# 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

# Abstract

DNA metabarcoding is a DNA analysis technique that combines DNA barcoding and high-throughput sequencing technologies to identify target species in complex samples. Metabarcoding is used in a variety of fields such as food safety, forensics and biodiversity conservation. This technique uses regions within DNA, known as universal barcodes, which are conserved among species from a target community but contain enough variations to differentiate at genus or species level. Selecting the correct barcode that can distinguish between the species of interest remains a challenge and can lead to delays and profit loss in wet-lab experiments if an unsuitable barcode is chosen.

BLOOM (Barcode Linker for Organism and Ontology Mapping) is a Python application that aims to help biologists select more efficient and accurate universal DNA barcodes *in-silico*. Barcode selection in BLOOM includes the plants barcodes trnL-UAA and trnL-P6. BLOOM uses the National Centre for Biotechnology Information (NCBI) database to retrieve barcode sequences for a target species. After the user selects a suitable sequence, BLOOM uses the Basic Local Alignment Search Tool (BLAST) to align the sequence with the same barcode from other species. Finally, BLOOM provides results in the form of graphs, a taxonomy tree and a detailed CSV file that evaluate the selected barcode efficiency.

The application was tested using *Coffea arabica* as the target species, given the current interest within the food safety industry. The information that BLOOM provided were further analysed to evaluate trnL-UAA and trnL-P6 efficiency for a range of taxonomy ranks and BLAST modes.

As a result, BLOOM successfully evaluated both barcodes and assessed their efficiency prior to any wet-lab experiment. Furthermore, the application successfully provided taxonomy information of each barcode at the selected taxonomy ranks. Although BLOOM has proven to be a useful tool, future improvements and additional features are considered to further expand the application’s capabilities.

# Introduction

Polymerase chain reaction (PCR) is a DNA copying based technique that has been central to species identification within a variety of fields (de Oliveira *et al*., 2022; Kang, 2019; Zambianchi *et al*., 2023). One of the applications of PCR is to amplify a short DNA region known as a barcode. This region can be used to identify the species through a method known as DNA barcoding.

DNA barcoding provides a higher accuracy and reliability when compared to other traditional methods (Uncu et al., 2017). However, the DNA barcoding approach which is largely based on traditional Sanger sequencing has some limitations . For instance, species identification is restricted to a single species within a sample, DNA barcoding cannot resolve multiple species from complex mixtures (e.g. soil, processed food and eDNA samples) (Raclariu *et al.,* 2018). These limitations require additional financial and laboratory resources. The declining cost of NGS sequencing technologies, led to the development of DNA metabarcoding. DNA metabarcoding is a technique that identifies multiple species or higher taxonomic levels from complex mixtures. By combining DNA barcoding with high-throughput sequencing, DNA metabarcoding enables simultaneous species identification (Lui *et al.,* 2019). This approach is particularly advantageous for detecting all species in complex or processed samples (Lanubile *et al.,* 2024). DNA metabarcoding has diverse applications ranging from biodiversity monitoring, ancient ecosystem reconstruction, food adulteration and diet analysis (Ruppert *et al*., 2019).

As DNA metabarcoding utilises universal DNA barcodes for identification, choosing the correct barcode region can be challenging. Not all DNA regions make for suitable barcodes, a good barcode should have conserved flanking regions for universal primer design (You *et al.,* 2009). The length of a barcode is another important consideration, as too short of a region will lack sufficient variability, and too long of a region becomes susceptible to DNA degradation (Francesco Ficetola *et al.,* 2010). This becomes problematic when analysing processed food or environmental DNA, as the DNA is typically highly degraded. Lastly, a good barcode should have sequence data availability in a large public database such as GenBank or BOLD (Barcode of Life Data System) (Ratnasingham & Hebert, 2007).

DNA barcodes used for plant taxa identification do not provide the high standard of accuracy compared to DNA barcodes belonging to other kingdoms. Plant chloroplast DNA has a highly conserved structure, due to the slow evolving DNA, which consequently leads to low resolution of plant DNA barcodes (Monterisi *et al.*, 2023). The steady pace of plant chloroplast DNA evolving, could be due to factors including the long lifespans of plants, and genetic factors: DNA repair and metabolic rate (Yoshiyama *et al.,* 2013).

