

Introduction

The Evolutionary Trace Viewer (ETV) was developed to allow users a quick and easy method to view and run Evolutionary Traces. Prior to development of the ETV, users were required to use separate applications to analyze ET results. The ET Viewer integrates visualization of all the available ET output (and input) into a single application to quickly evaluate Evolutionary Trace results. Furthermore, the ET Viewer provides the ability to initiate new traces and modify existing traces via the ET Wizard.

The Evolutionary Trace method was developed by Olivier Lichtarge to determine evolutionarily conserved residues within a protein family[1]. The method was further developed resulting in the real valued ET [2]. ET residues cluster significantly to elucidate functional portions of a protein [3, 4]. Identifying these important residues offers distinct advantages to protein researchers by narrowing their focus to specific regions of their protein. ET clusters provide specific targets for mutational analysis, competitive inhibition with peptides and other experimental techniques. ET can also be utilized to identify sub-family specific residues by use of the Differential Trace Analysis [5]. This technique compares the ET results of a large super-family trace with a smaller sub-family trace (consisting of sequences present in the super-family) and determines which residues are more conserved within the sub-family.

ET can be performed from a variety of starting points. Typically, a new trace is run by providing ET with a structure file (in PDB format) that is representative of the protein of interest. A trace can also be initiated with an existing multiple sequence alignment (GCG format), phylogenetic tree (NHX format), list of sequence names from BLAST, or a list of sequences in FASTA format. In all of these cases it is most useful to provide a PDB structure since the power of ET is visualization of ET residues on a 3-dimensional structure. In fact, when running ET via the ET Wizard, you are required to provide a structure and/or PDB code to successfully complete ET.

ETV was implemented in Java and consists of four major modules: The Structure Viewer, Tree Viewer, MSF Viewer, and ET Wizard. These four modules are closely integrated so that ET results can be quickly viewed and edited or modified and rerun. Output in the Structure Viewer is synchronized with the Tree Viewer and the MSF Viewer. Interactive selection of sequences in the Tree Viewer can be simultaneously shown in the MSF Viewer (and vice versa). Selection of a set of sub-family sequences can be used to run modified traces, as may be required for sub-family analysis. ETV is launched via Java Web Start technology which uploads the initial application to the client machine and automatically checks for updates during subsequent start ups when a network connection is available. Although the ET Viewer is a Java program and designed to be platform independent, it has only been tested extensively on Windows XP machines.

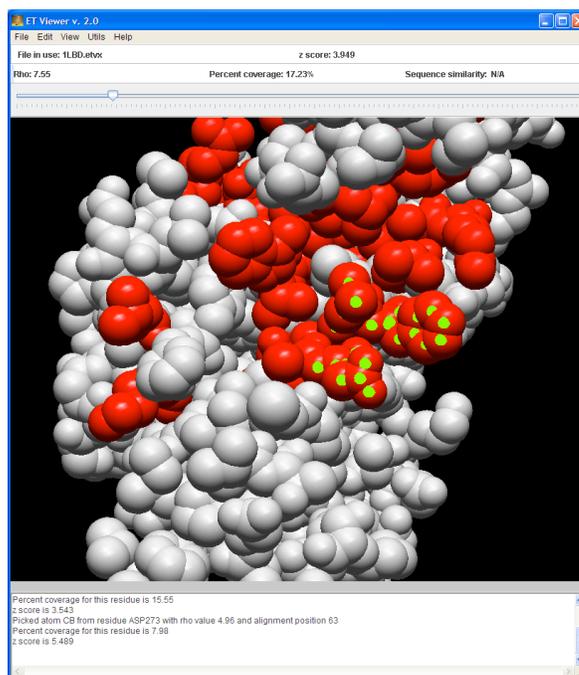
1. Modules

The ET Viewer consists of four main modules: The Structure Viewer, Tree Viewer, MSF Viewer, and ET Wizard. These modules provide closely integrated viewing of ET results and a method to modify existing traces.

1.1. Structure Viewer: The Structure Viewer serves as the entry point for ETV and acts as the parent application for all other content. This module starts up and organizes all of the components for the GUI in their default configuration (i.e. no molecule loaded). At the top of the frame is the menu bar containing the various menu items followed by informative text describing the loaded file and ET rank statistics. The majority of the frame consists of the Structure Viewer panel which displays the loaded trace results. The bottom section displays text messages, updating the user on file loading and descriptions of picked atoms.

