1 User-Level Documentation

1.1 Getting Started

The AliShredderWeb application has been deployed on the CIBIV Server and is accessible via the following link: `http://alishredder.cibiv.univie.ac.at/`. This interface is a web-based tool for measuring phylogenetic quality. The following web browsers were tested on the site:

- Chrome (Version 122.0)

- Firefox (Version 123.0)

- Microsoft Edge (Version 121.0)

This documentation will serve as a step-by-step guide for first-time users.

1.2 Main Interface

When accessing the AliShredder Web interface (see Figure 1), users will land on the main page, which features two primary sections: the Upload Section (1.) and the Options Section (2.).

Figure 1: AliShredderWeb Mainpage

In the Upload Section, users can upload files for analysis, with supported formats including PHYLIP and FASTA. Following successful upload, users can select from various computation and window-related options in the AliShredder software. Regarding window options, users can specify window size, step size, and start and end positions. Additionally, there's an option to configure windows to strictly adhere to the specified start and end positions or expand beyond the subalignment. If no options are selected, the Al-

iShredder software will default to its standard settings for analysis. For the analysis the user then specifies which computation option to proceed with:

- Select all options: analyzes all options

- Compute the information type: Displays different types of each site in the entire alignment.

- Completeness score for the alignment/windows: Outputs a single line indicating the completeness of the entire alignment.

- Completeness score for individual sequences: Computes and displays the informative sites for each sequence in the alignment.

- Completeness score for individual sites: Outputs a logfile and respective files, distinguishing completeness between each individual site in the alignment

- Completeness score for individual patterns: Outputs a logfile and respective files, showing different patterns in each site in the alignment.

- Complentess/Incompleteness score for pair of sequences: Compares each possible sequence pair and computes their completeness/incompleteness with each other.

- P-Distance for pair of sequences: Computes the similarity/dissimilarity between each possible combination of sequence pairs in the entire alignment.

After selection, users can submit their choices by clicking the submit button and will be directed to the Processing Page for further action.

1.3 Processing

Upon completion of file upload and selection of analysis options, users will be redirected to a processing page (see Figure 2). Here, users have the option to either wait for the analysis to finish and be automatically redirected to the success/fail page, or they can choose to obtain the data later via provided link for analysis within a 24-hour timeframe.
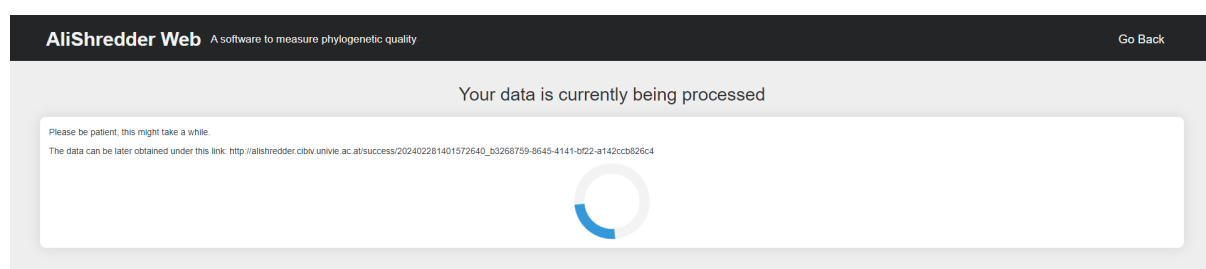

Figure 2: Processing Page

1.4 Success Page

Upon successful completion of the analysis, the processing page will redirect users to the Success page (see Figure 3), which serves as a dashboard for the analyzed data.

Figure 3: Success Page

The success page features a dropdown menu for changing visualizations , line plots displaying the data and a download area. The interactive line graphs allow users to pan, isolate traces, hover for details, zoom in/out, draw inside the graph, toggle spike lines, and download the adapted graph as a PNG file.

The dropdown menu on the success page provides various types of visualizations for the data, enhancing user experience by eliminating the need to return to the main page to change the computation option. This feature allows users to seamlessly explore different visual representations of the analyzed data without interruption.

