Editing Panel | 15 |
|  | The mRNA Translation Panel | 19 |
|  | The File Menu | 21 |
|  | The Edit Menu | 22 |
|  | The Annotation Menu | 24 |
|  | The View Menu | 26 |
| CHAPTER 2: ANNOTATION STATION COOKBOOK | | 28 |
| GETTING STARTED | | 30 |
|  | How do I log in to the *Annotation Station* and open a project? | 30 |
|  | How do I log in as a new user or open a new project? | 30 |
|  | How do I change the Login Dialogue Panel’s “user name” and “project” defaults? | 30 |
|  | How do I switch between file mode and http mode? | 31 |
|  | How do I alter the *Annotation Station* color scheme and set other user preferences? | 31 |
| VISUALIZATION – VIEWING YOUR DATA | | 33 |
|  | How do I visually drill down into the genome? | 33 |
|  | How do I view features on both the forward and reverse DNA strands? | 33 |
|  | How do I view the linear relationship between curations and evidence? | 34 |
|  | How do I determine if evidence agrees or disagrees with a curation? | 34 |
|  | How do I rapidly view the ID and genomic coordinates of evidence and curations? | 34 |
|  | How do I view the start and end coordinates for all of the exons in a gene model? | 34 |
|  | How do I view aligned splice-sites for all of the exons in a gene model? | 35 |
|  | How do I view the mRNA sequence for a working gene model? | 35 |
|  | How do I view start and stop codons in all six frames of a gene model? | 35 |
|  | How do I view the protein sequence for a gene model translated in three reading frames? | 36 |
|  | How do I output cDNA and protein sequences to the clipboard? | 36 |
| CREATING AND MODIFYING CURATIONS | | 37 |
|  | How do I open an existing curation for editing? | 37 |
|  | How do I open an existing curation into a “read only” Gene Editor Window? | 37 |
|  | How do I create a working model from a committed curation or gene prediction from within the Gene Editor Window? | 37 |
|  | How do I add exons and remove exons from a curation? | 38 |
|  | How do I add a completed curation to the database? | 38 |

Foreword

W

e stand today at the crossroads of a scientific and technological revolution. The emergence of Molecular Biology – with its fundamental discovery that DNA is the fabric of life – is causing us to radically rethink our approach to science and to life in general. Mammalian cloning is now possible. Once untreatable diseases such as cancer, Parkinson’s, and bipolar disorder, are rapidly being brought under our control. The Human Genome Project – a multinational initiative to sequence the entire human genome – plans to release an initial draft of complete human genome by May 2000. As additional resources are being poured into the sequencing of numerous other genomes – with the National Center for Biotechnology Information (NCBI) database currently reporting on over 47,000 different organisms – the sheer quantity of data becoming available to us is truly mind-boggling.

Never before have there been so many exciting scientific possibilities to explore. In order to choose from amongst these various options, the mass of genomic data must be analyzed. Fortunately, a concurrent revolution is occurring in the computer industry. More computational power is now available to the average desktop user than was available on the most advanced supercomputers 15 years ago. The recent explosion of the Internet and database technologies has allowed all of this biological data to be placed online in a form available to the general public. Bioinformatics, the powerful new frontier emerging from the combination of these scientific and computational approaches, is now enabling us to transform this raw biological data into useful scientific knowledge.

Despite this great explosion of information and computational resources, there has historically been a lack of technological tools available to organize and analyze this data. This is exactly the problem that was being faced by The Institute for Genomic Research (TIGR) one year ago. With the full sequence of the *Arabidopsis thaliana* genome scheduled to come out in several months, Steven Salzberg, the Director of Bioinformatics at TIGR, approached Neomorphic looking for a tool that would help TIGR’s scientific team find genes and other meaningful features within the *Arabidopsis* data. TIGR was interested in Neomorphic’s well-established visualization technologies, and Neomorphic was looking to partner with one of the top biological research groups in the country in order to create an annotation tool that would address the most relevant biological issues. And so the *Annotation Station* was born.

It is with great pride that we present to you Version 1.0 of the *Annotation Station*; a product that combines unparalleled scientific expertise with the power of the Internet revolution. We are confident that the *Annotation Station* will serve your Bioinformatics needs, no matter how big or how small. We look forward to meeting you down the road to discovery as you convert your biological data into scientific knowledge!

Best wishes,

Cyrus Harmon, President and CEO, Neomorphic Software, Inc.

Preface

This book is designed for use with the *Annotation Station: Integrated Analysis and Annotation* software package. This preface describes what you can expect from this user manual, highlights some of the conventions that are used in this book, and provides pertinent customer support information.

What’s in this Book

This User’s Guide is divided into two chapters. Chapter 1, “Annotation Station Components”, describes the feature set of the *Annotation Station*. The chapter is designed to help you become more familiar with the various windows and panels in the *Annotation Station* and with some of the terminology that is used to describe the annotation process. Screen shots are provided of the various parts of the *Annotation Station* and their corresponding functions are described in detail. Chapter 2, “Annotation Station Cookbook”, presents a series of use cases on how to complete various everyday annotation tasks using the *Annotation Station*. The chapter presents step by step instructions with the hope that it will facilitate your task of using the *Annotation Station* to complete your work.

Conventions Used in this Book

Several different symbols, icons, and type sets are used throughout this book to separate out different types of information for ease of reference and for the sake of improved clarity. They are:

🕭 Take “note” – This icon signals a useful side-point of which you might wish to take note. These points will educate about additional features of the *Annotation Station*.

Color key – The *Annotation Station* encodes visual information through the use of both different shapes and different colors. Since understanding the meanings of the different colors in the *Annotation Station* is so important, this icon signals information that describes the meanings of the various colors that you will encounter.



Menu item key – Indicates the definition of a menu item. This symbol is also in Chapter 2 any time you are asked to select a given menu item.



Action key – Used in Chapter 2, this icon indicates that you must perform a specific action.



Input key – Used in Chapter 2, this icon indicates that you must type something or input a command into the computer.



Bold courier font – Used for information that regards files or requires computer interaction outside of the *Annotation Station* program.

Who We Are

Neomorphic is a computational biology company located in Berkeley, California. Neomorphic’s mission is to develop cutting-edge tools that create meaning out of biological data. Neomorphic’s proprietary analytical techniques and its interactive graphical user interfaces enable researchers to identify novel genes and to determine their role in disease and other biological processes. We hope that the *Annotation Station* will enable you to assign meaning to *your* biological data.

We have tested and verified the components of the *Annotation Station* and the information contained herein to the best of our ability, but certain errors may have slipped past our notice. Please let us know if anything is amiss, or if you have any suggestions for future versions, by contacting us at:

Annotation Station Team

Neomorphic Software, Inc.

2612 8th Street • Berkeley, CA 94710

Phone 510.704.1030 • Fax 510.704.1013

For questions relating to the use of the *Annotation Station*, please feel free to contact me directly at: ‘burns@neomorphic.com’. I look forward to hearing from you and learning more about the work that you are doing!

Best wishes,

Erik Burns, Bioinformatics Project Manager, Neomorphic Software, Inc.

Acknowledgements

There are a number of people without whom the *Annotation Station* and corresponding *Annotation Station User’s Guide* would not have been possible.

First and foremost, our thanks go to all of those individuals at TIGR who provided invaluable feedback on which features to include and (thankfully!) which features to leave out. Steven Salzberg, the Director of Bioinformatics at TIGR, and Chris Town were very generous in structuring a collaborative relationship that would allow us to provide TIGR with a product that would meet its needs while still leaving us with something to show for our efforts. Xiaoying Lin, Staff Scientist at TIGR, provided great biological insight and guidance about the most relevant feature set for the *Annotation Station*. Hanif Kalak and Maria Ermolaeva gave helpful feedback about underlying design and database issues and Todd Creasy served as a tremendous mediator to make sure that TIGR’s needs were receiving a voice. My personal thanks go to Todd for being such a positive voice throughout the project; his enthusiasm helped keep us all feeling confident that we were moving in the right direction.

As much as we (the management) would like to think so, software projects aren’t so much the result of weekly meetings, schedule revisions, and issue tracking, but rather are pulled from the blood, sweat, and tears of their engineers. I don’t think we could have asked for a more dedicated and knowledgeable engineering team. Thanks to Shaw Sun for being a tremendous technical lead on the project and for keeping everyone’s priorities straight. Shaw came in numerous weekends to work on the project, and was a source of inspiration to us all. Steve Chervitz provided some very valuable input on ensuring that the I/O of the program would actually work, and also provided some solid documentation and organizational assistance early on in the project. Eric Blossom, a Senior Engineer here at Neomorphic, was a steadfast designer and builder and kept us on our toes. Cyrus Harmon and Moses Cesario helped keep us on track by providing necessary project vision. Thanks as well to Ann Loraine and Daniel Green for coming in to the project at a late stage and providing much-needed support.

It is impossible to mention all of the contributors on any project. To those people who helped out in the countless ways not mentioned here, we are equally indebted. Thank you!

Chapter

1

Annotation Station Components

This chapter provides a detailed overview of the primary windows in the Annotation Station and their corresponding menu items.

T

he *Annotation Station* consists of two main components: a Main Window and a Gene Editor Window. Each of these windows contains a variety of menu options that provide access to the various features of the *Annotation Station*. These windows are also divided into a number of sub-panels or “viewers” that differ substantially in their uses and functions, as well as in the types of input that they require from you. In order to become fluent in using the *Annotation Station*, it will be necessary for you to gain an understanding of the information provided by these panels and to learn how to use the corresponding menu items.