Once an ETV file is loaded, the user can move the slider to view trace results as they pertain to each rank. Slider position adjusts rank and percent coverage output. Percent coverage is the number of ET residues divided by the total number of residues in the structure. The rank is indicated by either “Rank” for integer value traces or “Rho” for real value traces. Most current traces have a statistical measure (z-scores) associated with them to provide quality assessment of each trace rank. Typically maximum z-scores are obtained at around 15%-30% coverage which corresponds to the best clustering statistics. We have found that z-scores > 3.0 that change smoothly with no large spikes are indicative of a good quality trace[6].

ET Viewer files (*.etv or *.etvx) are simply a concatenation of ET files. Users can modify this file if needed if they adhere to the following rules ([Figure 2](#)). ETV files begin with the PDB structure file. ET can only be run on single chains of a PDB structure and therefore the ETV file contains only a single chain. Some users will want to visualize their trace results with other portions of the PDB file present (such as ligands or other subunits in multimers). This is accomplished by replacing the single chain PDB with the full length original or any portion of the full length original. Keep in mind, however, that the PDB has no formatting rules prohibiting duplicate residue numbers in different protein chains. This can cause problems with the ET Viewer since it attempts to match residue numbers from the .ranks file with residue numbers in the structure file. Another thing to consider is the total number of atoms present in the structure. If there are too many, ETV will suffer in performance, depending on the user’s computer. After the structure information comes the .ranks file information. This contains the rank data from ET for each residue in the structure. It is not recommended you change this as it represents the actual results from ET. Some users will, however, want to renumber the residues in the .ranks file to match the residue numbering of an alternate structure file. The ETV will not load the file unless the residue numbering of the structure and .ranks file matches. Users may notice that some of the rows in the ranks portion contain “-” rather than residue numbers. These are positions from the multiple sequence alignment that do not map onto the structure. The final required part of the ETV file is the NHX formatted tree as output from ET. Every trace has a tree file associated with it and should never be manually altered. An optional addendum to ETV files is the cluster_report.summary file, which provides data on the z-scores for each trace rank and residue. This is generated automatically when each trace is run to aid the user in evaluating ET quality while viewing trace results in the ETV. If, for some reason, the user obtains ET results that do not include an .etv or .etvx file, ETV has a utility that will make this file automatically. When used, the Make ETV utility will query the user for the directory that contains the required files and concatenate them into a .etv or .etvx file. The user should make sure that the directory contains only one each of a .pss.nhx file, PDB file, .ranks file, and (optionally) .cluster_report.summary file. Additionally, the user should ensure that these files are from the same ET otherwise the results will be unreliable.



- 1.1.1. Mouse Control: Molecule manipulation is controlled via the mouse (see [Table 1](#)). The left mouse button controls protein rotation and atom picking. The right mouse button controls translation. On Windows computers, holding the left and right mouse buttons down together will zoom the protein. On other operating systems, holding the left mouse button and the Alt key will provide zooming capability.
- 1.1.2. Viewing: Selection of a few different viewing modes is available via the menu bar ([Table 2](#)). They are designed to aid the user in more easily identifying important residues, but ETV does not have more complicated viewing options such as ribbons. We

do provide macro writing utilities for PyMol and MolMol to export particular rank views for further modification as desired. **Any text output by ET Viewer utilities is viewed in the Report Viewer. If the user wishes to save the output, a single menu item is associated with the Report Viewer that allows the file to be saved. Alternatively, the user can Copy the text and paste it into another text editor.**

Table 1

Left Mouse button (hold down)	Protein rotation on x and y axis.
Left Mouse button (Click at bottom of viewing window and drag left or right)	Rotate protein on z-axis.
Left Mouse button + Right Mouse button	Moving the mouse up or down will zoom the molecule in or out.
Left Mouse button (Click)	When positioned over an atom, a left mouse button click will illicit descriptive text to be output into the status text box.
Right mouse button (hold down)	Used for translating the protein along x and y axis.

