USER MANUAL

BLOOM (Barcode Linker for Organism and Ontology Mapping): *In Silico* Evaluation of DNA Barcode Accuracy and Efficiency for Taxa Identification in Plant DNA Metabarcoding.

Neudek, M., Serrano Velazquez, M., Spriggs, M., Twigger, K.,

Cranfield University, Cranfield, MK43 0AL, England

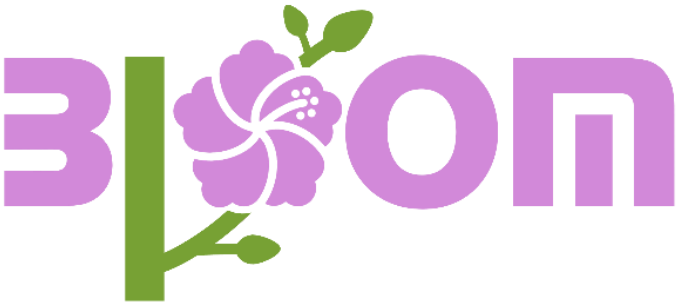

Table of contents

1. Introduction

The Barcode Linker for Organism and Ontology Mapping (BLOOM) tool provides the user with a graphical user interface (GUI) to assess the accuracy and efficiency of universal DNA barcodes for identifying the desired taxa. The underlying pipeline of BLOOM starts with searching the National Center of Biotechnology Information (NCBI) databases to match sequences based on nucleotide position and primers, to return a list of barcode sequences. An NCBI BLAST search is followed to identify the sequences in the BLAST core nucleotide database, which are similar to the barcode sequence. BLOOM presents the results in a results tab and provides the user the option to create a CSV or visualise the taxonomical relationships between the similar sequence species and the original species of interest. The purpose of BLOOM was designed to be used prior to wet lab work of metabarcoding method, and to be user-friendly to install and navigate for the user.

2. Download and Installation

BLOOM was designed to be easy for the user to access, therefore the tool has been compiled into an executable file, which can be downloaded by the user form the following link: https://github.com/ms2206/BLOOM/raw/refs/heads/main/BLOOM.exe?download=.

Figure 1 demonstrates the display of the 'BLOOM.exe' in your downloaded folder on your local machine. To activate the BLOOM app, double click on the 'BLOOM.exe' file.

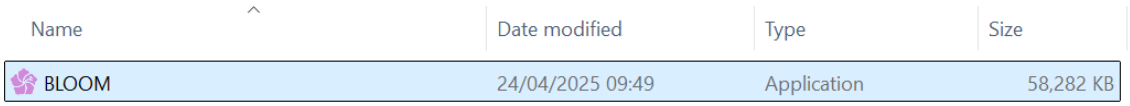

Figure 1. Display of the BLOOM.exe executable file on a local machine. *This figure shows the 'BLOOM.exe' file as it appears in the user-specified folder following a successful download.*

Once BLOOM has been activated on the user's local machine, a separate pop-up window, as shown in Figure 2 is displayed to the user. This indicates the BLOOM app is fully installed and ready to be used. The BLOOM application is a Windows PC application only, so installation should occur on a Windows only PC. There is also an option to access the BLOOM application for a non-Windows PC (MacOs and Linux) through a virtual environment, which has been assumed the user already has a virtual environment set up on their local machine.

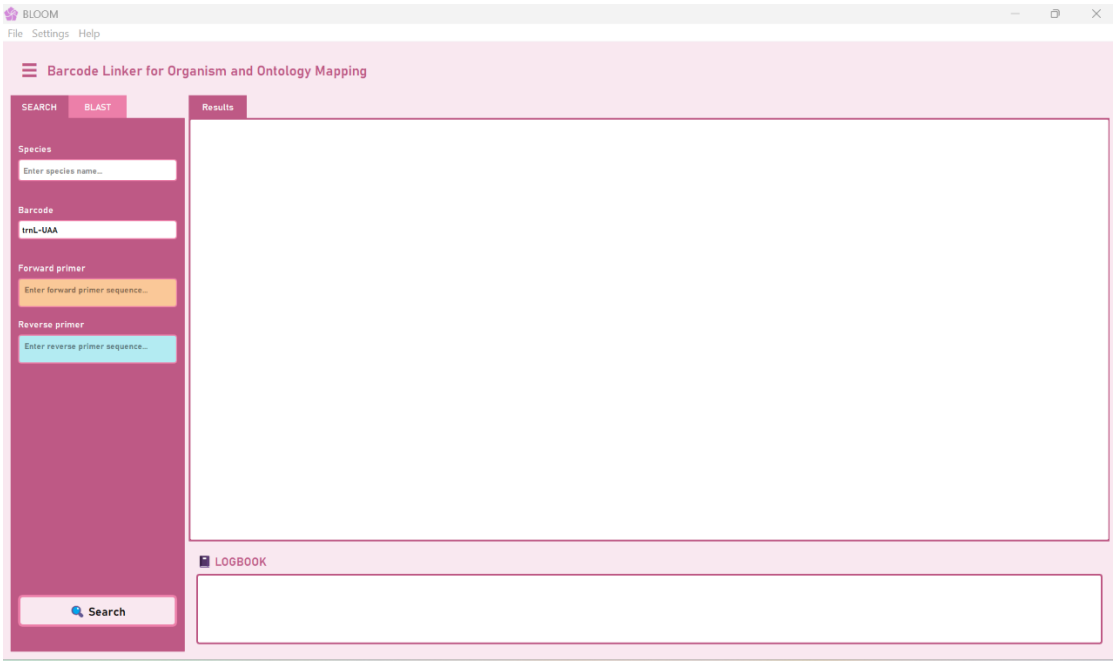

Figure 2. Graphical user interface (GUI) of the BLOOM application. *This separate pop-up window will appear once the BLOOM app has been successfully downloaded and is ready for use.*