The plant DNA barcode, trnL-UAA intron (300-500 bp) belongs to the chloroplast DNA coding for the transfer RNA-leucine (trnL) gene (Sugita *et al.*, 1995). The UAA is the anticodon in the tRNA, which will pair with the UUA codon in mRNA, specifying the amino acid leucine (Sugita *et al.*, 1995). The trnL-UAA intron is also found in most cyanobacteria and the chloroplast of algae as well as plants (Kaasalainen *et al.,* 2015) . As trnL-UAA encodes a tRNA, it will fold into a cloverleaf structure with the acceptor stem binding to the leucine amino acid, and the anticodon loop is the location of the UAA. The P6 loop is a shorter fragment of trnL-UAA, which is located in the region of the group I intron, which is a self-splicing RNA sequence. Group I introns can fold into complex secondary structures, with the P6 loop being one of the cloverleaf structures. Importantly, the P6 loop's main function is in

proper folding of the intron RNA, for the self-splicing properties, and due to its variability, it is being used as a phylogenetic marker (Olsson *et al.,* 2012).

Specifically, plant DNA barcodes such as the chloroplast intron trnL-UAA and its shorter fragment trnL-P6 provide biologist a popular and reliable DNA barcode for plant exploration Taberlet *et al.*, 2007). The trnL-UAA intron provides a conserved secondary structure and PCR amplification durability (Taberlet *et al.,* 2007). These properties of the DNA barcode provide biologists ease for tracking sequence changes, primer design, and generate consistent and reproducible results (Monterisi, S. *et al.*,2023). Above all these advantages, the underlining issue of poor resolution remains a consistent issue. There is a variety of different plant DNA barcodes available, and errors can arise from inadequate selection. Inconclusive results can be presented as the same DNA barcode can match multiple species making it difficult to distinguish between them. This is particularly problematic in taxa where closely related species share highly conserved regions, resulting in overlapping or indistinguishable barcode signatures (Burgess and Freeling, 2014).

A major complication is that researchers cannot always predict these limitations in advance. Although a select number of bioinformatic tools have been developed to tackle this issue, there remains no unified approach to determine whether a particular barcode will successfully differentiate between species (Geneious Prime, n.d.; Agostinetto et al., 2022; Boyer et al., 2016). Often, scientists must carry out laboratory experiments, such as PCR amplification, sequencing, and comparative analysis, before realising that the selected barcode region lacks the resolution required. This trial-and-error process consumes valuable time and resources and may lead to delays or even setbacks in studies involving food safety, biodiversity assessment, or ecological monitoring.

Food safety is an active research area where DNA barcoding and metabarcoding aid research in species identification. Food adulteration is a pressing issue within the agrifood sector, “the World Health Organization estimates unsafe food causes a loss of 33 million healthy life years annually” (Das *et al*., 2025). For example, *Coffea arabica*, which is considered as the luxury *Coffea* species, is commonly adulterated with *Coffea robusta* (Combes *et al.,* 2018). Detecting food adulteration can be challenging, often due to the expensive requirements and technical expertise needed to conduct the tests. DNA based techniques offer an advantage over other methods, characterizations often need to go beyond species identification, usually requiring the identification of a particular variety or cultivar (Das *et al*., 2025). DNA-based methods are unique in that, in principle, they can provide this level of granularity. DNA based methods are the only methods that can detect genetically modified organisms (GMOs) which fall under strict regulations in many regions. To address these challenges, the Barcode Linker for Organism and Ontology Mapping (BLOOM) application, offers a computational solution that supports researchers in selecting the most appropriate barcode for their specific study. BLOOM aligns user-selected sequences, by accessing and analysing data from the National Centre for Biotechnology Information (NCBI) (Sayers *et al.,* 2025). The tool uses BLAST to identify similar entries providing similarity percentages across taxonomy levels (Altschul *et al*., 1990) BLOOM aggregates this information, to detect species which cannot be resolved with the user-selected barcode. BLOOM not only reduces the likelihood of selecting ineffective barcodes but also minimizes the costs and labour associated with trial-and-error testing in the lab.