Table 2

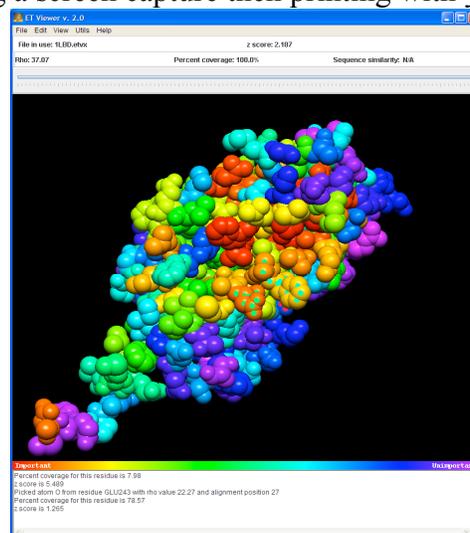
File	
Open ETV Results	Pops up an Open File dialog in the user's home directory. A file filter is defaulted to <filename.etv> or <filename.etvx>
Open Alignment File	Open a GCG or FASTA formatted multiple sequence alignment. An alignment can be opened and viewed independently of ET results. If an MSF file is opened while trace results are being viewed, ETV will check to see if they are from the same trace and attempt to synchronize the structure viewer with the MSF viewer.
Close ETV File	Removes the current file from memory and resets ETV to its default configuration.
Image Export	Currently supports exporting to jpeg and png formats. Currently this utility doesn't work very well. It is recommended that users use a screen capture or utilize the PyMol or MolMol scripts to generate pictures.
Print and Print Preview	Provides a method to print out to a printer, although this doesn't work so well either. It is recommended to capture the current view then use photo editing software to print or edit the image (as an image export). Future versions of ETV should contain better output control.
Exit	Shuts ETV down.
Edit	
Search	Inactive. Will be used in future releases.
Residue Selection	Lists every residue in the current structure. Selecting by residue number allows the user to quickly obtain ET information on a specific residue and alleviates attempting to find the residue by clicking on the structure.
View	
Gobstopper Color	Colors trace residues according to a prismatic coloring scheme. Most important residues are red. Least important residues are magenta.
Color by Cluster	Trace residues are colored in order of cluster size. The largest cluster is colored red. Individual residues (no cluster affiliation) are colored black. Other clusters are colored according to decreasing size as follows: blue, orchid, green, purple, cyan, turquoise, brown, coral, magenta, orange, sky blue, violet, gold
Segment Focus	Turns off unselected atoms to allow viewing of specific residues only.
Background	Dialog to change background color. Default is black, but a white background may be preferred for posters or journal figures. Every other color is available as well, but these are used only at the user's discretion.

Backbone	Makes the side chain atoms invisible.
Bonds	Shifts view mode so that the protein residues are represented as bond lines if they are not at the current rank. Residues that are important at the current rank are shown as spheres.
Spacefill	Returns view mode to spheres rather than bond lines.
Residue Ranks	Outputs the currently open ET results ranks file to a text viewer. This file lists the rank of every alignment position of the current trace along with the structure's residue numbers. If the user wishes to save the output, a single menu item is associated with the Report Viewer that allows the file to be saved. Alternatively, the user can Copy the text and paste it into another text editor.
ET Tree	Selecting this toggles the Tree Viewer on and off.
Utils	
Make ETV File	Starts dialog to automatically generate an ETV or ETVX file from existing ET results. This is normally not needed as all traces generated from the ET Wizard come packaged with the ETVX file.
ET Wizard	Launches the ET Wizard to perform an Evolutionary Trace.
Write PyMol/MolMol Macro	Outputs a macro script of the current ETV state. This can be saved and used as input to PyMol or MolMol. This is useful when more advanced molecular viewing modes are desired based on ET results. Any text output by ET Viewer utilities is viewed in the Report Viewer. If the user wishes to save the output, a single menu item is associated with the Report Viewer that allows the file to be saved. Alternatively, the user can Copy the text and paste it into another text editor.
Peptide Builder	Inactive. Will be available in future releases.
Help	
Tutorials	A web based tutorial to aid the user with common ETV tasks. These are also available directly from the ET Viewer website (http://mammoth.bcm.edu/traceview/).
About	Provides author and version information about ETV.

1.1.3. Image export: ETV has a rather rudimentary image export capability (File menu). It produces a fair reproduction of the current view in ETV. Two file formats are available: jpeg and png. The image export does not work correctly on Macintosh computers in this release of ETV (the color mappings are wrong).

1.1.4. Printing: Printing and Print preview are also rather rudimentary at this time. We recommend adjusting the view in ETV and doing a screen capture then printing with your favorite photo editing software. Alternatively, the user can output scripts (for PyMol or MolMol) and print via those applications.

1.1.5. Residue selection: Residue selection can be accomplished two ways. The first is via the residue selection dialog (found under the Edit menu). The selection dialog allows the user to select specific amino acids by residue number. The other method is by mouse selection. Either method will highlight the selected amino acid and give ET details of that amino acid in the status window. Currently, the mouse picker works best when the protein is shown in full space fill mode. Clicking the protein in



backbone or bonds mode will still select atoms that are not visible, causing misleading information to be output in the status window. To deselect a residue, merely click the amino acid again, or uncheck the box next to it in the selection dialog.