3. Usage Instructions

3.1 Searching for Barcodes

The ‘Search’ tab on the left hand-side of the BLOOM app, requires the user to input information to search for barcodes. The user is required to input:

1. **Species name, following the binomial nomenclature.**
2. **Select the desired barcode type.**
3. **Forward primer (default option available).**
4. **Reverse primer (default option available).**

The Search tab includes an auto-complete feature for the species name, which appears as a drop-down menu. Additionally, the barcode selection is offered as a down drop menu, with trnL (UAA) and its P6 loop as the barcodes available. For primer selection, there is an auto-populate option, which will appear in the forward and reverse primer input boxes, upon barcode selection. If the user would prefer to supply their own primers, the forward and reverse primer input boxes are editable for the user to remove the default primers and input their own primers. As an example, Figure 3 demonstrates the user input for ‘Avena sativa’ (oat) species, trnL (UAA) DNA barcode and the auto-complete forward and reverse primers provided by BLOOM. After the user is satisfied with the inputs provided to BLOOM, the user can activate the search request for barcode by pressing the ‘Search’ button at the bottom of the tab. This query will be sent to the NCBI servers to search the NCBI nucleotide database to retrieve the sequences for the barcode that match the users input request. Once the search for barcodes has been activated by the user, a message will appear in the logbook informing the user their request is being processed.

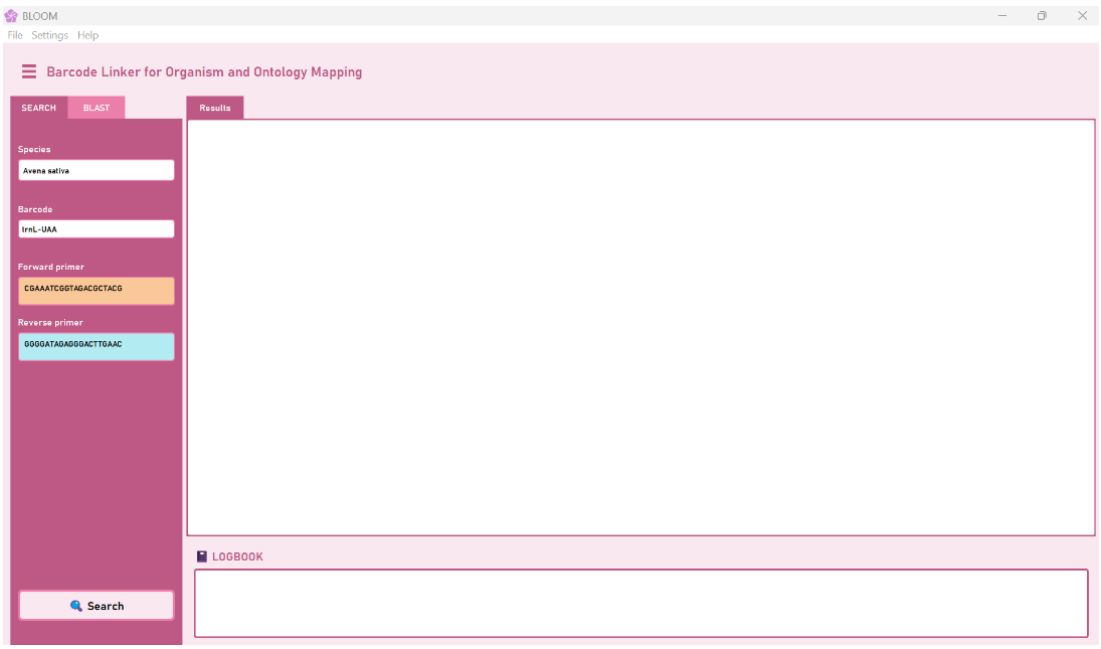

Figure 3. User input interface for barcode search in the BLOOM application. *This figure illustrates the input fields where users enter the required parameters to initiate a barcode search within BLOOM.*

3.2 Viewing Barcode Sequences

Once the request for barcodes has completed, BLOOM will return sequences which match the parameters inputted into the search tab. Sequences will be displayed as barcode cards in the barcode tab of BLOOM, shown in Figure 4. Each barcode card will contain:

1. **FASTA header information: the header for each sequence extracted from the FASTA file.**
2. **Number of duplicates: the number of duplicate sequences the BLOOM application found in the NCBI database.**
3. **Sequence length: the number of bases within each sequence.**
4. **Sequence: The sequence for the barcode found in the NCBI database.**