Chapter Contents:

|  |  |  |
| --- | --- | --- |
| THE MAIN WINDOW | | 2 |
|  | The Login Dialogue Panel | 3 |
|  | The Genome Viewer and Chromosome Viewer Panels | 4 |
|  | The Curation-Analysis Tree and Annotated Bio-Sequence Panels | 6 |
|  | The File Menu | 8 |
|  | The Edit Menu | 10 |
|  | The View Menu | 11 |
|  | The Help Menu | 12 |
| THE GENE EDITOR WINDOW | | 13 |
|  | The Annotation Editing Panel | 15 |
|  | The mRNA Translation Panel | 19 |
|  | The File Menu | 21 |
|  | The Edit Menu | 22 |
|  | The Annotation Menu | 24 |
|  | The View Menu | 26 |

# The Main Window

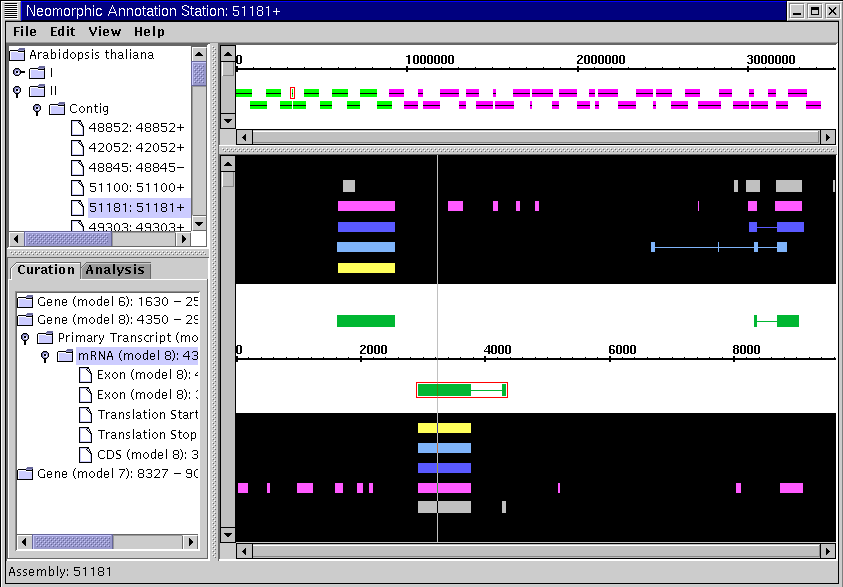
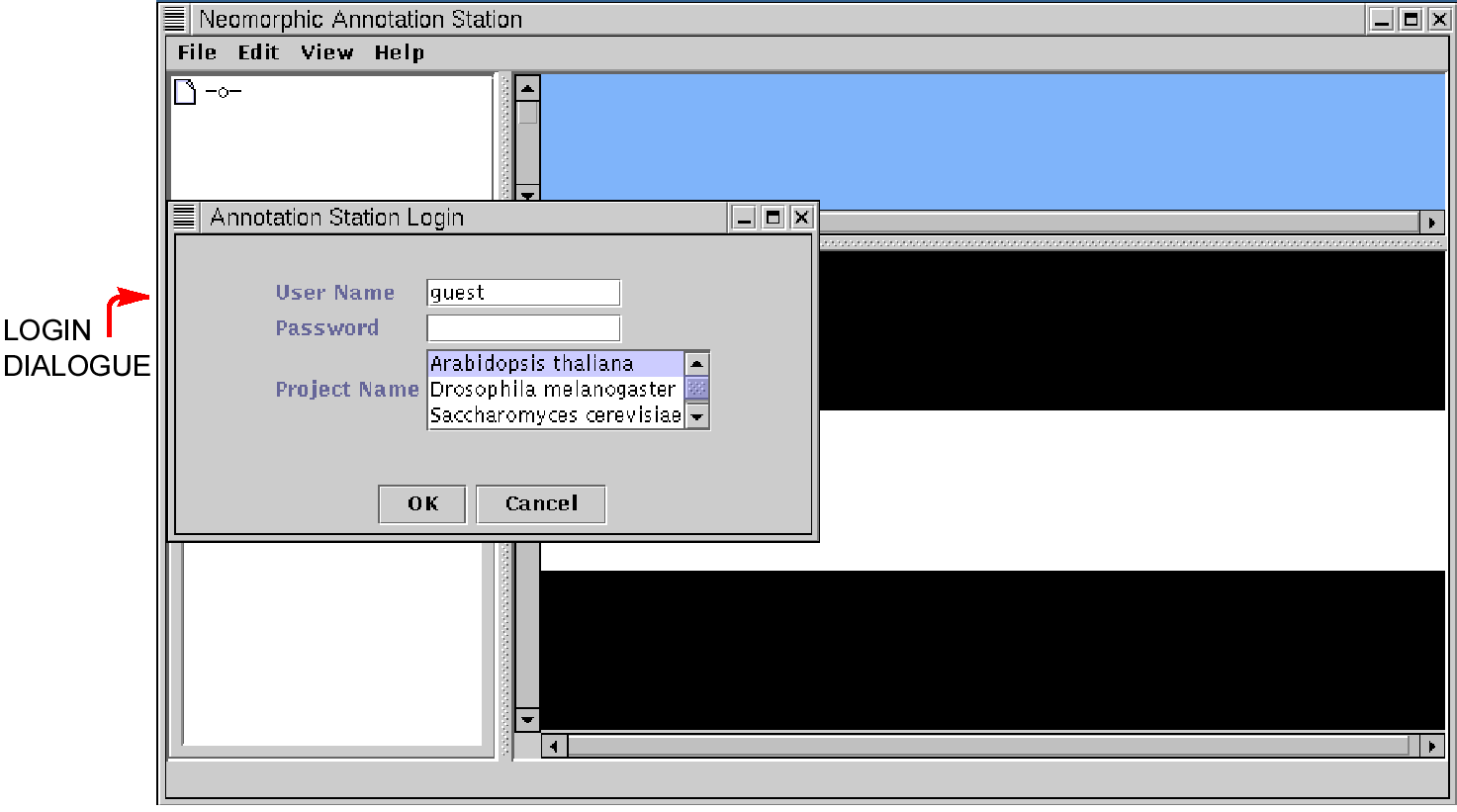


Figure 1.1 – The Main Window.

The Main Window is where the majority of the coarse-grained annotation work gets done. The Main Window handles user login, logout and project selection; enables you to “drill down” into a selected genome from the broad chromosomal level down to the DNA sequence; and displays the linear relationship between different types of analytical evidence and any associated curations in the database. The Main Window also allows you to choose a genomic feature of particular interest (e.g., a gene) to import into the Gene Editor Window and subsequently modify.

The Main Window consists of five different panels: the Login Dialogue, the Genome Viewer, the Chromosome Viewer, the Annotated Bio-Sequence, and the Curation-Analysis Tree. The Main Window also contains four different menus: File, Edit, View, and Help. The remainder of this section will be dedicated to describing the uses of these panels and menus.

The Login Dialogue Panel



**Fig. 1.2 – The Login Dialogue Panel.**

The Login Dialogue Panel is automatically launched when you run the *Annotation Station*. The Login Dialogue contains three input boxes that allow you to enter your user name, password, and project name. Your choice of user name determines what types of privileges you are granted when using the *Annotation Station*. For instance, if you log in as a “guest”, you will be able to view project data but will not be able to commit any changes to the database. Additionally, your choice of project will affect the type of data that you are able to view with the *Annotation Station*.

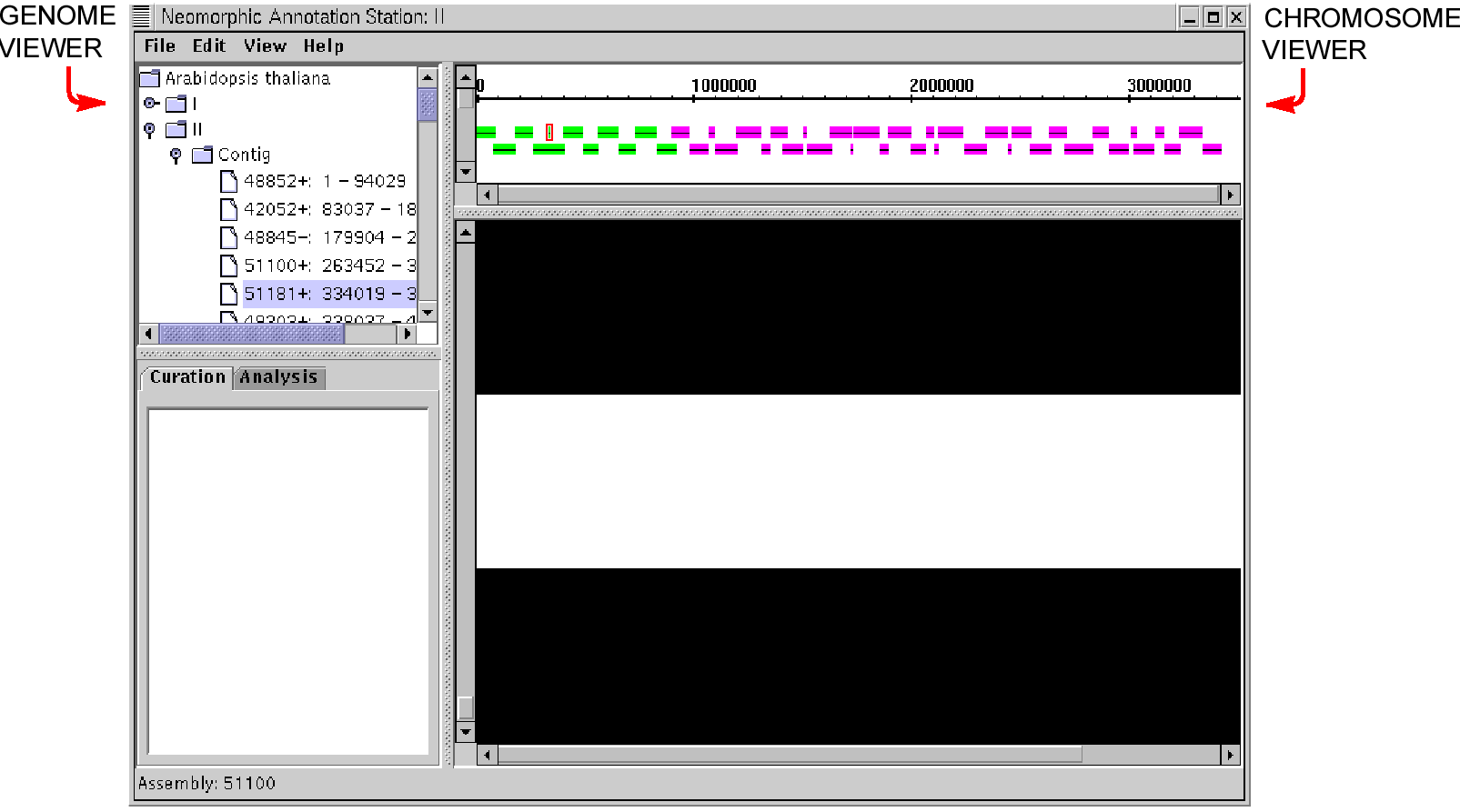
🕭

## Altering Login Dialogue Defaults

The default values for the user name and project fields are “guest” and “Arabidopsis thaliana”, respectively. If you wish to alter these defaults, you will need to alter the **stayshun.properties** file in the **preferences/** sub-directory of the *Annotation Station* main directory. Since this file is a text file, you will only need access to a basic text editor to edit the file. Altering the **user-name** and **project-name** fields will alter the corresponding values in the Login Dialogue Panel.

Once you have entered the required information into the Login Dialogue Panel and selected “OK”, the selected project will appear in the Genome Viewer Panel (upper left-hand panel). A list of other users logged-in to the same project is displayed in the Main Window status bar. Since one of the safety features of the *Annotation Station* is that it does not allow two users to concurrently commit edits to a gene model, it will be useful to keep track of the other annotators logged in to your project.

The Genome Viewer and Chromosome Viewer Panels



**Figure 1.3** **– The Genome Viewer and Chromosome Viewer Panels.**

Once you have logged in, the project you selected will become visible in The Genome Viewer Panel. For instance, if you selected the *Arabidopsis thaliana* project, the Genome Viewer will display the five *Arabidopsis* chromosomes and a category “unknown” which stores unassigned BACs. The Genome Viewer Panel displays information in a “tree view” which means that the information is displayed hierarchically. To drill down into this information, you merely need to expand one of the “nodes” of the tree. To do this, click on one of the metal “key” icons next to a given chromosome. This will show a list of all of the contigs and single BACs that have been assigned to that chromosome. You may further drill down into a contig to shown the constituent BACs for that contig.