1.1.6. Viewing modes: ETV has a few basic viewing modes to aid in evaluation of trace results.

1.1.6.1. Gobstopper color: This coloring scheme is best visualized with the rank slider all the way to the right. Prismatic coloring allows the user to identify the relative ranks of every residue in the protein by their color. The order is red-orange-yellow-green-blue-magenta, with red amino acids being the most important residues and magenta residues being the least important.

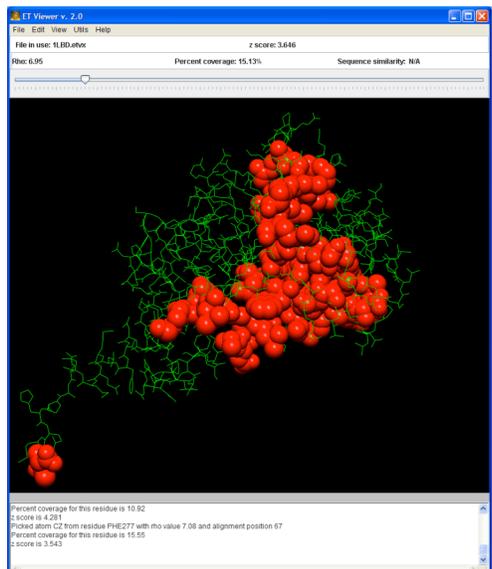
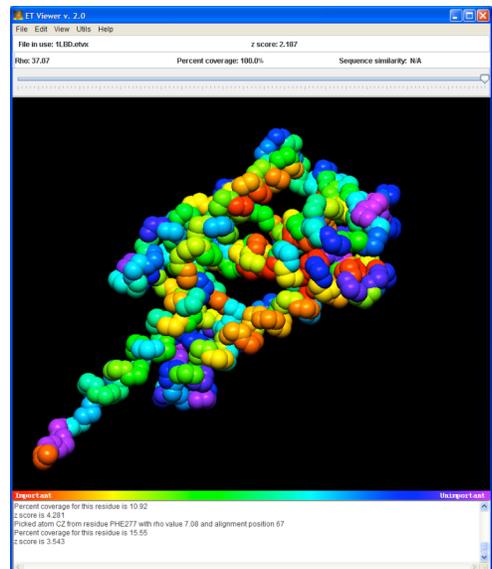
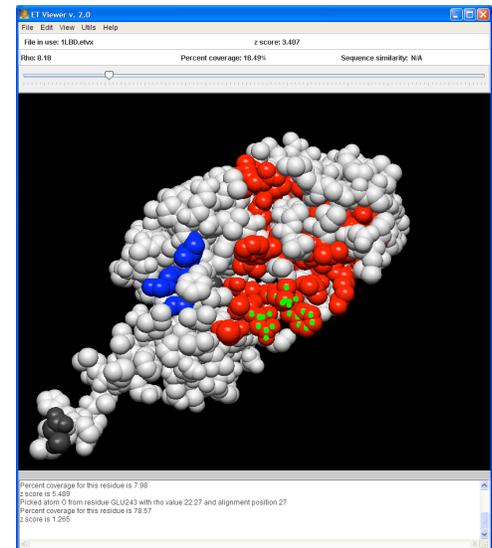
1.1.6.2. Color by cluster: When the protein is visualized with this coloring scheme, trace residues are colored according to how large the clusters are at a particular rank. The clusters are colored in order of descending size as follows: red, blue, orchid, green, purple, cyan, turquoise, brown, coral, magenta, orange, sky blue, violet, gold. When a residue is not of a cluster, it is shown as black.

1.1.6.3. Background: The background dialog allows the user to customize the background color of the viewer.

1.1.6.4. Backbone: To view trace residues that reside in the interior of the protein, users can select backbone viewing mode. This mode shows only the backbone atoms of each residue.