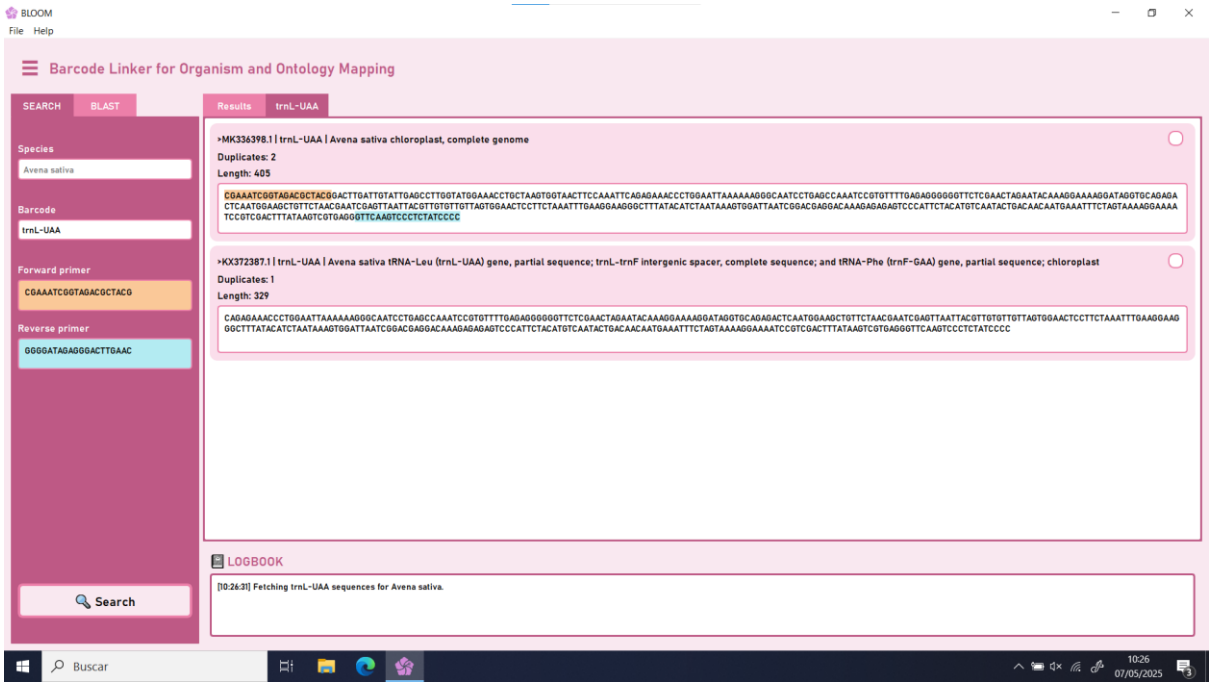

Figure 4. Barcode search results displayed in the BLOOM application. *It shows the output interface under the 'trnL-UAA' tab, where barcode results are presented as individual cards. Each card displays key information including sequence length, number of duplicates, FASTA header and nucleotide sequence. Users can hover over a sequence to highlight primer regions within the sequence. To select a barcode for BLAST analysis, the user must click the circle on the right side of the desired barcode card.*

BLOOM requires the user to select one of the sequences by clicking the circle on the far-right hand-side of the barcode card to BLAST. An additional feature BLOOM provides the user to aid the decision of the barcode sequence, is by identifying the primers within each sequence. To see the highlighted primers, the user must hover their mouse over the sequence display in each barcode card. The highlighted colour identifies the forward and reverse primer location, and primer completeness within each sequence. Moreover, there is a sequence alignment option. By clicking the circle on the far right-side of the barcode card and right clicking another barcode card to align, BLOOM will align the two sequences selected (Figure 5).

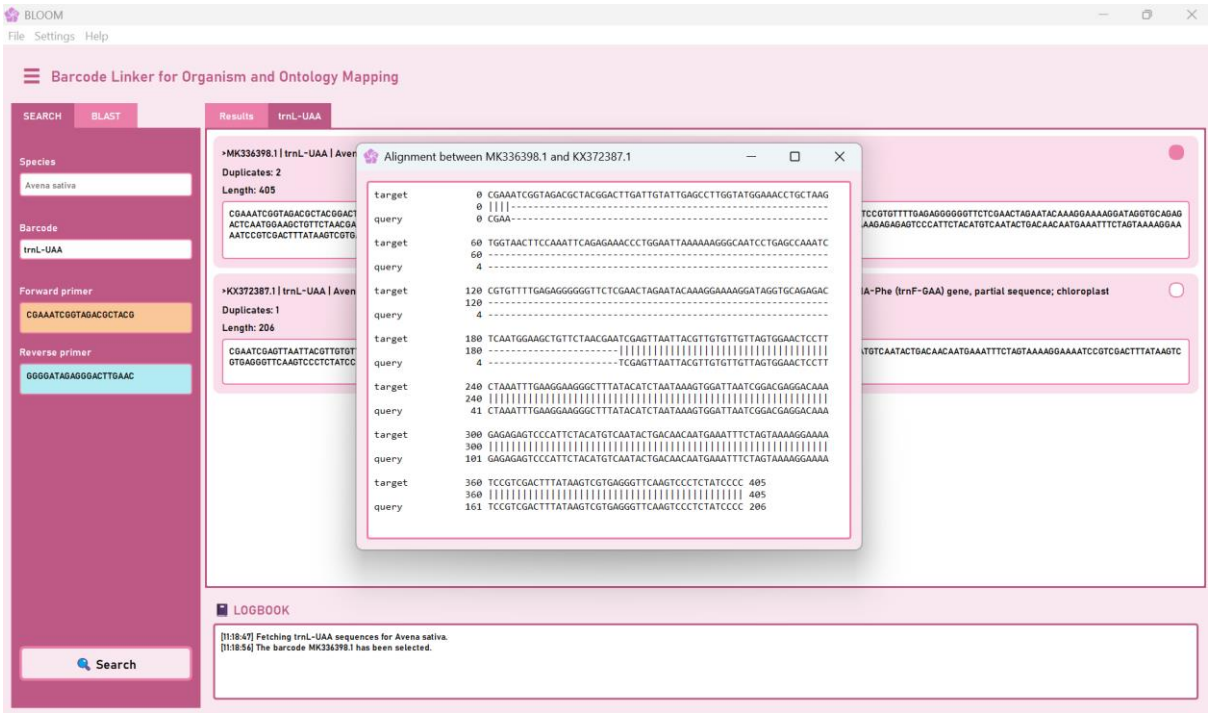

Figure 5. Sequence alignment view in the BLOOM application. *This separate pop-up window appears when a user selects one barcode card and right-clicks on another. The aligned sequences are displayed side-by-side to help the user compare them and choose the most appropriate sequence for BLAST analysis.*

3.3 Running BLAST Analysis

Once the user has selected one appropriate barcode sequence to BLAST, the user is required to navigate to the BLAST tab within BLOOM application. For BLAST to run, BLOOM requires the following information:

1. **Selected barcode card, using the circle button on the right-hand side of the barcode card.**
2. **The chosen BLAST mode.**
3. **A selected taxonomy rank.**

BLOOM provides the user with two different BLAST modes: Megablast and discontiguous blast. Additionally, the taxonomy rank ranges from species to family. Figure 6 demonstrates the field required for the user to input information before activating the BLAST request. Once the user is satisfied with the input information, the BLAST request is activated via clicking on the BLAST button in the middle of the BLAST tab, with the logbook displaying a message informing the user the request has been sent. It is important to highlight that there will be a few minutes processing time delay on generating the BLAST output, as the BLOOM application runs the query off the NCBI BLAST servers.