# Methods

## Data Sources

Research on DNA metabarcoding was conducted to gather information from freely accessible platforms, such as PubMed and Google Scholar. This consolidated the foundations of the design of BLOOM, with the aim to provide biologists an insight into the most accurate DNA barcode for species identification. BLOOM accesses the National Center for Biotechnology Information (NCBI), a publicly available database, to retrieve sequence data and taxonomical data to determine the accuracy of the barcode sequence for taxa identification (National Center for Biotechnology information (n.d)). NCBI's taxonomical database was accessed to fetch the taxonomical ID, based on the user input of species name (Federhen, 2015). Additionally, the taxonomical database was acquired to retrieve the lineage to build the taxonomy tree, based on matching the taxonomical ID supplied. The second NCBI database BLOOM accesses to get data is the nucleotide database, which selects sequences based on taxonomical ID and barcode for BLAST analysis (Sayers *et al.*2022). . Lastly, the BLAST core nucleotide database was utilised to match sequences to the original barcode sequence selected by the user, based on barcode taxonomical rank name (Altschul *et al.*,1990). BLOOM provides the user an auto-complete option for default forward and reverse primers for each barcode. This auto-completion for primers is hard-coded in BLOOM.

## Software & Environment

BLOOM was developed using Python programming language (Version 3.9.13) and was designed and created to operate on a Windows PC local machine (Python Software Foundation., 2024). The scripts are compiled into a single executable file, which can be activated through a double click, created using PyInstaller (version 2.46.0) (PyInstaller Development Team., 2025). Table 1 contains the file formats the BLOOM application utilises for graphical user interface (GUI) display, extra features, and creating a readable file for the user from the BLAST results.

## Pipeline Breakdown

The BLOOM application pipeline is divided into barcoding fetching, barcode comparison and analysis, BLAST search, results visualisation, tree generation and CSV generation. Figure 1 displays the flow chart for the pipeline of the BLOOM application. Table 2 contains the relevant package names, version numbers, parameter modifications, and citation for all the packages used to build the BLOOM application.

The first stage within the BLOOM application is barcode fetching. The outcome of this stage is to retrieve the barcode sequences from the NCBI nucleotide database and display them to the user within the barcode tab of the GUI. Modules from the Biopython library are utilised to access NCBI, align sequences and parses information to BLOOM. The Biopython Entrez module was used to access the NCBI nucleotide database. The Entrez module gets the barcode sequences based on taxonomical ID and barcode. For the data acquired from the NCBI nucleotide database to be harnessed to BLOOM, the Biopython SeqIO module parses the information Entrez sought from the NCBI web app to the BLOOM app. BLOOM uses the Biopython Align module to create a pairwise align object, which identifies the primers within the different barcode sequences the Entrez module previously acquired. The data needs to be converted from a string of characters to a Seq object, which is the preferred format for bioinformatics analysis. The conversion utilises the Biopython Seq module,

which results in BLOOM calculating the reverse compliment of the barcode sequence for the reverse primer.

Once BLOOM has acquired the appropriate barcode sequences, they are presented to the user as 'barcode cards', for user selection. The barcode cards contain NCBI information including: FASTA header name for the sequence, the number of duplicates per sequence, sequence length and the sequence. BLOOM highlights the primers within each barcode sequences in the barcode cards, using the PyQt5 QTextCharFormat and QTextCursor modules. The QTextCharFormat module manipulates the text style, by highlighting the primers within the barcode sequence, and the QTextCursor module tracks the cursor events to display the highlighted primers upon cursor hovering. The Biopython Align module visually aligns two selected barcode sequences in a separate pop-up window. The Align module aligns sequences based on nucleotide location and not based on primer matching.

BLAST takes place once the user has selected the appropriate barcode card, BLAST mode and taxonomy rank. The BLAST function utilises the Biopython Blast module, specifically the NCBIWWW module to access the BLAST web services, and the NCBIXML module for parsing the NCBI BLAST results to BLOOM. The BLAST search runs on the NCBI BLAST servers, to compare the selected barcode sequence with other species within the BLAST core nucleotide database. The RegEx package (version 2.2.1) is used for the BLAST search to find the scientific name within the BLAST description of each hit.

Once the BLAST search is complete, the results are filtered to retrieve the unique species data. The unique species data will not contain repeated species names. The results are displayed in the results tab in the BLOOM application, after calculating the percentage identity and number of differences statistics. For the BLAST hits and the unique species results, BLOOM displays a donut chart for the total number of species and a bar chart for count frequencies. The number of differences will display the nucleotide number of base differences when compared to the original barcode sequence. This is calculated by filtering the number of hits with the distinct number of differences, culminated to a histogram. The percentage identity displays the percentage of similarity between each BLAST species sequence and the original barcode sequence. This percentage identity is calculated based on the individual NCBI BLAST score divided by the sequence length. Similar results can be viewed, which are all the hits above and including 98%, whereas dissimilar hits display the hits below 98% as well as above.