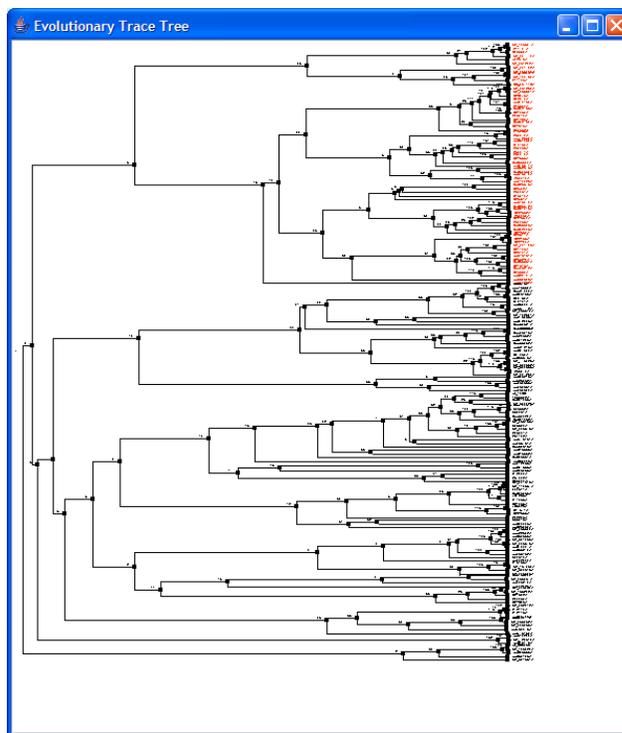
1.1.6.5. Viewing modes: The user has a choice of two viewing modes: space fill or bonds. Space fill simply draws a sphere of Van der Waals radius for each atom while bonds mode draws a line between each atom. Residues that are ranked at the current rank slider position will appear as spheres while the remainder will be shown as bonds.

1.1.6.6. Residue Ranks: One of the files output from the Evolutionary Trace is a ranks file. This file includes the calculated rank of every alignment position from the alignment file that is input into ET. Also include in this file are the structure's residue numbers and variability information. Selecting 'Residue Ranks' from the 'View' menu will output this text file in a report viewer.



1.2. Tree Viewer

The Tree Viewer is utilized to view the phylogenetic tree that is associated with each Evolutionary Trace. Our Tree Viewer uses code from ATV [7] to read in tree data that conforms to .nhx formatting rules. A modified ATV is also used to display the ET Tree. The Tree Viewer is integrated with the ET Viewer and MSF Viewer. When viewing integer trace results, the current rank is shown on the tree by means of a vertical blue bar. This bar does not appear when viewing real value traces (the default) since the tree does not directly correlate with the ranks (ρ) [2]. In either case, examination of the underlying tree that each trace produces gives the user the ability to visualize the associations of each sequence that appear in the supplied multiple sequence alignment. The presence of single branches that appear to be widely separated from the other sequences might indicate that this branch could be pruned.



- 1.2.1. Node and sequence selection: Individual sequences or sub-branches can be selected by left mouse clicking the nodes (as indicated by small boxes at each branch point). Selected sequences are displayed as red text. Selections are linked to the MSF Viewer ([section 1.3](#)) when the associated multiple sequence alignment is loaded.
- 1.2.2. Zooming: Many traces have trees that are quite large. Viewing the entire tree causes the sequence names to become unreadable. You can make the tree larger by pressing the '=' key. Zooming back out is accomplished by using the '-' key.
- 1.2.3. Printing: Printing in the Tree Viewer has not been enabled. Printing can be accomplished by screen captures. As some trees can be quite large, it might be necessary to zoom in ('=' key) and perform several captures of different parts of the tree. After printing out the panels in an image editing program, they can be taped together to form a banner.

1.3. MSF Viewer: The MSF Viewer provides a means to view multiple sequence alignments. This module is unique in that it can be used independently of ETV files so that new traces can

be run from existing alignments. The alignment loader is from PFAAT v1.2 [8] and is used to load GCG or FASTA formatted alignment files. When an alignment is opened with its corresponding ETV file, rank bars above each column in the alignment provide indication of each residue's relative rank. The sequence that matches the structure will be highlighted (if it is available).



- 1.3.1. Coloring schemes: The MSF Viewer has three coloring schemes available: Default (see [Table 3](#)), ClustalX [9, 10], and black and white. Depending on your circumstances, these schemes can aid you in evaluating your alignment.

- 1.3.2. Sequence selection and saving: Sequences can be selected by clicking on sequence names individually. Selected sequences are shown in red. If the current multiple sequence file is linked with the current ETV file, then selected sequences in the MSF Viewer will also be selected in the Tree Viewer. You can save the selected sequences (from the MSF Viewer File menu item) for input to run new traces. If you just need to remove a few sequences, choose 'Select All' from the MSF Viewer 'Edit' menu and then deselect the sequence names you want to remove. Alternatively, you may want to run traces on sub-branches of an existing trace. Sub-branches chosen in the Tree Viewer (section 1.21) will be selected to be saved in the MSF Viewer. Selected sequences are saved in FASTA format as a .input file. This file can subsequently be used as input for running a new trace using the ET Wizard (section 1.4). When used as input to the ET Wizard, these sequences will be realigned using ClustalW.