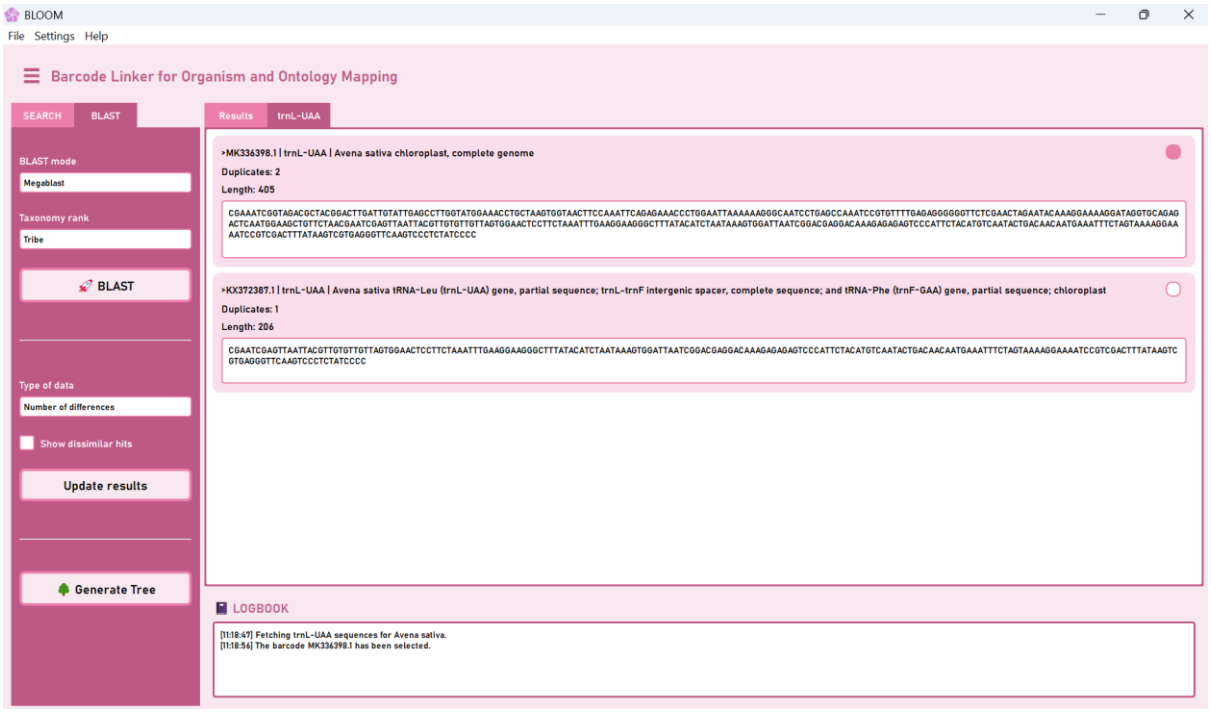

Figure 6. BLAST tab interface in the BLOOM application. *The blast tab allows the user to configure a BLAST search by selecting the BLAST mode, taxonomy rank and the barcode sequence to be queried. Once the parameters are set, the user initiates the search by clicking the BLAST button located at the centre of the tab.*

3.4 Viewing Results

After the BLAST request has completed successfully, a results tab will appear in the BLOOM application containing donut charts and bar charts as shown in Figure 7. The results tab informs the user on the best BLAST hits found for the barcode sequence of choice, the number of nucleotide differences between the barcode sequences and the retrieved BLAST sequences. Also, in the ‘types of data’ drop down menu, there is an option to display the number of species based on percentage identity. To implement the change between the number of differences of the percentage identity, the ‘Update results’ button is required to be clicked by the user. There is also an option for the user to display similar and dissimilar BLAST hits, by clicking the toggle button in the BLAST tab and clicking ‘Update results’. The option of showing dissimilar or similar hits is for the users aid in interpreting the bar chart results directly from the BLOOM application. Showing dissimilar hits will show the results of sequences that have high numbers of differences and low percentage identity, which is activated when the toggle is selected. However, the similar hits will display to the user results to sequences within a smaller range for the differences and percentage identity, making the visual interpretation easier and clearer for the user.

Figure 7. Results tab displaying BLAST hit comparisons in the BLOOM application. *This tab presents a visual summary of BLAST hits, including a donut chart showing the number of hits with and without unique species identification. Bar charts display the number of base pair differences (starting at 0) between each hit and the reference barcode sequence selected earlier. Users can customize the data shown by selecting different options from the 'Types of data' dropdown menu, such as percentage identity and number of differences. Additionally, a toggle allows users to filter for similar or dissimilar hits. To apply any changes to the display, the 'Update results' button must be clicked.*

The results tab is colour coordinated to represent the differences between either the number of different bases or the percentage identity. Table 1 identifies the different colour ranges for the number of differences or the percentage identity. The results colours will always be displayed as a linear gradient from red to green, which always passes through yellow. The number of differences does not always range from 0-9, the output is dependent on the number of differences needed to drop the identity percentage below 98%. Therefore, the number of differences bar charts can alter depending on species and barcode originally selected by the user.