An extra feature of BLOOM creates a taxonomy tree to visualise the taxonomical relationship between species barcodes BLAST identities. The ETE Toolkit NCBITaxa module was utilised to connect to the NCBI taxonomy database to retrieve the taxonomical lineage per taxonomical ID from the BLAST results. A taxonomy tree is built from the list of species names with unique BLAST hits, and the taxonomical lineage for each taxon. The ETE Toolkit Nodestyle, Faces and Textface modules was utilised to annotate each node with the appropriate species name, delete any empty nodes, and create node labels for the tree containing species name. The taxonomy tree was styled with the ETE Toolkit TreeStyle module, to display interactive faces and coloured nodes, which are dependent on the percentage identity per species, when the tree is rendered. The interactive faces, which contain the percentage identity per species will display in a separate pop-up window with the alignment of all the sequences for the BLAST species identified, with the original barcode sequence. The Biopython Align module will align the sequences in the pop-up window, which is displayed when the user clicks on the interactive percentage identity face. The taxonomy tree produced does not have distance metrics as the branches of the tree was built based on NCBI taxonomical lineage per taxonomical ID.

BLOOM creates a CSV file, which contains all BLAST results for the user to inspect further and perform downstream analysis. The CSV (version 1.0) package was utilised to read and write a CSV file from the parsed XML file containing the BLAST results.

***Table 1. File formats utilised by the BLOOM application.*** *This table contains the file formats the BLOOM application utilises to generate results to the user. Each file format has a description to explain the use of the file in BLOOM. The CSV file is created in the BLOOM application upon user request and the XML file is parsed from BLAST. Both CSV and XML file formats contain the data generated from the BLAST request.*

| File Format | Functionality |
|---|---|
| TXT | Contains the binomial names for the plant species found in NCBI consortium. Supports the auto-complete feature of the species name search tool in BLOOM. |
| ICO | Contains the image of the logo for the BLOOM application. |
| QSS | Contains the details for the style of the graphical user interface (GUI). |
| YAML | Contains the configuration parameters and stored values for the whole applications. |
| CSV | File created to store the BLAST results. |
| XML | File that is parsed containing the BLAST results. |

## Graphical User Interface (GUI)

The display and feature function of the GUI within the BLOOM application, utilises the PyQt5 package. The PyQt5 QtWidgets module was used to establish the widgets of the GUI and any graphical element that is stored in each of the widgets. For classes that are not graphical elements, such as regular expressions and event loops, the PyQt5 QtCore module was used. However, for the graphical classes such as event handling and basic imaging, the PyQt5 QtGui module was applied. The fundamental design of the GUI is for navigation ease and sufficient visual representation of the outcomes for the user. The BLOOM GUI contains a logbook at the bottom of the interface, which utilises the Datetime package, to display a date and time in the logbook for each action performed.

## Statistical Analysis

Downstream analysis utilising the CSV file created by BLOOM, was performed to determine whether changing the BLAST modes (MegaBLAST and Discontiguous BLAST) and taxonomy levels (Family, Subfamily, Tribe and Genus) impacted species discrimination for trnL-P6 and trnL-UAA. Bar charts were created to display the unique species results for BLAST modes and taxonomy levels using R and RStudio (version 4.4.2) (R Core Team., 2024). Additionally, three *Coffea* species were compared to *Coffea arabica* and *Avena sativa* compared with *Avena barbata* were further investigated to identify number the BLAST hits for the same species, as sometimes more than one hit can return per species. Each BLAST hit per species were displayed as a table and bar chart based on their identity percentage, to evaluate the overall identity percentage for each species.

## Accessibility and Reproducibility

The BLOOM application can be accessed through the following GitHub repository link: https://github.com/ms2206/BLOOM.git.