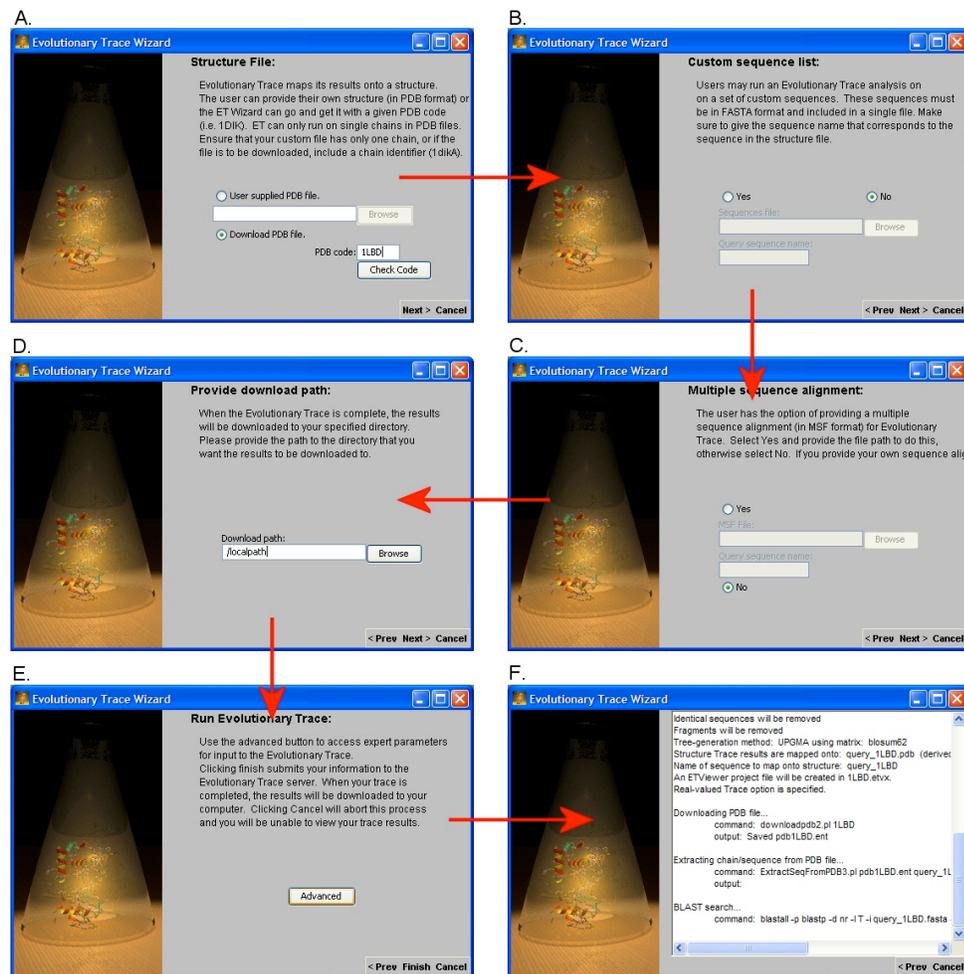
Table 3

Default Color	Residues
Orange	G, P, S, T
Red	H, K, R
Blue	F, W, Y
Green	I, L, M, V
No color	A, C, D, E, F, P

- 1.4. ET Wizard: The ET Wizard is used to gather required information for running traces on our server. Required and optional information is gathered from the user and submitted for processing via our ET Server. Status messages are displayed in the ET Wizard as the trace is being run and the resulting trace files are zipped and downloaded back to the user's computer when the trace is complete. The ET Wizard is responsible for the download and should stay open until the trace is complete in order to retrieve the trace results.
- 1.4.1. Minimum input: The minimum input for running ET is a PDB formatted file or valid PDB code ([Figure 1A](#)). The ET Wizard can check a supplied PDB code for validity and supply any existing chain information via the Protein Data Bank at <http://www.rcsb.org/pdb/>. The Evolutionary Trace can only run on single chains in the structure file thus the supplied PDB code needs to contain a chain specifier. If no chain is specified, ET defaults to the first chain in the PDB file (or chain A). When supplying a custom structure, if more than one chain exists in the file, then ET will always trace the first chain.