Table 1. Base differences and percentage identity (%) ranges associated with result colours in the BLOOM results and node colours in the taxonomy tree. *This table defines the colour ranges used in the bar and donut charts within the BLOOM's app results tab, and the percentage identity ranges in the taxonomy tree. Colours represent sequence similarity to the reference barcode and follow a linear gradient from red to green, passing through orange and yellow to represent intermediate values. Red indicates identical or near-identical sequences (0 or 1 base difference or 100% identity), while green represents lower similarity (<98% identity and >9 base differences).*

Colour	Percentage Identity (%)
Red	100
Orange	100-99
Yellow	99-98
Green	< 98

3.5 Taxonomy Tree Feature

BLOOM application provides an additional feature to display a taxonomy tree based on the selected organism. The purpose is to display the full lineage from BLAST taxa selected, down to species level. Figure 7displays the BLAST tab within the BLOOM application, which contains the ‘Generate Tree’ button at the bottom of the tab. When the user clicks the ‘Generate Tree’ button, a taxonomy tree will be generated and displayed in a separate pop-up window. Figure 8 shows the output of the taxonomy tree for *Avena sativa* at the tribe taxonomy rank for the trnl-UAA barcode. Table 1 highlights the different colours of the species node, which corresponds to the identity percentage, printed next to the species name Figure 9).

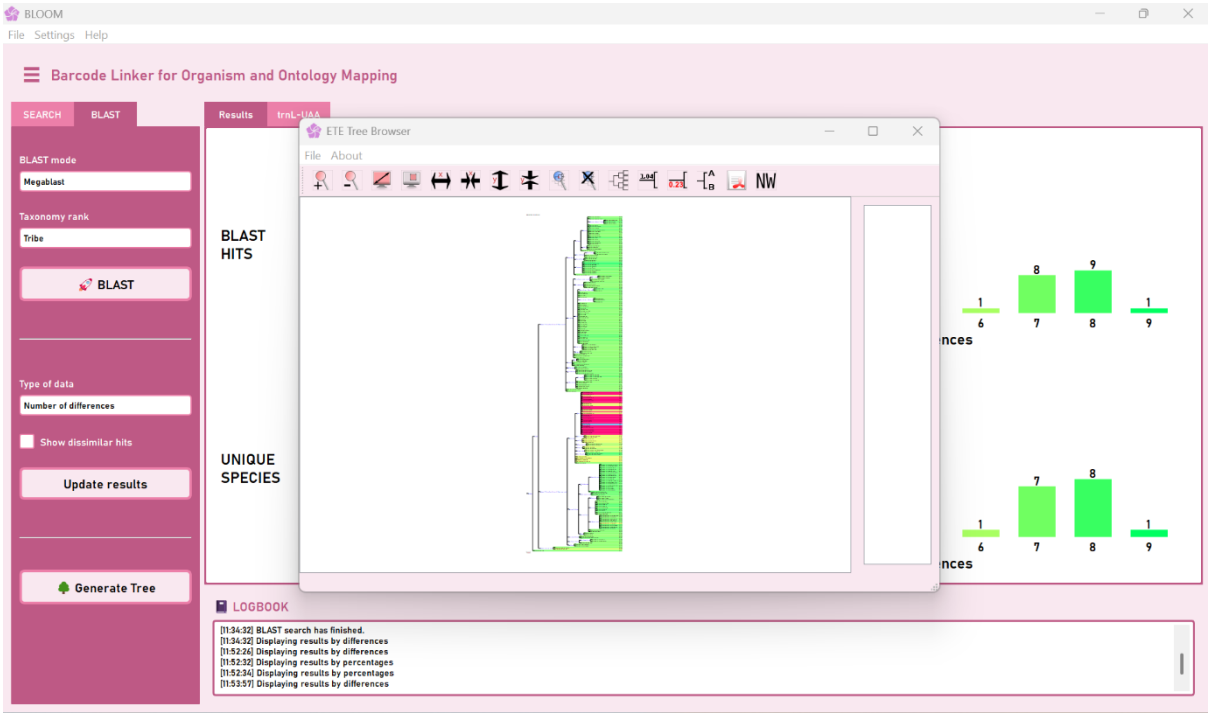

Figure 8. Taxonomy tree for Avena sativa in a separate pop-up window. *This taxonomy tree, generated based on the trnL-UAA barcode, is displayed in a separate pop-up window upon clicking the ‘Generate Tree’ button. The pop-up provides an interactive view of the taxonomic relationships specific to Avena sativa.*

Figure 9. Taxonomy tree close-up of the species nodes for Avena sativa. *This taxonomy tree displays an enlarged image of the species nodes for Avena sativa. The taxonomy tree labels nodes at every taxonomy rank. Each species node is colour-coded to correlate to the similarity of the species sequence to the reference barcode sequence for Avena sativa. Blue indicates the species (reference) barcode sequence that was searched and blasted. Red nodes indicate species sequences with an identical match to the reference barcode sequence. Red to orange node indicate percentage identity ranging between 98-100 percent. The percentage identities between 0-98 percent range from orange-yellow-green, following a linear gradient from red to green. Additionally, to the right side of each species node, the percentage identity for the sequences of each sequence compared to the reference barcode sequence is found.*

When the user clicks on the individual percentage identity in the taxonomy tree, a separate pop-up window will appear displaying sequences. The sequences displayed are all the sequences BLAST identified for each species at the node of the tree, aligned with the barcode sequence, which was selected by the user previously to BLAST (Figure 10).