In the download section, users can obtain the output data if necessary. The TSV files include the results of each analysis option and the corresponding data, while the log file contains the log output generated by the AliShredder software during the analysis process. This allows users to access and review the detailed analysis results and software logs for further examination or documentation.

1.5 Fail Page

If an error occurs during the analysis process, users will be redirected to a Fail Page (see Figure 4), which displays the respective log output generated by AliShredder. This provides users with detailed information about the encountered error, aiding in troubleshooting and resolving issues effectively.
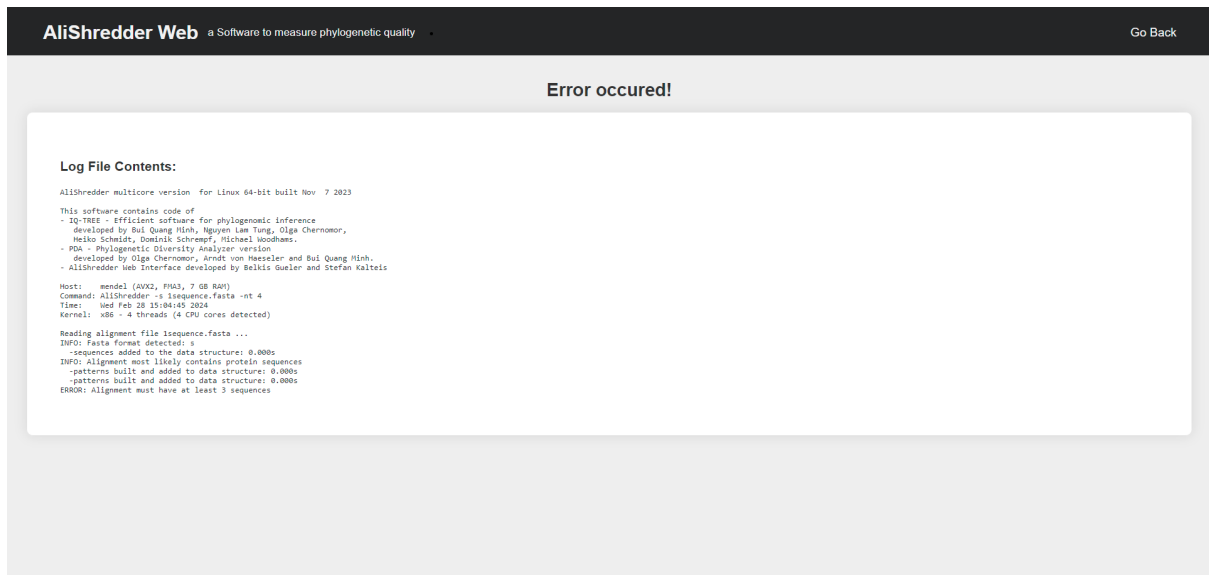
Figure 4: Fail Page

The log output contains the execution details and processing steps of the AliShredder program, including the input parameters, host system information, and any encountered errors during execution.

1.6 Troubleshooting

When using the AliShredder Web, it's essential to consider certain constraints in file upload and analysis options to prevent potential errors. The following checklist outlines key points to address when encountering errors:

- Ensure you are using a valid file format: AliShredder supports specific file types such as PHYLIP and FASTA. Verify the format of your file during upload to avoid errors.

- Confirm that your alignment contains at least four sequences: AliShredder requires a minimum of three sequences for analysis to proceed successfully. Check your alignment to ensure it meets this requirement before submission.
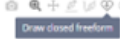
If errors persist despite validating the file format and alignment, refer to the error message provided by the software on the Fail page for further troubleshooting.