Example: 1AO1 is the PDB code for a deoxyhemoglobin A which has 4 chains: A, B, C, and D. In order to run a trace on chain B of this structure, the user should input 1AO1B or 1ao1b. In all cases, whether the user supplies the PDB file or not, the PDB code needs to be supplied to the ET Wizard.

Figure 1. The ET Wizard can be used to submit information for running traces. **A.** On the first page, the user is required to provide either the path to a PDB file or the code to be downloaded ([see example](#)). In either case a 4 character code should be provided. Supplied structures should contain a single chain to be traced. **B.** On this page the user can supply a custom set of sequences. These sequences MUST be in unaligned FASTA format and must be included in a single file ([see example Figure 3](#)). If the user already has an aligned set of sequences, item **B** should be left as “No” and the alignment provided on the next page (**C**). In both cases, whether the set of sequences is aligned or not, the user must provide the name of the sequence that corresponds with the structure file provided on page **A**. **This sequence must be identical to the structure file or the Evolutionary Trace will not work. Also ensure that the header for each sequence (in the .input file) follows the BLAST format, that is the greater-than symbol '>' followed by the sequence name or identifier followed by a space.** **D.** ET Wizard needs to know the local path to which the user would like the ET results downloaded to. **E.** Pressing the advanced button gives the user access to many features not fully supported at this time. Most notable, however, is the Real Value trace option. Default is to run a Real Value trace, but users can switch to integer value trace by un-checking the Real Value option. Pressing Finish takes the user to the status page (**F**) which displays the progress of the ET. The user should not close the ET Wizard until the trace is complete since it is responsible for downloading the results of the Evolutionary Trace. The results are downloaded as a .zip file. Included in the .zip file are all of the generated data files from the Evolutionary Trace and the .etv/.etvx file which is formatted so that the ET Viewer can read it. Also included is a log file which details the output from the Evolutionary Trace server. This file is important for keeping track of the settings used to run your trace as well as providing a means to troubleshoot the trace should it fail to run correctly.



PDB code. Also, the residues in the PDB file must match the reference sequence in the BLAST results file or tree file.

- 1.4.4. No structure solution: Very often researchers are interested in a protein which has not had a structure determined yet. In this scenario, searching the protein's sequence against the PDB using BLAST will return the closest homologs which can be submitted as a representative structure. The ET Wizard requires a structure in order to submit a trace.