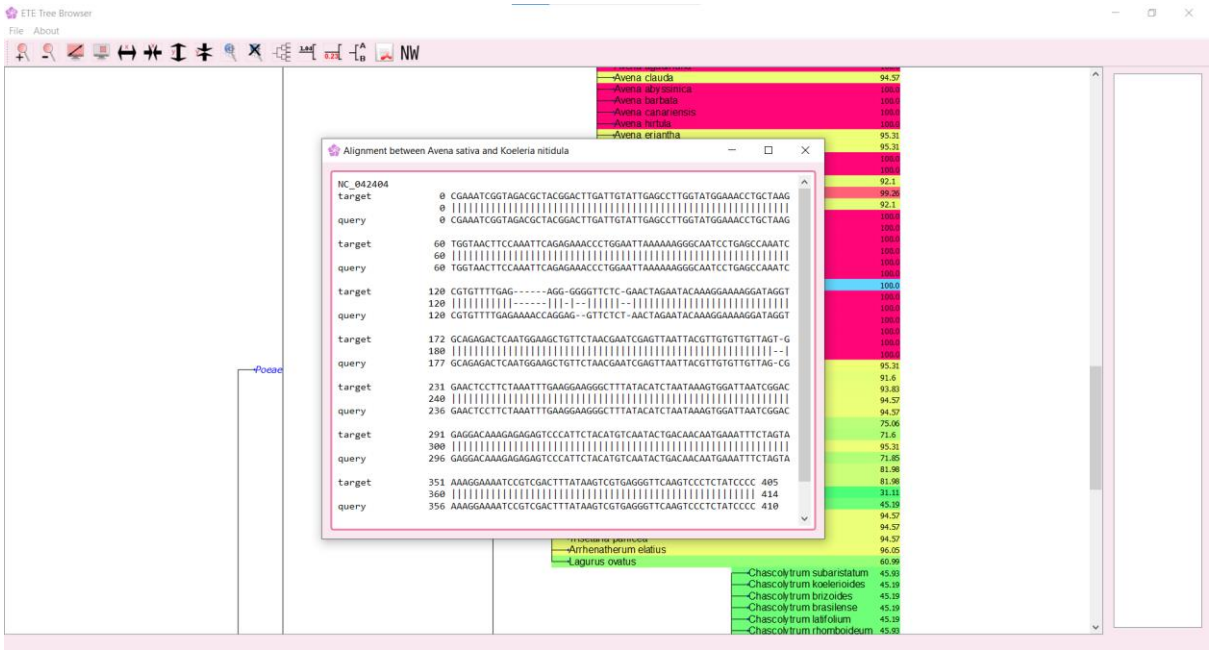

Figure 10. Sequence alignment pop-up window between Avena sativa and Koeleria nitidula. *This separate pop-up window displays the alignment between all available sequences of Koeleria nitidula and the reference barcode sequence of Avena sativa. The number of aligned sequences depends on how many Koeleria nitidula sequences are retrieved. To view an alignment, users can click on the percentage identity value displayed to the right of the species node name.*

4. Output Description

4.1 Exporting Results

To explore the results tab in further detail, the BLOOM application provides the user an option to create a CSV file, which contains informative data from BLAST. Figure 11 displays how to create a CSV file, by clicking the ‘File’ button in the top left-hand corner of the BLOOM GUI. An option for ‘Create csv file’ will be presented to the user, which will allow them to save into their local machine under the filename of their choosing. Once the user has downloaded the CSV, it can be explored further in another application, such as Excel.

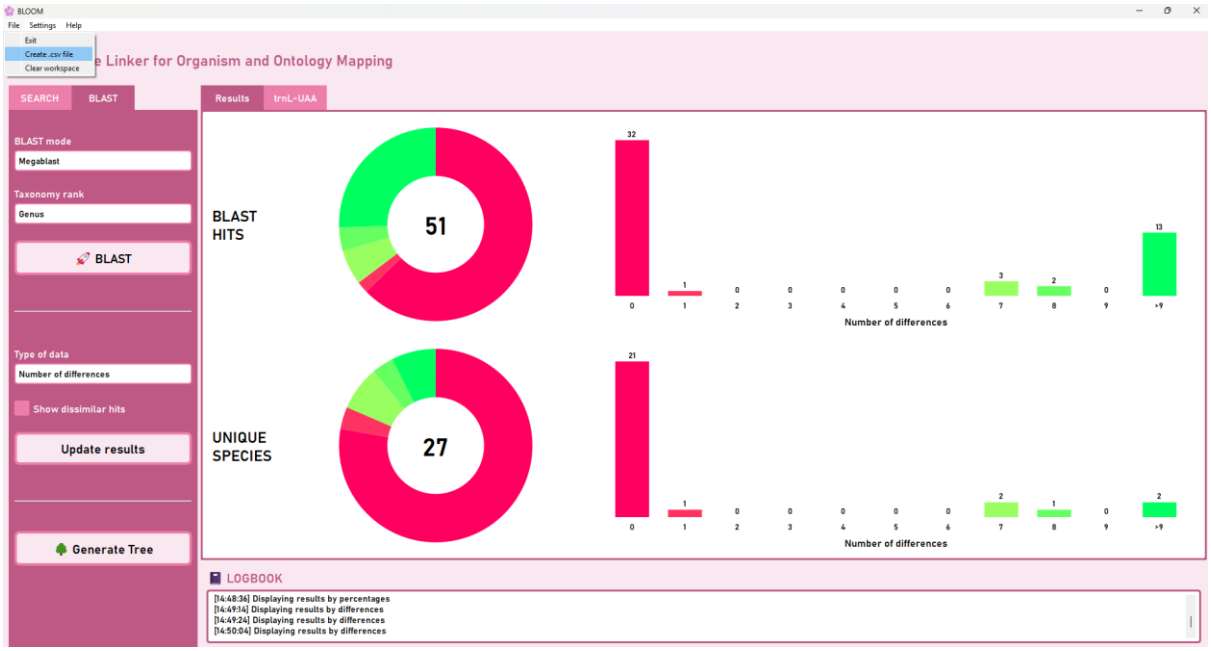

Figure 11. Exporting results as a CSV file from the BLOOM application. *This figure demonstrates how users can generate a CSV file from BLOOM. By selecting ‘File’ in the top-left menu bar of the GUI, users can click the ‘Create csv file’ option. A pop-up window then prompts the user to choose a filename and save location to their local machine.*

4.2 Managing Workspace

When the user is ready to use a different barcode, or try a new species, they can clear the workspace in BLOOM. Following Figure 12, clearing the workspace can be achieved by selecting ‘File’ in the top-left menu bar of the GUI, and clicking the ‘Clear workspace’ option. As a result, the BLOOM GUI input tabs will clear, and the application can be used again for a different species.