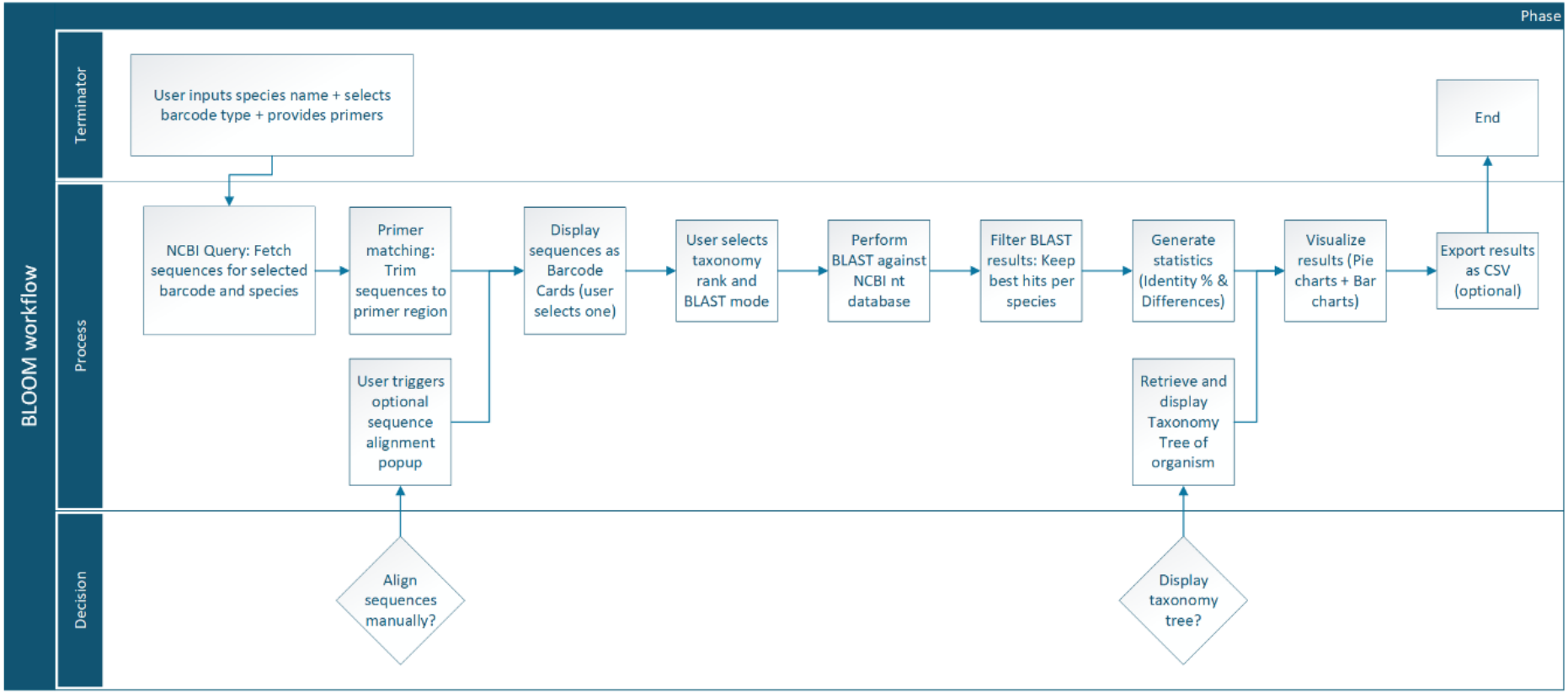



***Figure 1. Pipeline overview of the BLOOM tool.*** *This flowchart illustrates the complete data processing pipeline implemented in BLOOM, a tool designed to compare the accuracy of DNA barcodes to facilitate species identification. The process begins with the user input, including the species name and selected DNA barcode region. BLOOM constructs an NCBI query based on these inputs and retrieves the corresponding sequences. Primer alignment and trimming and then performed to isolate the relevant barcode region. The user selects a representative barcode sequence, which is subsequently used to perform a BLAST search. BLOOM calculates sequence similarity metrics, including percentage identity and the number of base differences. These results are visualised using bar charts and pie charts and can be exported as a CSV file for further analysis outside of BLOOM. Users can generate a phylogenetic tree based on taxonomy to explore the relationships between closely related species. Each stage is modular and designed to be customisable for specific use cases.*