Once you have highlighted a given piece of information in the Genome Viewer Panel – for example, one of the chromosome folder icons or a given contig – you will notice that information about your selection is displayed in the Chromosome Viewer Panel (upper right-hand panel). These two panels are tightly linked together and display different aspects of the same information; the Genome Viewer Panel is textual while the Chromosome Viewer Panel is graphical. You may select several different levels of information in the Genome Viewer Panel – a genome, a chromosome, a contig, or a BAC – and the Chromosome Viewer Panel displays each of these differently. Following is a list of the information displayed for each of these different types of selections:

|  |  |
| --- | --- |
| Genome Viewer Selection | ***Chromosome Viewer Display*** |
| Genome………………………. | No information is displayed |
| Chromosome……………….… | BACs present in that chromosome are displayed in a long list. |
| Contig………………………… | The BACs present in that contig are displayed along with their relative position along an axis. |
| BAC…………………………. | If the BAC is part of a contig, the contig is still displayed in the Chromosome Viewer Panel and a red box is drawn around the corresponding BAC. If the BAC is not assigned to a contig, the BAC is displayed by itself. |

Color encoding plays a large role in the display of information in the Chromosome Viewer Panel. The background color of the panel changes depending on the level of information displayed, and BACs are color-coded to reflect their sequencing status. The meaning of these colors is as follows:

Chromosome Viewer Panel Background Colors:



**Light Blue** – the entire genome is selected in the Genome Viewer Panel.

**Light Brown** – a chromosome is selected in the Genome Viewer Panel.

**White** – A contig or a BAC is selected in the Genome Viewer Panel.

BAC Colors:



**Magenta** – BAC has been ­*fully* sequenced, *fully* annotated.

**Green** – BAC is being edited by another annotator and is *locked*.

**Gray** – Annotation status is *unknown*.

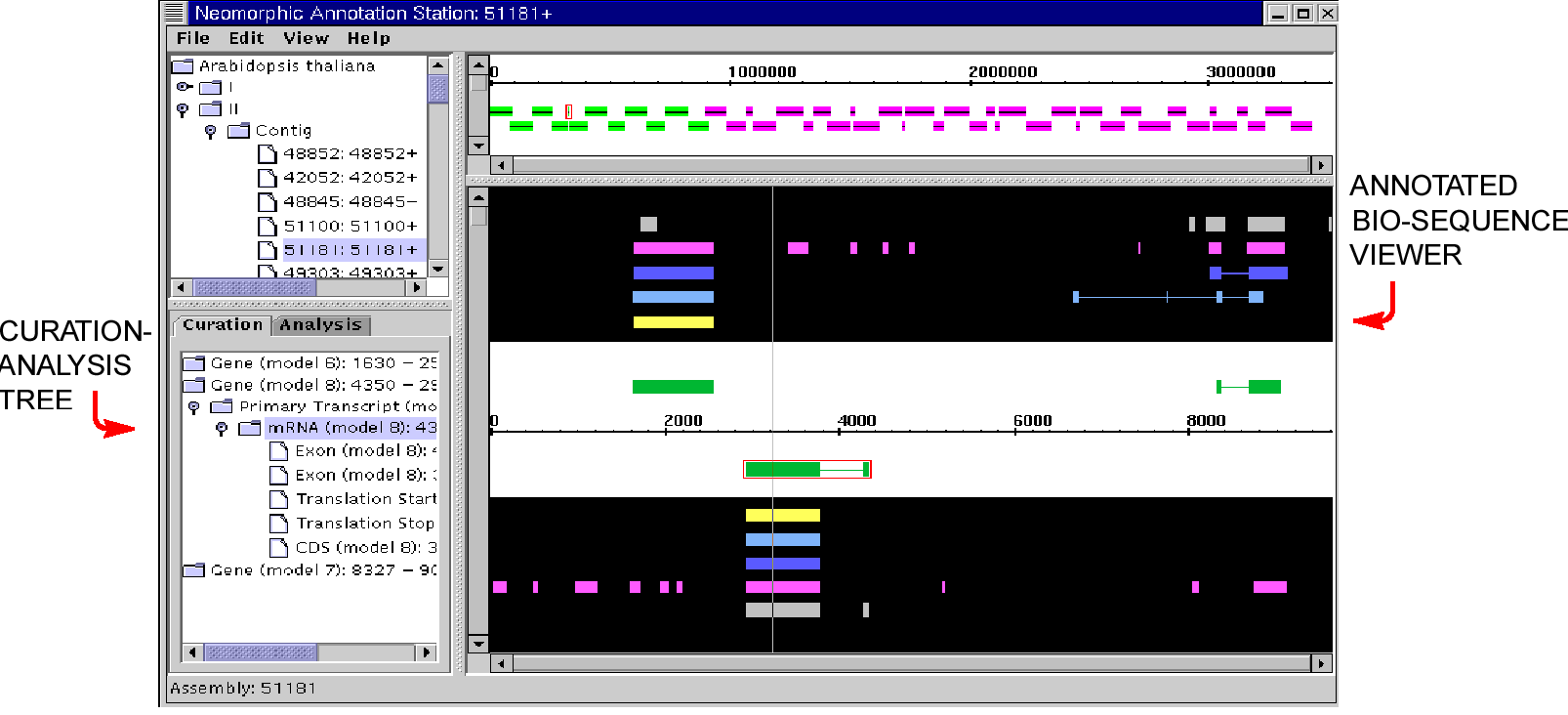
**Blue** – Gaps are being closed in the sequence.

**Red** – BAC sequencing is in progress.

**Yellow** – Annotation is in progress, but is not yet finished.

The main purpose of the Genome Viewer Panel and the Chromosome Viewer Panel is to select a BAC to view in greater detail. A BAC may be opened in one of two ways – either you may “double click” on a BAC in the Genome Viewer Panel or Chromosome Viewer Panel or you may open a selected BAC by choosing the “Open BACs” option in the “File” menu. Once you have opened a BAC, it will load into the Curation-Analysis Tree Panel and the Annotated Bio-Sequence Panel.

The Curation-Analysis Tree and Annotated Bio-Sequence Panels



**Figure 1.4 – The Curation-Analysis Tree and Annotated Bio-Sequence Panels.**

Once you open a BAC, it is displayed in full range in the Curation-Analysis Tree and Annotated Bio-Sequence Panels. Like the Genome Viewer and Chromosome Viewer Panels, these two panels display different aspects of the same information; the Curation-Analysis panel is textual while the Annotated Bio-Sequence Panel is graphical.

The Curation-Analysis Tree Panel consists of two selectable “tabs”, one entitled “Curation” and the other “Analysis”. These two tabs each contain hierarchical “tree”-based information displays enabling you to drill down into the information at various levels of depth. The first tab, “Curation”, contains a list of all of the genes that have been officially edited by a trained curator and added to the database. The other tab, “Analysis”, contains a list of all of the different types of evidence that are stored in the database. This list includes gene predictions from genefinder, genscan, and genscan+; exon predictions from grail; and database hits from a variety of different sources. Both of these tabs enable you to “drill down” into the data from the level of a gene/gene model to a primary transcript to an mRNA to the individual exons that make up the mRNA.

Every feature in the Curation-Analysis Tree Panel is displayed in a graphical form in the Annotated Bio-Sequence Panel. The Annotated Bio-Sequence Panel contains two distinct regions, an “evidence” region with a black background that displays the gene predictions and database hits, and a “curations” region with a white background that displays the committed curations. The upper half of the Annotated Bio-Sequence Panel displays genomic features on the forward strand, and the lower half displays features on the reverse strand. A thin gray “alignment hairline” provides a quick visual reference of the alignment of the evidence and curations with the axis.

Color encoding is used as a means to distinguish between the different types of evidence and curations displayed in the Annotated Bio-Sequence Panel. The meaning of these colors is as follows:

Evidence and Curations Colors (from top to bottom):



**Light Gray** – A conglomeration of DB hits from a variety of protein and EST databases.

**Pink** – Grail exon predictions

**Dark blue** – Genscan+

**Light blue** – Genscan prediction

**Yellow** – Genefinder prediction

**Green** – A curation; a feature that has been added to the database by a trained scientist

🕭

## Determining the Evidence Type

There are two quick ways to determine the identity of a selected piece of evidence. The evidence type is always displayed in the Main Window status bar. Additionally, you may zoom in to the Annotated Bio-Sequence Panel by moving the left-hand scroll bar downwards. Once you are zoomed far enough in, the evidence identity will appear in writing on the chosen piece of evidence.

Selection is two-way between the Curation-Analysis Tree Panel and the Annotated Bio-Sequence Panel. When you select an mRNA or one of its component exons in the Curation-Analysis panel, a red box is drawn around the corresponding item in the Annotated Bio-Sequence Panel. If you draw a red box around a feature in the Annotated Bio-Sequence Panel (by clicking on the feature), then the corresponding feature is highlighted in the Curation-Analysis Tree Panel.

The main purpose of the *Annotation Station* is to enable you to see aligned evidence in order to reveal new genes and other genomic features and to allow you to see ways in which existing curations should be edited and improved. However, the Annotated Bio-Sequence Panel does not supply editing capability directly. Rather, the main purpose of the Curation-Analysis Tab and Annotated Bio-Sequence Viewer Panels is to allow you to select a curation you wish to edit or a new area of a given BAC in which you wish to add a curation. Once you have selected a region to edit, you will open this region into the other primary window of the program called the Gene Editor Window, which is where all of the fine editing will occur. There are two ways to open a region: in “write enabled” mode or in “read only” mode. How you will open it depends on your level of user privileges, whether or not other curators are working on your project, and on the menu option you choose.

The File Menu

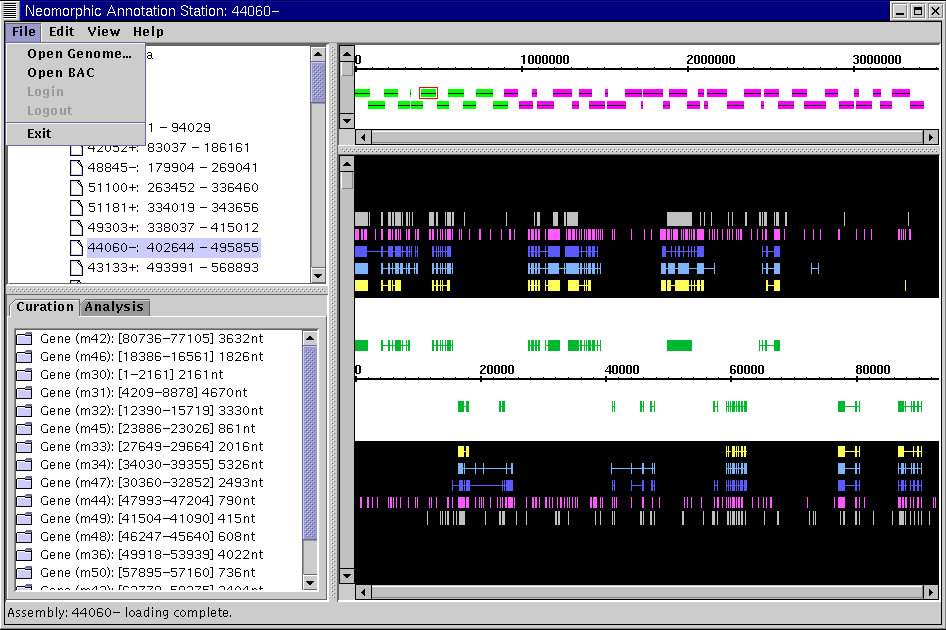


Figure 1.5 – The File Menu.