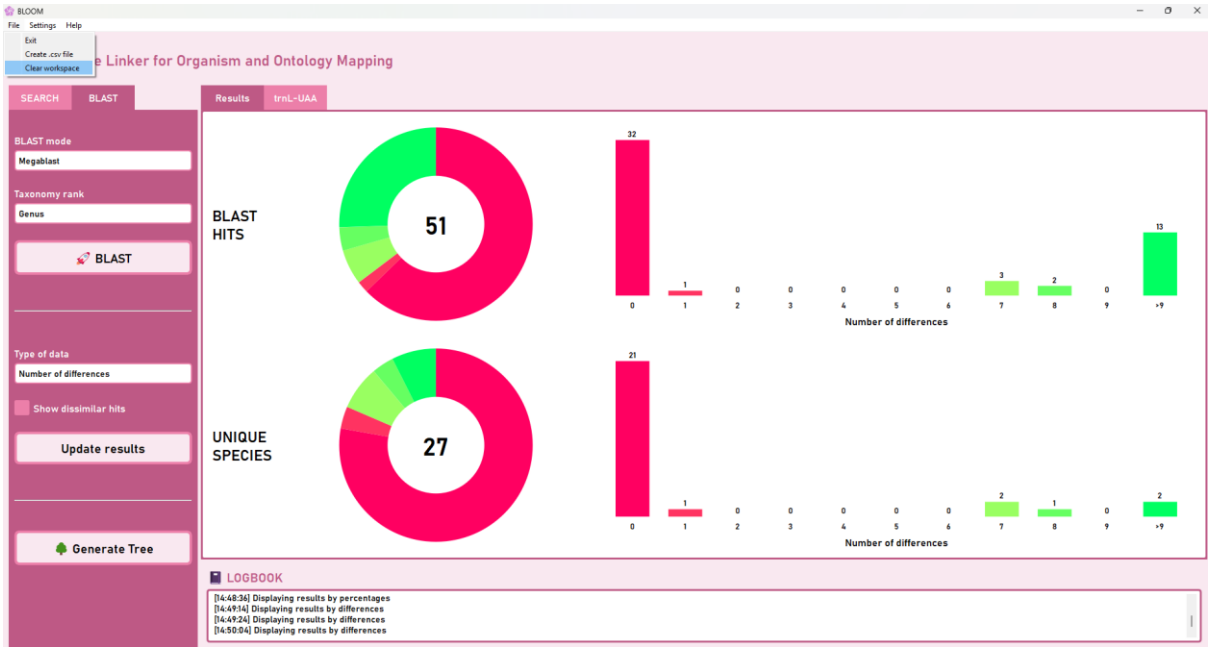

Figure 12. Clearing the workspace in the BLOOM application. *This figure illustrates how users can reset the workspace in BLOOM to begin a new query. To do so, the user must select 'File' from the top-left menu bar of the GUI, and then click 'Clear workspace' option.*

5. Troubleshooting

5.1 Downloading the BLOOM application on Google Chrome

Downloading the BLOOM application can sometimes lead to your local machine throwing a trust issue. Figure 13 identify the type of trust issue the local machine might occur, due to the application not being a commonly downloaded software. If this issue arises during downloading the application, click on the three dots on the right-hand side of the issue, and a menu will open. From this menu, if the user is comfortable with trusting the application and is ready to activate it, then ‘Keep’ needs to be selected. If any other option is selected, BLOOM application will not download to your local machine nor activate.

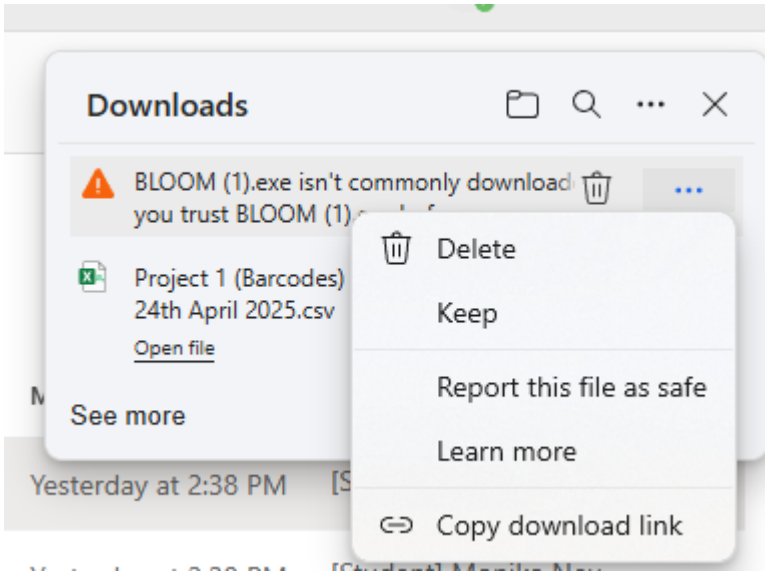

Figure 13. Trust warning when downloading the BLOOM application on a local machine. *This notification may appear during the download of the BLOOM application, as the system flags it as uncommon and potentially untrusted. To proceed with the download, the user must click the three dots beside the file and select ‘Keep’.*

After the user has selected ‘Keep’ when downloading the BLOOM application to their local machine, the user will be asked by their local machine if application is trusted (Figure 14). If the user clicks ‘Keep anyway’ the BLOOM application will successfully download to their local machine, as displayed in Figure 15. If the user selects any other option from Figure 14, the BLOOM application will not install nor be activated on the user's local machine. To activate the BLOOM application, the user must click on the BLOOM.exe file downloaded (Figure 15).