1.7 Using the Software

1.7.1 Understanding Plot Navigation

To effectively navigate the plots and customize the visualization of the analysis, various Plotly modebars (1) are provided. This table (Table 1) offers detailed explanations of each modebar option:

Table 1: Plot Modebar Options

Icons	Explanation
	Download plot as a png: After completing changes in the plot, it can be downloaded through this download button.
	Zoom: Clicking and holding with the mouse allows for zoom.
	Pan: to pan across regions of the graph, click and hold the mouse to explore the data. Double-click anywhere to return to the original view.
	Draw Line: allows drawing straight lines on the plot.
	Open Freeform: allows drawing freehand shapes or lines on the plot.
	Closed Freeform: similar to Open Freeform Mode, but it automatically closes the shape you draw.
	Draw Circle: this tool lets you draw circles or ellipses on your plot.
	Draw Rectangle: with this tool, you can draw rectangles or squares on the plot. Click and drag to define the starting point and size of the rectangle.
	Erase active shape: this tool allows removing shapes from the plot. Simply click on a shape you want to erase, and it will be deleted.
	Zoom in/out: You can zoom in and out by clicking on the + and - buttons.
	Autoscale: zooms the plot back to a setting that is optimized to include all the viewable data regardless of the Axes Range setting.
	Reset Axes: zooms to include the Axes Range, if this has been set. If it has not been set it zooms to a setting that is optimized to include all the viewable data, the same as if Autoscale had been clicked.
	Toggle spike lines: This tool enables users to show or hide vertical lines extending from data points to the axes for improved visualization of specific values.

1.7.2 Overview of available Plots

1. Informative Sites: When selecting the informative sites plot, users will view the proportions of different site types (e.g., Fully Informative Sites, Partly Informative Sites) represented as percentages of total sites. By default, the display of average distance and completeness across the entire alignment is deactivated. However, users have the option to activate and visualize these two calculations within the informative sites plot if desired.
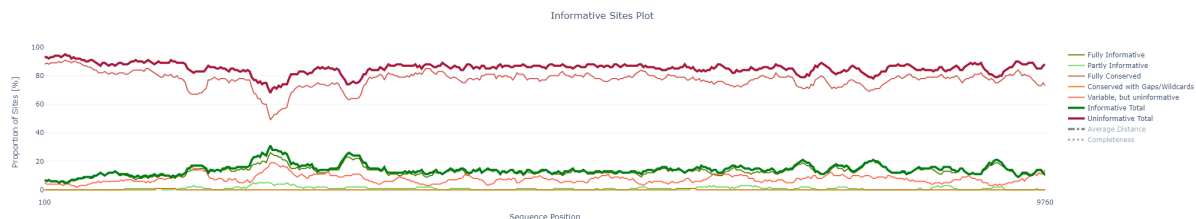
Figure 5: Informative Sites Plots

2. Informative Sites (IQ-Tree): This Plots shows the different types of sites of the entire alignment with the calculations generated by IQ-Tree.
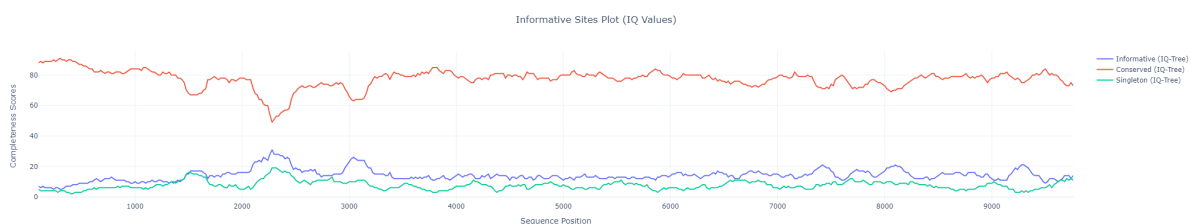

Figure 6: Informative Sites (IQ-Tree) Plot

3. P-Distance: By default, this plot visualizes the average distance across all sequence pairs, representing the similarity or dissimilarity between them. Users can toggle between viewing all sequences or individual sequences, with the option to display the average distance alongside the P-distance for each sequence.
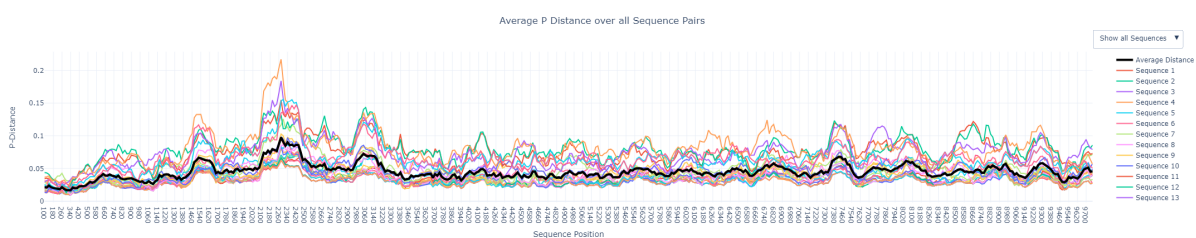