***Table 2. Python packages, classes, versions, modification and references used in the development of the BLOOM application.*** *This table summarises the Python packages utilised in building BLOOM, including specific classes (where applicable), version numbers, custom modifications or parameters applied, and relevant references. Each package contributed to a particular component of the application pipeline and was configured to support the generation of the final output.*

| Library Name | Package | Version number | Modifications/Parameters | Reference |
|---|---|---|---|---|
| Biopython | Align | 1.85 | The Align package is used to create two instances of the Align.PairwiseAligner class. The first one is used to detect primers inside nucleotide sequences. It has lower scores for mismatches and higher scores for gaps than the default ones. The second one is used to align two sequences between each other. This instance prioritises gaps over mismatches. | Cock *et al.,* 2009 |
| Biopython | Sequence | 1.85 | No modifications. | Cock *et al.,* 2009 |
| Biopython | Entrez | 1.85 | This package contains the functions Entrez.efetch and Entrez.esearch to retrieve sequencing data from NCB in text format. All uses of these functions employ a dummy email and an empty NCBI API key. | Cock *et al.,* 2009 |
| Biopython | Blast | 1.85 | Searches performed with the blast function target the “nt” database, use the “blastn” program and retrieve a maximum of 20000 hits to avoid not retrieving all hits since the default number is 5000. | Cock *et al.,* 2009 |
| ETE Toolkit | - | 3.1.3 | Since the ETE Toolkit interactive tree is based on PyQt the window style matches that of the whole app. | Huerta-Cepas *et al.,* 2016 |
| PyQt5 | - | 5.15.11 | To allow a more flexible use of PyQt5 widgets, different custom classes are created to inheriting regular PyQt5 classes properties. | Riverbank Computing (n.d.) |

# Results

## Performance of the trnL-UAA Barcode in Species-Level Discrimination

This section presents the types of results that can be obtained using the BLOOM application. To illustrate its capabilities, species examples have been selected related from the food authenticity industry, such as *Coffea arabica, Coffea liberica, Coffea canephora, Coffea stenophylla* and *Avena barbata*.

The BLOOM application was used to assess the trnL-UAA barcode region for *Coffea arabica* and its ability to differentiate between other species. *Coffea arabica* and *Coffea robusta* are both used for coffee production, but the former is considered of superior quality and demands higher prices. Differentiating them by DNA metabarcoding could be beneficial for certifying purity in the food adulteration industry. Using the pipeline, a phylogenetic tree and CSV file have been generated. From the CSV file a doughnut chart visualising species identity distribution, and a histogram of the differences in nucleotide will be generated. Each chart provides a valuable insight in how well the trnL-UAA marker can distinguish species and how sequence identity relates to their taxonomic grouping, allowing user to understand species-level differences.

The phylogenetic tree presented in Figure 2 ,Figure 2, displays all unique species identified by the BLAST search at the chosen taxonomy level (family) and matched to the NCBI database, for *Coffea arabica*. Figure 2a displays the overview of the entire tree, while Figure 2b magnifies a fragment of the tree, focusing on the *Coffea* genus. In Figure 2, the taxonomy tree is color-coded based on its percentage identity for each species BLAST returned. The colours represent the initial species of interest (colour-coded in blue), and the other coloured nodes correspond to the species BLAST identified as similar, which have a continuous colour gradient ranging from red (highest identity) through yellow (lower identity), to green (the lowest identity). This enables a useful and quick visual identification of species with high similarity to *Coffea arabica*. The tree shows 16 species with 100% sequence identity to *Coffea arabica*. The lowest percentage identity is 72.98% for *Coffea lebruniana*.