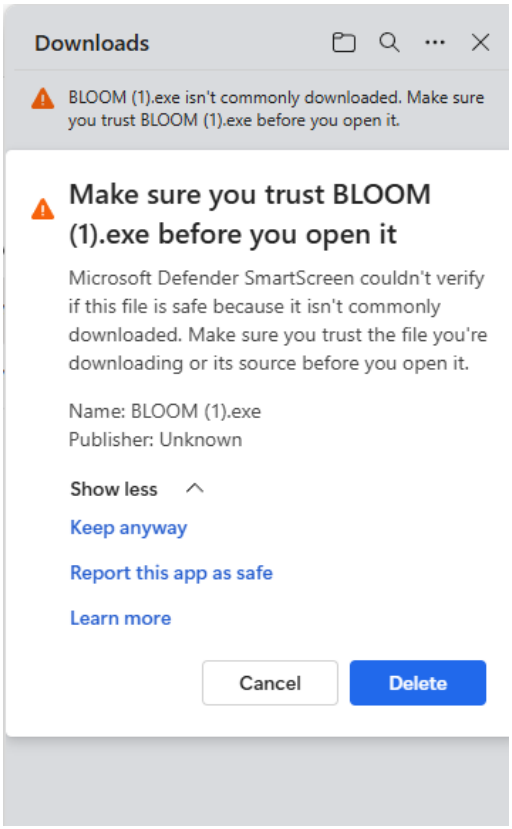

Figure 14. Final trust notification when keeping the BLOOM application file. *This notification appears when the user chooses to keep the BLOOM application despite the initial trust warning. To proceed with the download, the user must select 'Keep anyway'. Choosing alternative options will prevent the application from being downloaded or launched on the local machine.*

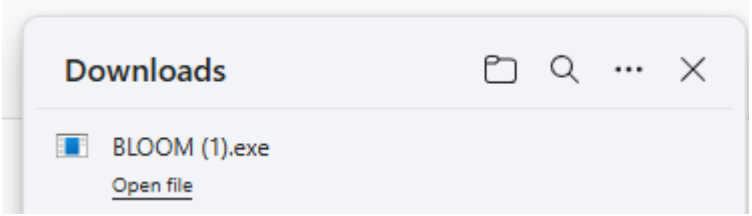

Figure 15. Successful download of the BLOOM application on the user's local machine. *This notification appears in the users download section, once successful download of the BLOOM application. To activate the BLOOM application, the user must click on the downloaded BLOOM.exe file.*

5.2 NCBI errors

Due to the BLOOM application being dependant on NCBI database, sometimes errors may occur when the database does not respond, or during any connection or internet stability issues between the database and the BLOOM application. The errors may present as shown in Figure 16 and Figure 17. In this cases it is recommended to restart the BLOOM application and try again. If the issue persists, it is recommended to retry the search and BLAST at different time, when NCBI database has been restored and working correctly.

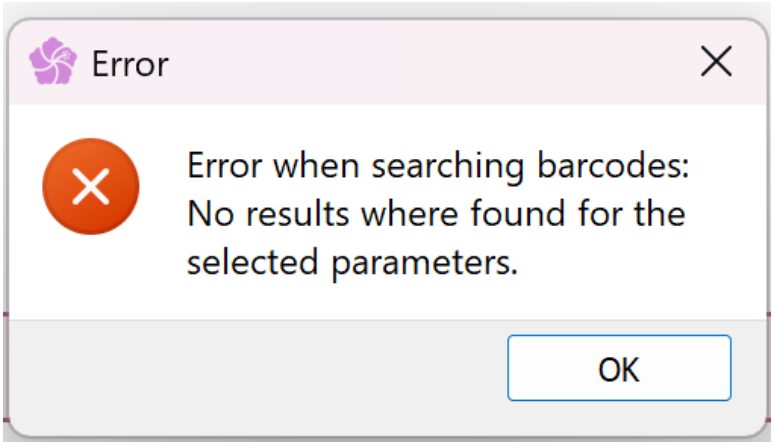

Figure 16. Error message due to NCBI database connection issue on Windows operating system.

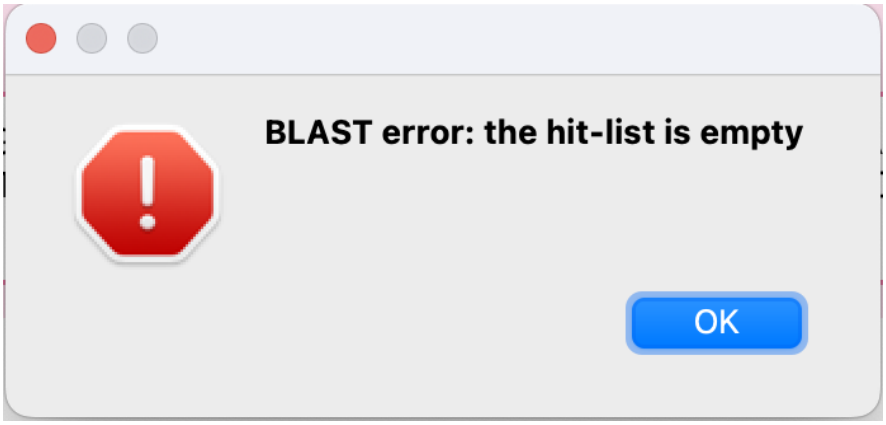

Figure 17. Error message due to NCBI database connection issue on iOS operating system.

5.3 BLAST

When the user is activating the BLAST request through clicking the BLAST button in the BLAST tab, the button must only be clicked once. If the user is clicking the BLAST button multiple times, then the BLOOM application will crash and close. To identify if the request has been sent to BLAST, the logbook in the BLOOM GUI will contain a message identifying the request has been sent to BLAST. Sometimes BLAST can take several minutes, due to the design of BLOOM running the BLAST request off the BLAST servers. If the BLOOM application crashes and closes, the user will have to re-launch the application again and start from the beginning to continue with their work.