Figure 7: P-Distance

4. Completeness of Sequence Pairs: This plot illustrates the completeness of each sequence across all sequence pairs, with values ranging from 0 to 1. A value of 1 indicates that the sequence position contains fully informative sites, while a value of 0 indicates the absence of informative sites at that position.
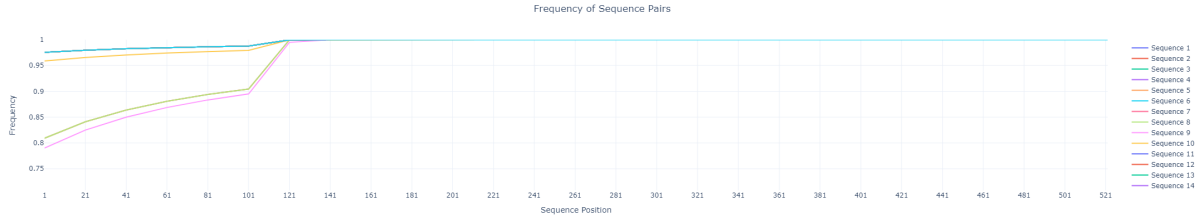
6

Figure 8: Completeness of Sequence Pairs Plot

5. Incompleteness of pair of sequences: This plot illustrates the inverse of the completeness of sequence pairs. A value of 0 indicates that the sequence position contains full information, whereas a value of 1 signifies the absence of any information.
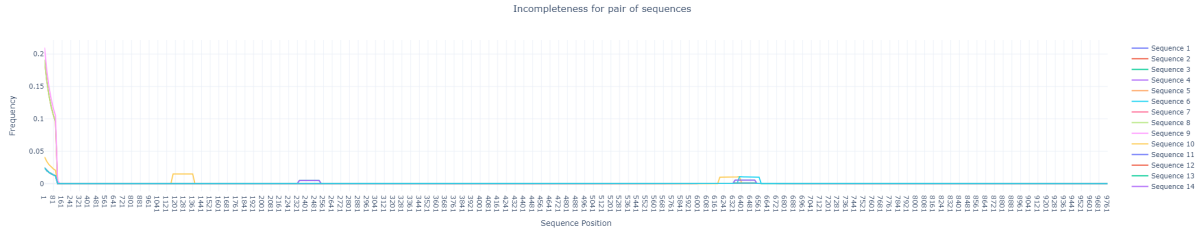

Figure 9: Incompleteness of pair of sequences

6. Informative Sites for Individual Sequences: In the last plot, each sequence's informative sites are depicted. A value of 1 indicates that the chosen sequence contains fully informative sites at that position, while a value of 0 signifies the absence of informative sites.
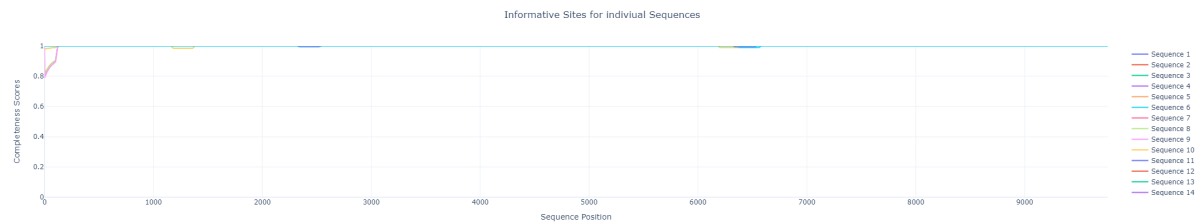

Figure 10: Informative Sites for Individual Sequences Plot

References

[1] Plotly Modebar Documentation: `https://plotly.com/python/reference/layout/`