Open Genome… – This option allows you to open a local **.xml** file into the *Annotation Station*. Choosing this option pulls up a “directory navigator panel” which allows you to search through your local hard drive for a **.xml** file to open. While most of the **.xml** files are stored in the **xml/** directory, the **.xml** file you choose may be stored anywhere on your hard drive.



🕭

## The Open Genome … Command and “file mode”

Open Genome… is only available when running the *Annotation Station* in “file mode”. When running under “http mode”, projects are opened using the “Login” command. For more information on the difference between “file mode” and “http mode”, see “How do I switch between file mode and http mode?” in Chapter 2 of this User’s Guide.

Open BACs – The option allows you to open one or multiple BACs that are highlighted in the Genome Viewer or Chromosome Viewer Panels. Choosing this option causes the selected BACs to be displayed in the Curation-Analysis Tree and Annotated Bio-Sequence Panels. “Double-clicking” of the selected BACs will achieve the same result.



Login – Pulls up the Login Dialogue Panel that allows you to specify your user name and password and to select a project to open. This option is only available if you are not already logged in to a project.



Logout – Allows you to logout of the current project you are working on. Logging out of a project session closes the project you were working on, but leaves you in the *Annotation Station* Main Window for subsequent login to a new project.



Exit – Permanently exits you from the program. Closes all open windows (including the Main Window and all open Gene Editor Windows) and logs you out of the program.



The Edit Menu

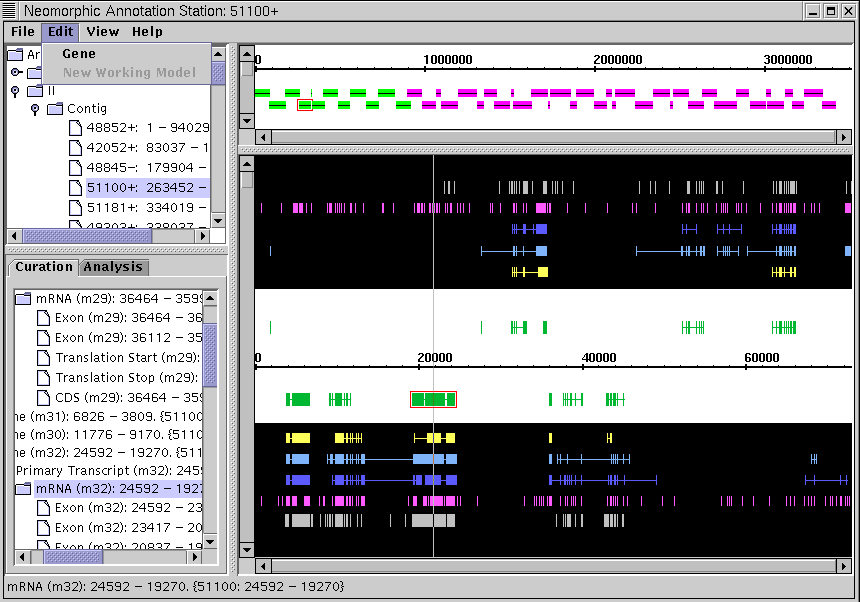


Figure 1.6 – The Edit Menu.

Gene – Allows you to import a curation into a “write enabled” Gene Editor Window for more detailed viewing and subsequent editing. To use this option, a gene must be highlighted in the Curation-Analysis Tree and Annotated Bio-Sequence Panels.



New Working Model – Allows you to import a piece of highlighted evidence into the Gene Editor Window for subsequent viewing and editing. The evidence selected may be a full gene prediction from genefinder, genscan, etc. or may be a single grail exon prediction. If there are no other Gene Editor Windows open, this option will open a “write enabled” Gene Editor Window; otherwise, a “read only” Gene Editor Window is opened.



The View Menu

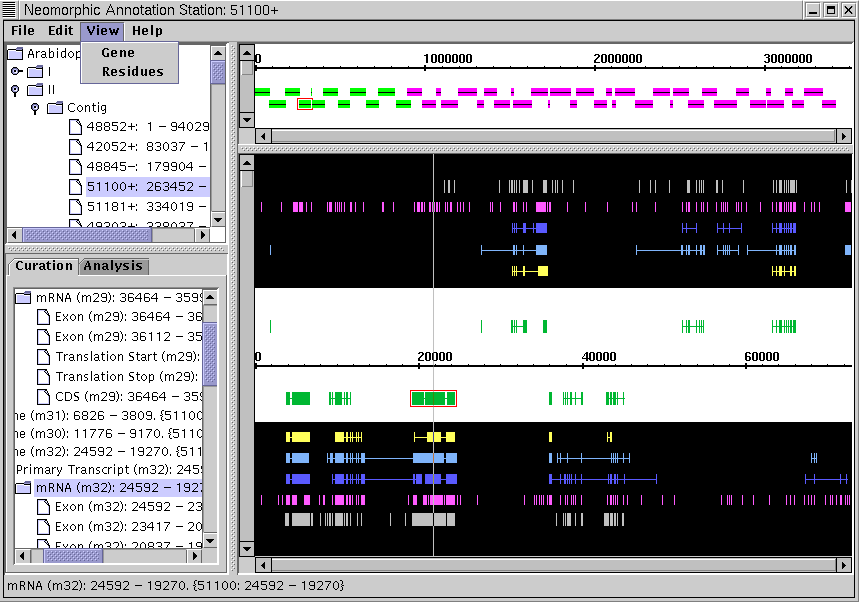


Figure 1.7 – The View Menu.

Gene – Allows you to import a highlighted curation into a “read only” Gene Editor Window. This option is available any time a curated feature is highlighted in the Curation-Analysis Tree and Annotated Bio-Sequence Panels. This menu option is especially useful if a “write enabled” Gene Editor Window is open, because this option provides the only way to open subsequent Gene Editor Windows.



Residues – Pulls up a text box that displays all of the nucleotides present within the current BAC. This option provides a convenient alternative to zooming all the way in to a BAC in the Annotated Bio-Sequence Panel in order to look at the constituent nucleotides. The residues in the text box are colored to reflect their alignment with the various types of evidence and curated features.



The Help Menu

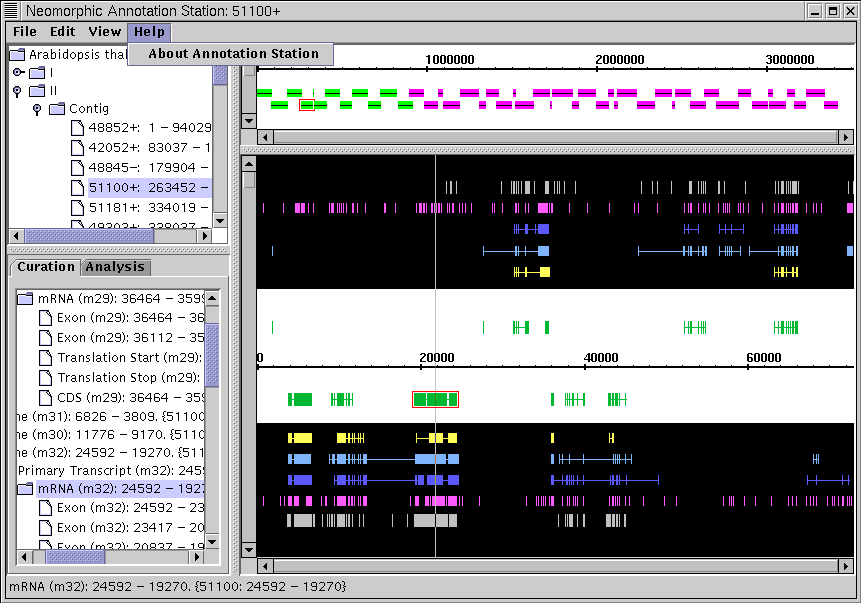


Figure 1.8 – The Help Menu.

About Annotation Station – Opens a text box that displays relevant information about the Neomorphic Annotation Station. The software version, build date, and web site address are displayed.



# The Gene Editor Window

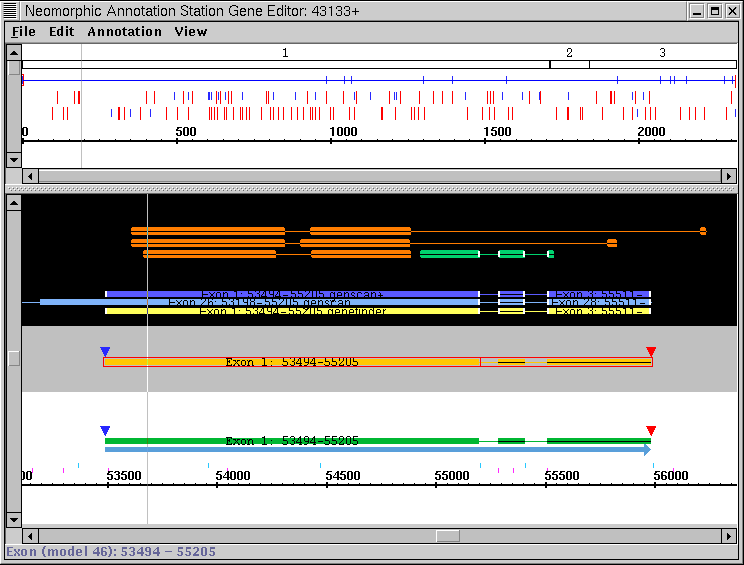


Figure 1.9 – The Gene Editor Window.

The Gene Editor Window is where the fine-grained annotation work gets done. The Gene Editor Window enables you to make accurate modifications to pre-existing curations and to create new gene models from scratch. In this window, you may view and modify the start and end coordinates of all of the exons in a gene model, you may view and modify splice sites and exon boundaries; and you may visualize the mRNA sequence, start and stop codons, and six-frame protein translations for a working model. You may output the cDNA and protein sequences from a gene model to the clipboard for incorporation into database searches or may save completed curations to the database. The Gene Editor Window incorporates all of the basic functionality required to do detailed genomic annotation.

🕭

## “Read only” Gene Editor Windows

You may open multiple Gene Editor Windows. However, only one “write enabled” Gene Editor may be open at any given time; all other Gene Editor Windows are launched as “read only”. This feature helps to prevent you from accidentally over-writing your work. The only functional difference between these two types of Gene Editors is that “read only” Gene Editor Windows may not be used to commit changes to database.

Each Gene Editor Window contains two different panels: an Annotation Editing Panel and an mRNA Translation Panel. Gene Editor Windows also contains four different menus: File, Edit, Annotation, and View. The remainder of this section will be dedicated to describing the uses of these panels and menus.