Figure 3: Sample .input file

```
>AAC15588-1
GAPPEMPLERILEAELRVESQTGTLSESAQQQDPVSSICQAADRQLHQVLQWAKHIPHFEELPLEDRMVLLKAGWNELLIAAFSHRSVDVDRDGI
LATGLVVQRHSAHGAGVGAI FDRVLTTELVAKMREMKMDRTELGLCLAVVLFNPEAKGLRTPCGSGGPEGESVSALEEHCRQYYPDQPGRFAKLLLR
LPALRSIGLKCLEHLFFFKLIGDTPIDNFFLLSMLEAFSD
>AAC03056-1
GPGRDITVERLMEADQMSEARCGDKSIQYLRVAASNTMIPPEYRAPVSAICAMVNKQVFQHMDFCRRLPHFTKPLNDQMYLLKQSLNELLILNI
AYMSIQYVEPDRRNADGSLERRQISQQMCLSRNYTLGRNMAVQAGVVQIFDRILSELSVKMKRLDLDATLCLLKSIVVFNPDVTRLDDRKSIDL
LRSRIYASLDEYCRQKHPNEDGRFAQLLLRLPALRSISLKCLDHLFFYQLIDDKNVENSIVIEEFHKL
>NP_571292-1
SFNEEMPVEKILDAELAVEPKTEAYMESSMSNSTNDPVTNICQAADKQLFTLVEWAKRIPHFSDDLPLDDQVILLRAGWNELLIASFSHRSVTVKD
GILLATGLHVHRSSAHSAGVGSIFDRVLTTELVS KM RDMQMDKSELGCLRAIVLFPNDAKGLSNPSEVEALREKVVYASLEGYTKHNYPDQPGRFAK
LLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQI
>AAA40081-1
GAPEEMPVDRILEAELAVEQKSDQGVGPGATGGGGSSPNDPVTNICQAADKQLFTLVEWAKRIPHFSDDLPLDDQVILLRAGWNELLIASFSHR
IDVDRGILLATGLHVHRNSAHSAGVGSIFDRVLTTELVS KM RDMRMDKTELGLCLRAIMFNPDAKGLSNPGEVEILREKVVYASLETYCKQKYPEQQ
GRFAKLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQI
>P28701-1
GGSEDMPVERILEAELAVEPKTEAYS DVNTESSNDPVTNICHAADKQLFTLVEWAKRIPHFSDDLPLEDQVILLRAGWNELLIASFSHRSVSQD
GILLATGLHVHRSSAHSAGVGSIFDRVLTTELVS KM RDMQMDKSELGCLRAIVLFPNDAKGLSSPSEVESLREKVVYATLEAYTKQKYPEQPGRFAK
LLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPLQV
>AAG24886-1
SSVRDVTIERIHEAQLSEQKSGDNAIPYLRVGSNSMIPPEYKGA VSHLCQMVNKIYQLIDFARRVPHFINLPRDDQVMLLRGCWNEMLIAAVA
WRSMEYIETERSDDGSRITVRQPQLMCLGPNFTLHRNSAQAGVDTLFDRI LCELGIKMKRLDVTRAEGLVKAIILFNPDIRGLKCQKEIDGMR
EKIYACLDEHCKQHPSEDGRFAQLLLRLPALRSISLKCLDHLNFI RLLSDKHLDSFIVEMLDMP
>NP_033133-1
SSHEMMPVERILEAELAVEPKTESYGMNVENSTNDPVTNICHAADKQLFTLVEWAKRIPHFSDDLPLEDQVILLRAGWNELLIASFSHRSVSQD
GILLATGLHVHRSSAHSAGVGSIFDRVLTTELVS KM RDMQMDKSELGCLRAIVLFPNDAKGLSNPSEVETLREKVVYATLEAYTKQKYPEQPGRFAK
LLRLPALRSIGLKCLEHLFFFKLIGDTPIDSFLLMEMLETPLQI
>AAC15589-1
PGSPDMPLERIEAEMRVEQPAPSVLAQTAASGRDPVNSMCQAAPLHELHVQWARRIPHFEELPIEDRTALLKAGWNELLIAAFSHRSVAVRDGI
VLATGLVVQRHSAHGAGVGDI FDRVLAELVAKMRDMKMDKTELGLCLRAVLFNPDAGLNRATREALREKVVYAALEEHCRRHHPDQPGRFGKLL
LRLPALRSIGLKCLEHLFFFKLIGDTPIDSFLLNMLEAPD
>P54779-1
SSVQELSIERLLEIESLVADPPEEFQFLRVGPESGVPAKYRAPVSSLCQIGNKQIAALVVWARDIPHFGQLELEDQILLIKNSWNELLFIAIWR
SMEYLTDERENVDSRSTAPPQLMCLMPGMTLHRNSALQAGVGQIFDRVLSLSLKMRTL RMDQAEYVALKAIILLNPDVKGKLNKPEVVVLRK
FSCLEDEYVRRSRCAEEGRFAALLRLPALRSISLKCFEHL YFFHVLVADTSIASYIHDALRNHAPS
```

References

1. Lichtarge, O., H.R. Bourne, and F.E. Cohen, *An evolutionary trace method defines binding surfaces common to protein families*. J Mol Biol, 1996. **257**(2): p. 342-58.
2. Mihalek, I., I. Res, and O. Lichtarge, *A family of evolution-entropy hybrid methods for ranking protein residues by importance*. J Mol Biol, 2004. **336**(5): p. 1265-82.
3. Yao, H., et al., *An accurate, sensitive, and scalable method to identify functional sites in protein structures*. J Mol Biol, 2003. **326**(1): p. 255-61.
4. Madabushi, S., et al., *Structural clusters of evolutionary trace residues are statistically significant and common in proteins*. J Mol Biol, 2002. **316**(1): p. 139-54.
5. Madabushi, S., et al., *Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions*. J Biol Chem, 2004. **279**(9): p. 8126-32.
6. Mihalek, I., et al., *Combining inference from evolution and geometric probability in protein structure evaluation*. J Mol Biol, 2003. **331**(1): p. 263-79.
7. Zmasek, C.M. and S.R. Eddy, *ATV: display and manipulation of annotated phylogenetic trees*. Bioinformatics, 2001. **17**(4): p. 383-4.
8. Johnson, J.M., et al., *Protein family annotation in a multiple alignment viewer*. Bioinformatics, 2003. **19**(4): p. 544-5.
9. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
10. Thompson, J.D., et al., *The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools*. Nucleic Acids Res, 1997. **25**(24): p. 4876-82.