The Annotation Editing Panel

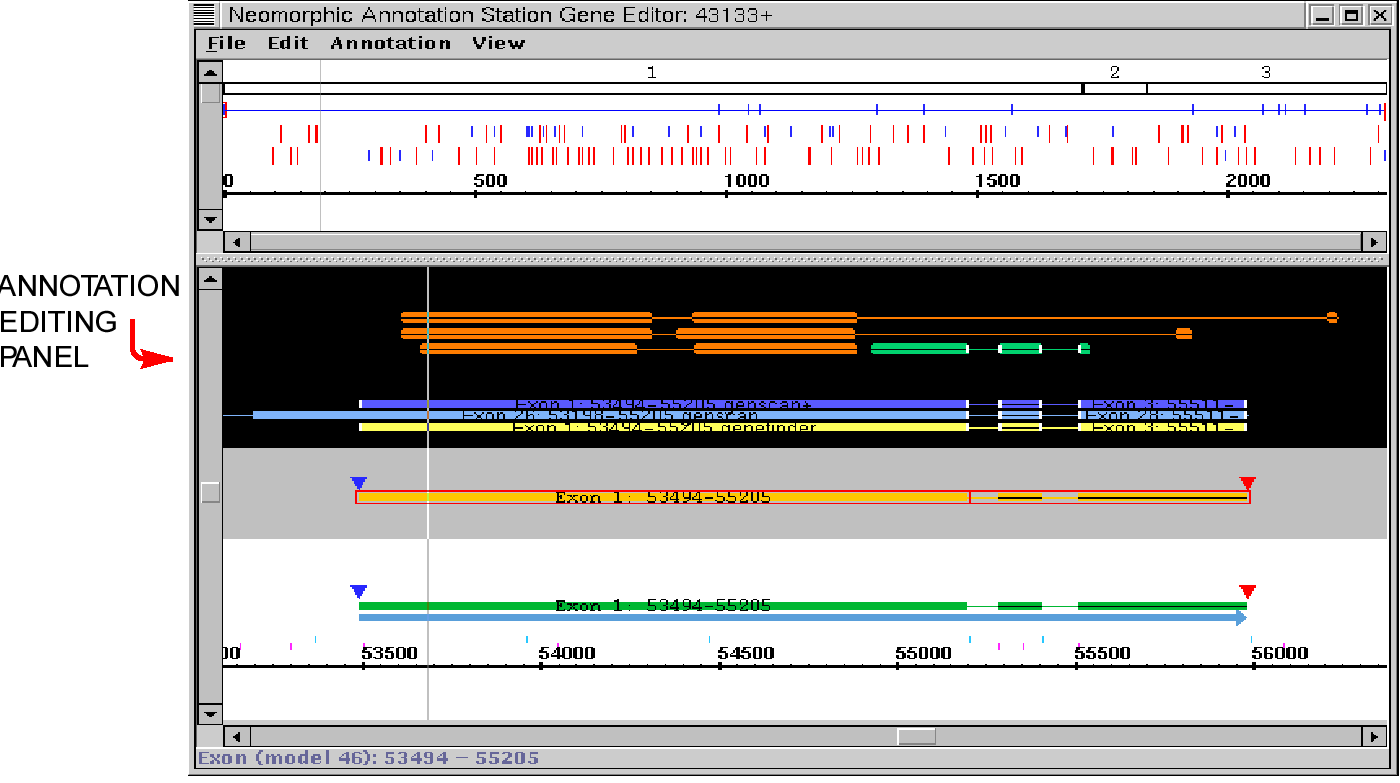


Figure 1.10 – The Annotation Editing Panel.

The Annotation Editing Panel is the main panel in the Gene Editor Window. The Annotation Editing Panel allows you to view and edit features along a primary transcript, add and remove exons from a gene model, view splice sites and adjust exon boundaries, and view the agreement between a gene model and its associated evidence. The Annotation Editing Panel also enables you to specify different translation start and stop sites via its interaction with the mRNA Translation Panel. The Annotation Editing Panel is probably where you will spend most of your time within the *Annotation Station*.

The Annotation Editing Panel consists of three distinct regions – a white “curations region” at the bottom of the panel, a gray or cyan “working model” region in the middle of the panel, and a black “evidence region” at the top of the panel. These regions are described in detail below:

The Curations Region

The curations region is the white region at the bottom of the Annotation Editing Panel. The curations region displays a more detailed view of the corresponding curations region in the Annotated Bio-Sequence Panel in the Main Window. Just like in the Annotated Bio-Sequence Panel, curations stored in the database are displayed along the genomic axis. However, there are a number of additional color-coded features displayed in the curations region of the Annotation Editing Panel. These features are described below:

Curations Region Colors (from top to bottom):



**Blue triangle** – Selected start codon for the curated gene model.

**Red triangle** – Selected stop codon for the curated gene model.

**Yellow-orange connected rectangles** – A curated gene that is currently being used as the basis for a working gene model in the “working model” region.

**Green connected rectangles** – Curated genes that are present in the database.

**Dark gray connected rectangles** - Curated genes that have been marked for deletion from the database.

**Blue arrows** – Primary transcripts for curated genes.

**Green pointed rectangles** – Curated tRNA genes present in the database.

**Magenta tick marks near axis** – Splice site acceptors; indicated by the presence of “AG’s” in the genomic sequence. The splice site acceptors are aligned with the nucleotide immediately *following* the “AG” in the sequence.

**Blue tick marks near axis** – Splice site donor; indicated by the presence of “GT’s” in the genomic sequence. The splice site donors are aligned with the nucleotide immediately *before* the “GT” in the sequence.

**Light gray pointed rectangles below axis** – Repeat regions in the genomic sequence.

By default, the curations region in the Annotation Editing Panel displays gene features along the selected strand only. However, you may also view features that lie on the opposite strand by clicking on the “View > Other Strand” check box.

The Working Model Region

The working model region is in the middle of the Annotation Editing Panel; it has a gray background in a “write enabled” Gene Editor Window and a cyan background in a “read only” Gene Editor Window. The working model region enables you to edit existing creations or to create new curations from scratch, and then commit these curations to the database. The working model region is, therefore, the most important part of the Gene Editor Window.

All of your interaction with the working model region will be through the use of “working models”. Working models are editable copies of pre-existing curations, or are new curations that are made from scratch or based on gene predictions. These “working models” are color-coded to reflect their status, and there is additional color-coded information displayed in the working model region of the Annotation Editing Panel. These features are described below:

Working Model Region Colors (from top to bottom):



**Blue triangle** – Selected start codon for the working gene model.

**Red triangle** – Selected stop codon for the working gene model.

**Yellow-orange connected rectangles** – The “current working model”; the model that is currently selected and is being edited.

**Green connected rectangles** – A working model other than the current working model.

**Dark orange connected rectangles** – Working models that have been multiply selected by “Ctrl-clicking” on them.

You may create an unlimited number of working models in the working model region, and there are a variety of ways to create them. You may replicate an entire curated gene or gene prediction, or you may choose only to replicate a single curated exon, exon prediction, or database hit. To learn how to perform these tasks, please refer to the command listings in *The Annotation Menu*.

Once a model has been created, there are a number of ways that it may be edited. You may add and remove exons from a working model or edit the boundaries of existing exons. You may join multiple working models or you may choose to delete a working model. There is an extensive list of commands that can be used to perform these functions. For a complete description, see the command listings in *The Edit Menu* and *The Annotation Menu*.

🕭

## Verifying your model with the Exon Table

A major part of editing working models is ensuring that your splice sites are correct. The *Annotation Station* is designed to facilitate this process. By selecting View > Exon Table, a table is displayed of all of the exon boundaries in the current working model. The two 3’ nucleotides of all of the introns in the current working model are color coded to reflect their agreement with the sequence “AT”, while the six 5’ nucleotides of all of the introns in the current working model are color coded to reflect their agreement with the sequence “GTAAGT”. Agreement is indicated in **black**, while disagreement is indicated in **red**.

Once you have created a working model and edited it to your satisfaction, the final step is to save your work. You may either save your work to a local file or commit your curations directly to the database. For a description of these option, please see the command listing in *The File Menu*.

The Evidence Region

The evidence region is the black region at the top of the Annotation Editing Panel. The evidence region displays a more detailed view of the corresponding evidence region in the Annotated Bio-Sequence Panel in the Main Window. The display of gene prediction is that same as it was in the Annotated Bio-Sequence Panel, although there is a more complete display of database hits, as described below:

Evidence Region Colors (from top to bottom of region):



**Orange** – Protein database hits using a TIGR internal non-redundant protein database.

**Green** – EST database hit.

**Aqua** – Protein database hit.

**Pink** – Grail exon predictions.

**Dark blue** – Genscan+.

**Light blue** – Genscan prediction.

**Yellow** – Genefinder prediction.

There are also two additional features present in the evidence region of the Annotation Editing Panel. To make the evidence panel less cluttered, the display of gene predictions may be turned off at any time by de-selecting the “View > Gene Predictions” option box. Additionally, the ends of evidence are color-coded to indicate their agreement with the current working model. If a piece evidence is drawn with a white end, it indicates that the evidence is aligned with the current working model. If a piece of evidence is drawn without a white end, it indicates that the evidence is *not* aligned with the current working model.

The mRNA Translation Panel

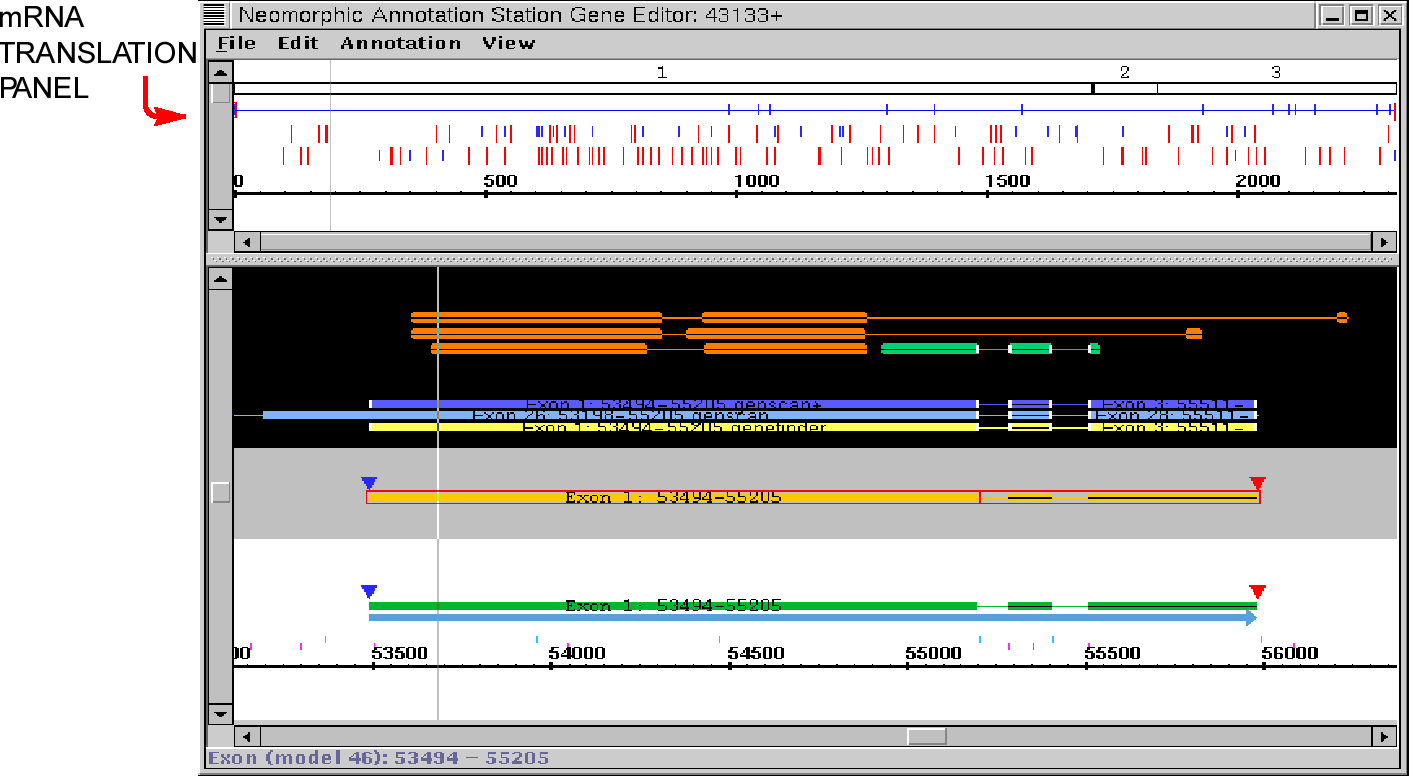


Figure 1.11 – The mRNA Translation Panel.

The mRNA Translation Panel is drawn at the top of the Gene Editor Window. You may use the mRNA Translation Panel to view the mRNA sequence for a working gene model, to specify the start and stop codons for a current working model, and to visualize the mRNA protein sequence translation in an open reading frame (ORF).

The mRNA Translation Panel displays the mRNA sequence for the currently selected working model in the Annotation Editing Panel. The mRNA sequence is just the genomic sequence of the exons in the working model with the introns spliced out (and with the “T’s” replaced by “U’s”). The numbers drawn at the top of the mRNA Translation Panel indicate the numbers of the corresponding exons in the working model, and the black vertical “dash marks” drawn below the exon numbers indicate the locations of the spliced-out introns. Start and stop codons are displayed according to the following color scheme:

mRNA Translation Panel Colors:



**Blue pointed rectangles** – Start codons that have the sequence AUG.

**Red rectangles** – Stop codons that have the sequence UAA, UGA, or UAG.

The main action required of you in the mRNA Translation Panel is to select a given start codon. To select a start codon, click on it in the mRNA Translation Panel. If you are having trouble doing this, you can zoom in to the mRNA Translation Panel to make the start codons bigger. The zoom bar is on the left-hand side of the mRNA Translation Panel; moving the bar down zooms you further in to the mRNA Translation Panel. Once you have selected a given start codon, you may use it to specify the start and stop codons in the current working model. For more information on this option, see the command listing in *The Annotation Menu*.

Selecting a start codon in the mRNA Translation Panel draws a blue line between that start codon and the next stop codon. This blue line indicates an ORF beginning at the selected start codon. To view the translated protein sequence for the selected ORF, you must zoom all the way in to the mRNA Translation Panel. Zooming all the way in will show you the actual mRNA sequence for the current working model. The translated protein sequence is displayed below the mRNA residues. You may then output this protein sequence and the corresponding cDNA sequence to the clipboard for incorporation into database searches. For more information on these options, please refer to the command listing in *The Edit Menu*.

The File Menu

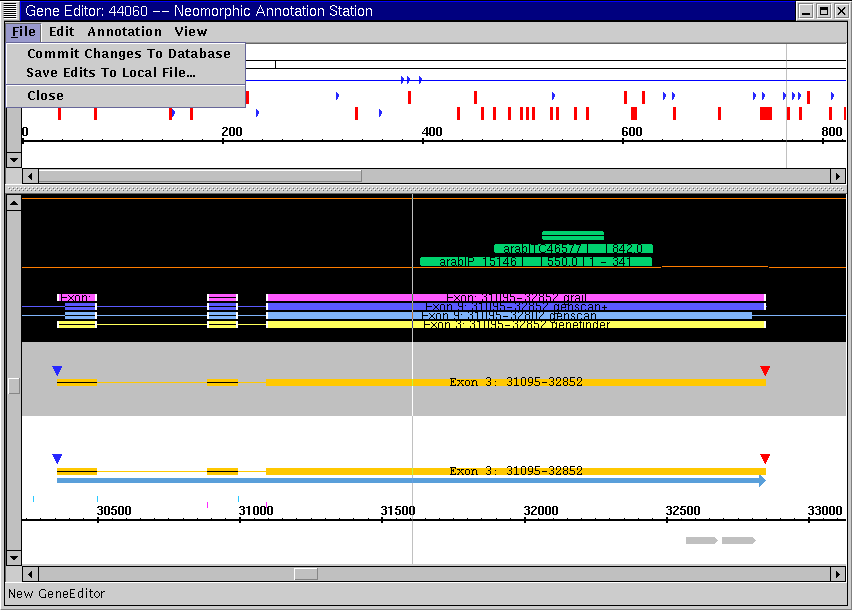


Figure 1.12 – The File Menu.

Commit Changes To Database – Allows you to save any changes that you made in the Gene Editor Window to the database. This option is only available in a “write enable” Gene Editor Window.



Save Edits to Local File…– This option allows you save your edits to a local file stored on your hard drive. This option is available in both “write enabled” *and* “read only” Gene Editors, and is the only way to save your work from a “read only” Gene Editor Window.



Close – This option closes the Gene Editor Window. If the Gene Editor is “write enabled”, you will be asked if you wish to commit your changes to the database before the Gene Editor Window closes.



The Edit Menu

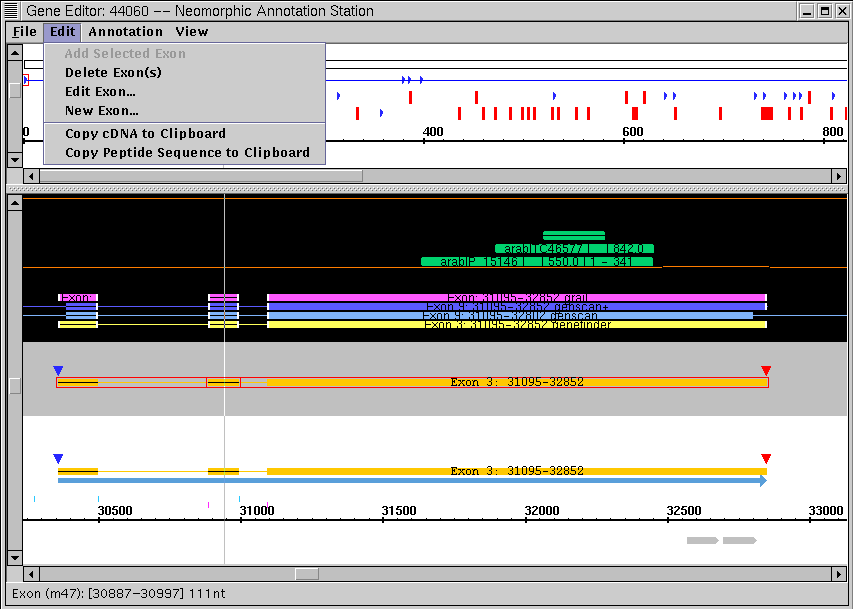


Figure 1.13 – The Edit Menu.

Add Selected Exon – Adds a selected exon to the current working model. To use this option, you must have an exon selected from either a committed curation, a gene prediction program, or a database hit.



Delete Exon(s) – Deletes an exon from the current working model. To use this option, you must have an exon selected in the current working model.



Edit Exon … – Allows you to edit a selected exon in a working model. A dialogue box is launched with the start and stop coordinates of the selected exon. You may edit these coordinates by hand and then select ‘OK’ to finish editing the exon. The exon will be updated to reflect your changes. In order to use this option, an exon in the current working model must be selected.



New Exon… – Allows you to add a new exon from scratch to the current working model. A dialogue box is launched with the arbitrary start coordinate “100” and stop coordinate “200”. You may edit these coordinates by hand then select ‘OK’ to add a new exon. This menu option is always available.



Copy cDNA to Clipboard – Turns the mRNA sequence from the current working model into a cDNA (by changing mRNA “U’s” to “T’s”) and then copies this sequence to the clipboard. You may then paste this sequence into database queries, etc.



Copy Peptide Sequence to Clipboard – Takes the “selected peptide sequence” for the current working model and pastes it to the clipboard. The “selected peptide sequence” is based on your choice of start codon in the mRNA Translation Panel. When you select a start codon in the mRNA Translation Panel, a blue line is drawn from that start codon to the next stop codon, thereby defining an ORF. This ORF is then translated into a peptide sequence, which is the protein sequence that gets copied to the clipboard.



The Annotation Menu

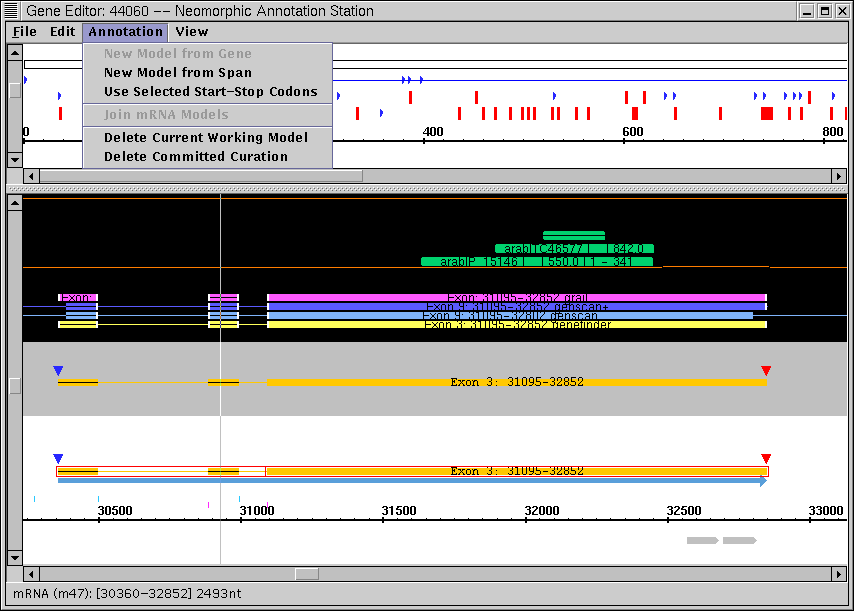


Figure 1.14 – The Annotation Menu.

New Model from Gene – Allows you to create a new working model from a curated gene or a gene prediction. The curated gene then appears as a current working model in the working model region of the Annotation Editing Panel. This menu option is only available when an entire curation or gene prediction is selected.



New Model from Span – Allows you to create a new working model consisting of only one exon from a gene prediction or database hit, or from part of a curated gene. The chosen exon then appears as a current working model in the working model region of the Annotation Editing Panel. This menu option is only available when a predicted exon, a database hit, or an exon from a curated gene is selected.



Use Selected Start-Stop Codons – Allows you to specify new start and stop codons for the current working model. To use this option, you must click on one of the blue start codons in the mRNA Translation Panel. That start codon and the next stop codon encountered in the mRNA Translation Panel can then be used as the “selected start-stop codons” for the current working model. This menu option is only available if you have selected start-stop codons in the mRNA Translation Panel *and* if these start-stop codons are not already being used as the start-stop codons for the current working model.



Join mRNA Models… – Enables you to merge two or more gene models into one gene model. You may select multiple working models by holding down the “Ctrl” button and then clicking on several different working models. Using this menu option then joins all of these working models together. This menu option is only available if multiple working models have been selected.



Delete Current Working Model – Removes the current working model from the working model region. The previous “current working model” then becomes the new “current working model”. This option is only available if there is a working model selected in the Annotation Editing Panel.



Delete Committed Curation/Restore Curation – This menu item changes depending on the item that you have selected. If you have selected a curation from the curations region in the Annotation Editing Panel, this menu item reads “Delete Committed Curation”. Using this menu item turns the curation from green to dark gray, thereby marking it for subsequent deletion from the database. If you later commit your changes to the database, this curation will be deleted.



If you have selected a curation that has been marked for deletion, then this menu option will read “Restore Curation”. Using this menu item will turn the curation from dark gray back to green and will unmark it for deletion from the database.

The View Menu

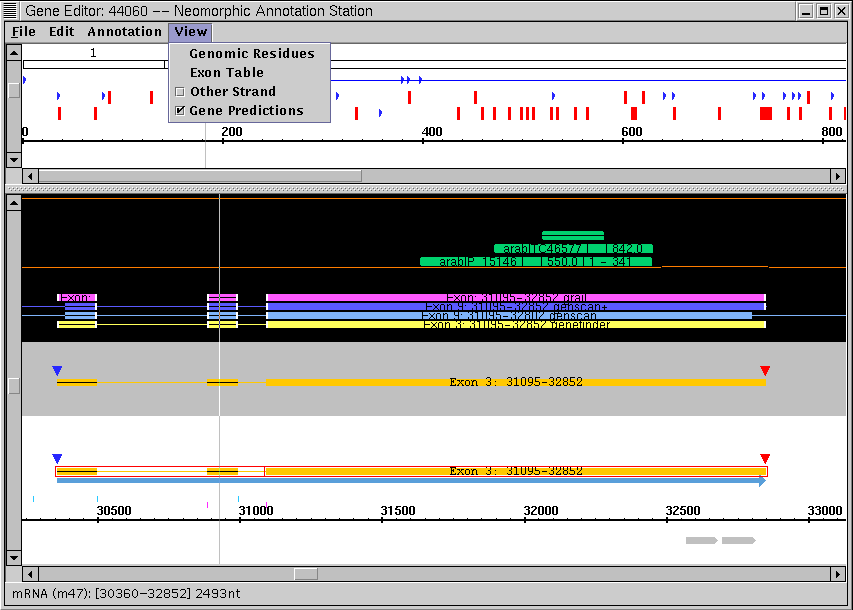


Figure 1.15 – The View Menu.

Genomic Residues – Pulls up a text box that displays all of the nucleotides present within the open BAC. This option provides a convenient alternative to zooming all the way in to the Annotation Editing Panel to display the genomic sequence.



Genomic Residues Text Box Colors:



**White** – The default color of the displayed genomic residues.

**Yellow** – The color of the genomic resides of the exons present within the current gene model.

Exon Alignments – Pulls up a text table that displays a detailed view of all of the exons within the current working model. This feature is designed to facilitate the process of editing exons. The two 3’ nucleotides of all of the introns in the current working model are color coded to reflect their agreement with the sequence “AT”, while the six 5’ nucleotides of all of the introns in the current working model are color coded to reflect their agreement with the sequence “GTAAGT”. Agreement is indicated in **black**, while disagreement is indicated in **red**. This option is only available if there is a working model present in the Annotation Editing Panel.



Other Strand – Displays the reverse strand in the Annotation Editing Panel and the mRNA Translation Panel. Checking this box displays the reverse complement; un-checking this box only allows display of the selected strand.



Gene Predictions – Enables you to toggle the display of gene predictions in the evidence region of the Annotation Editing Panel. Checking this box displays the gene predictions in the evidence region; un-checking this box removes the gene predictions from the evidence region display.



Chapter

2

Annotation Station Cookbook

The chapter will show you how to perform everyday annotation tasks using the Annotation Station.

S

ince the purpose of the *Annotation Station* is to help you – the curator – get your work done, the purpose of this chapter is to teach you how to use the *Annotation Station* to perform specific annotation-related tasks. For ease of reference, this chapter is divided into three functional sections. The first section, “Getting Started”, will show you how to log in to the *Annotation Station* and specify which project you wish to work on. You will learn how to determine who else is working on your project. You will also learn how to alter your preferences files to change visual and functional aspects of the *Annotation Station* user interface.

In the second section, “Visualization – Viewing Your Data”, you will learn how to drill down into your data from a high-level view of the genome down to the DNA sequence itself. You will learn how to visualize the alignment of different types of computational evidence and evaluate their agreement with curated genes and other features. You will learn how to view a curated gene up-close and how to view exon boundaries and splice sites. You will learn how to view the mRNA and protein sequences for a gene model in six frames, and how to copy these sequences to the clipboard for incorporation into database searches.

In the third section, “Creating and Modifying Curations”, you will learn how to open curations that are present in the database and how to create new curations. You will learn how to add and remove exons from a curation and edit exon boundaries. You will learn how to join and separate gene models, and how to commit curations to the database.

Chapter Contents:

|  |  |  |
| --- | --- | --- |
| GETTING STARTED | | 30 |
|  | How do I log in to the *Annotation Station* and open a project? | 30 |
|  | How do I log in as a new user or open a new project? | 30 |
|  | How do I change the Login Dialogue Panel’s “user name” and “project” defaults? | 30 |
|  | How do I switch between file mode and http mode? | 31 |
|  | How do I alter the *Annotation Station* color scheme and set other user preferences? | 31 |
| VISUALIZATION – VIEWING YOUR DATA | | 33 |
|  | How do I visually drill down into the genome? | 33 |
|  | How do I view features on both the forward and reverse DNA strands? | 33 |
|  | How do I view the linear relationship between curations and evidence? | 34 |
|  | How do I determine if evidence agrees or disagrees with a curation? | 34 |
|  | How do I rapidly view the ID and genomic coordinates of evidence and curations? | 34 |
|  | How do I view the start and end coordinates for all of the exons in a gene model? | 34 |
|  | How do I view aligned splice-sites for all of the exons in a gene model? | 35 |
|  | How do I view the mRNA sequence for a working gene model? | 35 |
|  | How do I view start and stop codons in all six frames of a gene model? | 35 |
|  | How do I view the protein sequence for a gene model translated in three reading frames? | 36 |
|  | How do I output cDNA and protein sequences to the clipboard? | 36 |
| CREATING AND MODIFYING CURATIONS | | 37 |
|  | How do I open an existing curation for editing? | 37 |
|  | How do I open an existing curation into a “read only” Gene Editor Window? | 37 |
|  | How do I create a working model from a committed curation or gene prediction from within the Gene Editor Window? | 37 |
|  | How do I add exons and remove exons from a curation? | 38 |
|  | How do I add a completed curation to the database? | 38 |

🖳

# Getting Started

* How do I log in to the *Annotation Station* and open a project?

Once you start up the *Annotation Station*, a query box – called the Login Dialogue Panel – is automatically launched. The Login Dialogue Panel contains input areas that allow you to enter your user name and password and to choose a project. To launch a project from the Login Dialogue Panel, follow these steps:

Enter your user name and password, and highlight your desired project



Click on ‘OK’



The Login Dialogue Panel may also be accessed at any time from the Main Window by using the following menu option:

File > Login



Note: you may only use the “File > Login” command if the *Annotation Station* is running and you are not already logged in to another project.

* How do I log in as a new user or open a new project?

The *Annotation Station* only allows you to be logged in to one project at a time. To log in as a new user or to open a new project, you will need to log out of your current project, and then log in again. From the Main Window, follow these steps:

File > Logout



File > Login



Using the File > Login command will bring up the Login Dialogue Panel. To use the Login Dialogue Panel to open a project, please see the help topic “How do I log in to the *Annotation Station* and open a project?”

* How do I change the Login Dialogue Panel’s “user name” and “project” defaults?

The default settings for the Login Dialogue Panel’s “user name” and “project” fields are “guest” and “Arabidopsis thaliana”, respectively. These defaults values are stored in a file called **stayshun.properties** in the preferences sub directory of the *Annotation Station* main directory. By editing this file, the corresponding default values may easily be changed. To edit this file:

Locate the **stayshun.properties** file in the preferences sub-directory of the *Annotation Station* main directory.



Open the **stayshun.properties** file with a text editor of your choice.



Change the **user-name=guest** field to read **user name=*new default name*** and change the **project-name=Arabidopsis thaliana** field to read **project-name=*new default project name***.



Save the modified **stayshun.properties** file.



* How do I switch between file mode and http mode?

The *Annotation Station* may be run in two modes: “http mode” which allows interaction with the database and “file mode” which allows you to run the *Annotation Station* off of files stored locally on your hard drive. One advantage of “file mode” is that it does not rely on database connectivity, thereby allowing the *Annotation Station* to run considerably faster. However, since it is usually desirable to be connected to the database when performing annotation work, the default is set to “http mode”.

To switch to “file mode”, you must alter the **stayshun.properties** file in the preferences sub directory of the *Annotation Station* main directory:

Locate the **stayshun.properties** file in the preferences sub-directory of the *Annotation Station* main directory.



Open the **stayshun.properties** file with a text editor of your choice.



Change the **mode=http** field to read **mode=file**.



Save the modified **stayshun.properties** file.



* How do I alter the *Annotation Station* color scheme and set other user preferences?

*Annotation Station* color properties are stored in the **stayshun.ass** file in the **preferences** sub-directory of the *Annotation Station* main directory. You can modify the **stayshun.ass** file to alter panel background colors, the appearance of assemblies and clones, and the appearance of gene features and evidence. To edit this file:

Locate the **stayshun.ass** file in the preferences sub-directory of the *Annotation Station* main directory.



Open the **stayshun.ass** file with a text editor of your choice. The beginning of the **stayshun.ass** file contains an extensive description of the various modifications one may make to achieve the desired results.



Modify the **stayshun.ass** file according to the instructions in the beginning of the file.



Save the modified **stayshun.ass** file.



While the **stayshun.ass** file defines visual aspects of the program, there is another file entitled **stayshun.properties** that defines more low-level session preferences. Together, these two files determine all of the user preferences settings. For more information on the **stayshun.properties** file, see “How do I switch between file mode and http mode?” “How do I change the Login Dialogue Panel’s ‘user name’ and ‘project’ defaults?” or read the **stayshun.properties** file itself for a more extensive description of the changes that you can make to this file.

🔍

# Visualization, Viewing Your Data

* How do I visually drill down into the genome?

When you first open a project, it will appear as a “drill down-able tree” in the Genome Viewer Panel in the upper left-hand corner of the *Annotation Station* Main Window. To drill down into the genome, you merely need to expand one of the “nodes” of the tree until you encounter a BAC that you wish to view in greater detail. You must then open this BAC into the Curation-Analysis Tree and Annotated Bio-Sequence Panels where it can be viewed in greater detail. Perform these steps:

Open a project into the Genome Viewer Panel (for more information see “How do I log into the *Annotation Station* and open a project?”)



Expand one of the nodes of the Genome Viewer Panel by clicking on one of the metal “key” icons next to a given chromosome. This will show a list of all the contigs and single BACs that have been assigned to that chromosome.



Click on a BAC that you wish to view in greater detail.



File > Open BAC



Zoom in to the Annotated Bio-Sequence Panel. You may move the left-hand scroll bar up and down to zoom out and in, respectively. If you zoom all the way in, you will see the genomic residues for that BAC. This is the highest level of detail supported by the *Annotation Station*.



* How do I view features on both the forward and reverse DNA strands?

The Main Window is hard-coded to shows features on both the forward and reverse strands. Features on the forward strand are displayed above the axis in the Annotated Bio-Sequence Panel, while features on the reverse strand are displayed below the axis. However, the Gene Editor Window enables you to toggle display of the reverse strand. By default, the reverse strand is not displayed. To display the reverse strand in the Gene Editor Window:

Open a “write enabled” or “read only” Gene Editor Window.



View > Other Strand



Checking the “Other Strand” check box option will display the reverse strand in *both* the Annotation Editing Panel and the mRNA Translation Panel. To stop displaying the reverse strand, simply un-check the “Other Strand” check box.

* How do I view the linear relationship between curations and evidence?

The Main Window and Gene Editor Windows by their very nature display the linear relationship between curations and evidence. However, an additional feature of the *Annotation Station* called the “alignment hairline” was designed to enhance your ability to view the linear relationship between curations and evidence. The “alignment hairline” stretches vertically across the Annotated Bio-Sequence Panel in the Main Window and the Annotation Editing Panel and mRNA Translation Panels in the Gene Editor Window. This “alignment hairline” can be used at any level of zoom to help reveal the alignment between curations and evidence.

* How do I determine if evidence agrees or disagrees with a curation?

While you can always visually inspect a curation and its associated evidence to see if the boundaries are aligned, this is a somewhat tedious and time-consuming process. To rapidly visualize the agreement between a curation and its associated evidence, you must open the curation into the Gene Editor Window. The selected curation will then be represented as a “current working model” in the Annotation Editing Panel, and the ends of all pieces of evidence associated with it will be color-coded to reflect their agreement. Follow these steps:

Click on a curation of interest in the Main Window.



View > Gene



The curation will appear as a “current working model” in the Annotation Editing Panel in the Gene Editor Window. Evidence that agrees with the current working model will be displayed with white ends; evidence that disagrees with the current working model will not be displayed with white ends.



* How do I rapidly view the ID and genomic coordinates of evidence and curations?

The status bars in the lower left-hand corners of the Main Window and Gene Editor Window display a variety of information that you may find relevant. One of the types of information they may be used to display is the ID and genomic coordinates of evidence and curations. To display this information in the status bar:

Click on a curation or piece of evidence in the Main Window or the Gene Editor Window.



Look at the status bar in the lower left-hand corner of the window. It will display the name of the curation/evidence type followed by [yy – zz]. The number ‘yy’ indicates the start coordinates and the number ‘zz’ indicates the end coordinates of the selected curation or evidence.



* How do I view the start and end coordinates for all of the exons in a gene model?

While you may click on individual exons and view their coordinates in the Main Window or Gene Editor Window status bar, this would be a somewhat tedious way to view the start and end coordinates of all of the exons in a 11-exon gene model. For this reason, the *Annotation Station* includes a feature in the Gene Editor Window that can be used to instantly displays a list of the coordinates of all of the exons in a given gene model. To use this feature:

In the Gene Editor Window, make a “current working model” from a gene model of interest. (For more information on performing this task, see “How do I create a working model from a committed curation or gene prediction within the Gene Editor Window?”)



View > Exon Table



The exon table displays a numbered list of exons that includes the exons’ start and end coordinates within the contig, their start and end coordinates within the assembly, and their genomic sequence near the intron/exon boundaries.

* How do I view aligned splice-sites for all of the exons in a gene model?

There is a special feature in the *Annotation Station* designed specifically for this task. This feature is called the Exon Table. In order to pull up the Exon Table, follow these steps:

In the Gene Editor Window, make a “current working model” from a gene model of interest. (For more information on performing this task, see “How do I create a working model from a committed curation or gene prediction within the Gene Editor Window?”)



View > Exon Table



The exon table displays a numbered list of exons. The genomic sequence near the intron/exon boundaries is displayed at the right-hand side of the table. This information may be used to rapidly view and confirm the splice sites for all of the exons in a given gene model.

* How do I view the mRNA sequence for a working gene model?

The mRNA sequence for a working gene model is displayed in the mRNA Translation Panel drawn at the top of the Gene Editor Window. To view the mRNA sequence for a working model, you merely need to zoom all the way in to the mRNA Translation Panel until the mRNA sequence is displayed. Follow these steps:

In the Gene Editor Window, make a “current working model” from a gene model of interest. (For more information on performing this task, see “How do I create a working model from a committed curation or gene prediction within the Gene Editor Window?”)



In the mRNA Translation Panel, move the left-hand scrollbar downwards to “zoom in” to the mRNA sequence. Once you have moved the scrollbar all the way down and are maximally “zoomed in”, the mRNA sequence will be displayed below the axis.



* How do I view start and stop codons in all six frames of a gene model?

The mRNA Translation Panel displays the start and stop codons for a current working model. By default, only the start and stop codons for the three reading frames on the current strand are displayed. To additionally display the start and stop codons for the three reading frames on the opposite strand, you must ask the Gene Editor Window to display the opposite strand:

In the Gene Editor Window, make a “current working model” from a gene model of interest. (For more information on performing this task, see “How do I create a working model from a committed curation or gene prediction within the Gene Editor Window?”)



View > Other Strand



Checking the “Other Strand” check box displays the start and stop codons for the current working model in all six frames in the mRNA Translation Panel.

* How do I view the protein sequence for a gene model translated in three reading frames?

The mRNA Translation Panel can be used to display the hypothetical protein sequence that would be generated by a reading the mRNA sequence in a given frame. To view this sequence, you must be all the way “zoomed in” to the mRNA Translation Panel and you must additionally select a desired reading frame. With a current working model open, follow these steps:

Select a blue start codon in the mRNA Translation Panel that specifies your desired reading frame. A blue line will be drawn between the chosen start codon and the next stop codon, specifying the ORF that will form the basis for the translated protein sequence.



“Zoom in” to the mRNA Translation Panel by moving the left-hand scroll bar as far down as possible in the panel. Once you are fully “zoomed in”, the protein sequence will be displayed below the mRNA sequence below the axis.



* How do I output cDNA and protein sequences to the clipboard?

You can easily output the cDNA and peptide sequences from a current working model to the clipboard for incorporation into database searches. To output the cDNA sequence with a current working model open:

Edit > Copy cDNA to clipboard.



This option will copy the cDNA sequence to the clipboard. In order to copy the peptide sequence to the clipboard, you will need to have selected an ORF in the mRNA Translation Panel. Follow these steps:

Select a blue start codon in the mRNA Translation Panel that specifies your desired reading frame. A blue line will be drawn between the chosen start codon and the next stop codon, specifying the ORF that will form the basis for the translated protein sequence.



Edit > Copy Peptide Sequence to clipboard.



🖊

# Creating and Modifying Curations

* How do I open an existing curation for editing?

To edit a curation, you must import it from the Main Window into a “write enabled” Gene Editor Window:

Choose a curation that you wish to edit by highlighting it in either the Curation-Analysis Tab Panel or the Annotated Bio-Sequence Panel in the Main Window.



Edit > Gene



Choosing this option will open a “write enabled” Gene Editor which will enable you to modify your curation and save the modifications to the database. However, since only one “write enabled” Gene Editor may be open at any time, this option will only be available if you do not already have a “write enabled” Gene Editor open. If you do have a “write enabled” Gene Editor open, you must open curations with “read only” Gene Editor Windows.

* How do I open an existing curation into a “read only” Gene Editor Window?

If you run the Gene Editor Window in “read only” mode, you will still be allowed to use the entire set of Gene Editor editing features. The only difference between “write enabled” and “read only” Gene Editors is that you are not able to commit changes to the database from a “read only” Gene Editor. To open a curation with a “read only” Gene Editor:

Choose a curation that you wish to edit by highlighting it in either the Curation-Analysis Tab Panel or the Annotated Bio-Sequence Panel in the Main Window.



View > Gene



* How do I create a working model from a committed curation or gene prediction from within the Gene Editor Window?

Your primary interaction with the Gene Editor Window is through the use of “working models”. Working models are editable copies of genes that allow you to add and remove exons and modify exon boundaries. You may then commit your working models to the database as formal curations. To create a “working model” from a committed curation or a gene prediction from within the Gene Editor Window:

Choose a committed curation or a gene prediction that you wish to edit by clicking on it the Annotation Editing Panel.



Annotation > New Model from Gene.



Selecting this menu option will create a current working model from the selected gene. You may also use the “Annotation > New Model from Span” option to create a current working model out of a single exon.

* How do I add exons and remove exons from a curation?

There are two ways to add an exon to a curation. You may either add a pre-existing exon from a committed curation, gene prediction, or database hit, or you may add a new exon from scratch. To add a new exon from a committed curation, gene prediction, or database hit:

Choose an exon from a committed curation, gene prediction, or database hit by clicking on it the Annotation Editing Panel.



Edit > Add Selected Exon



To add a new exon to a curation from scratch:

Edit > New Exon…



Choosing this menu option will pull up a text box which you can use to specify the start and stop coordinates of your new exon.

To remove an exon from the current working model:

Choose an exon from the current working model by clicking on it the Annotation Editing Panel.



Edit > Delete Exon



Once you are done editing your working model, you will need to save it to the database if you wish to see your changes reflected in the committed curations. For more information on this, see “How do I add a completed curation to the database?”

* How do I add a completed curation to the database?

Once you have edited one or more working models, you may wish to permanently commit your changes to the database. From a Gene Editor Window that contains edited working models:

File > Commit Changes to Database



Selecting this option will pull up a dialogue box that asks if you want to commit your changes to the database. Selecting ‘yes’ will save all of the edited working models to the database, while selecting ‘no’ will cancel the save. You are also asked if you wish to save your changes to the database any time you close a “write enabled” Gene Editor Window.