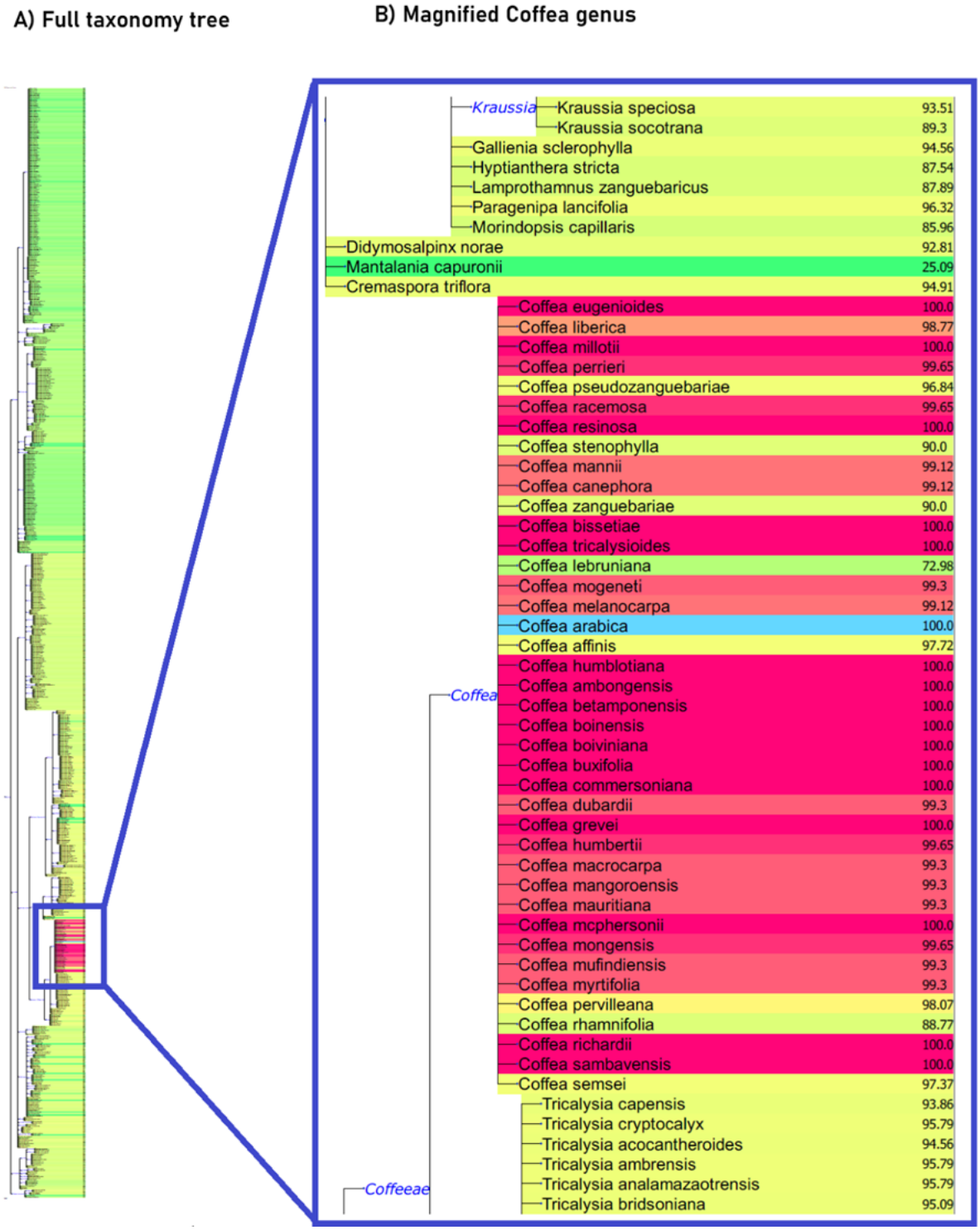



***Figure 2 Phylogenetic relationships of species identified via BLAST using the trnL-UAA barcode.*** *(A) An overview of the full taxonomy tree displaying all unique species matched to Coffea arabica, with branches color-coded by sequence identity (red = highest, yellow = lower, green = lowest). (B) Magnified view of the Coffea genus, highlighting 40 unique species with their corresponding identity percentages. Most Coffea species, exhibit high sequence similarity (≥97%), while more distantly related genera such show lower identity values. This visualization illustrates the barcode's resolution for species-level discrimination within the Coffea lineage and its relatives.*

The donut chart in Figure 3a, summarizes the distribution of percent identity values among all unique species from all BLAST results. The doughnut chart displays a total of 42 unique species were identified for *Coffea arabica* using the trnL-UAA barcode. The colour-coding of the doughnut chart correspond to the identity percentage for each species within the BLAST results. The chart reveals a wide range of identity scores, with a significant portion of species showing very high sequence similarity (large red and pink segments), and progressively fewer species falling into lower similarity ranges (indicated by smaller orange, yellow, and green segments). This visual summary allows for a quick assessment of how closely related the BLAST hits are to the target species, even in the absence of explicit numerical thresholds.

Figure 3b contains a bar chart displaying the number of nucleotide differences between *Coffea arabica* and all unique species from BLAST hit results. Each bar corresponds to a specific number of differences, and the height represents the number of unique species per category. As seen in this example, 18 species have an identical barcode sequence to the barcode sequence of *Coffea arabica* (indicated by the 0 bar), while 5 species had 1 nucleotide difference, and 10 species differed by 2 nucleotides. There has been one unique species found per category with exactly 3 and 7 nucleotide differences. There have been 3 unique species identified per category, which have a 10, 11 or 12 nucleotide base difference. Any species with more than 12 differences were excluded from this graph.

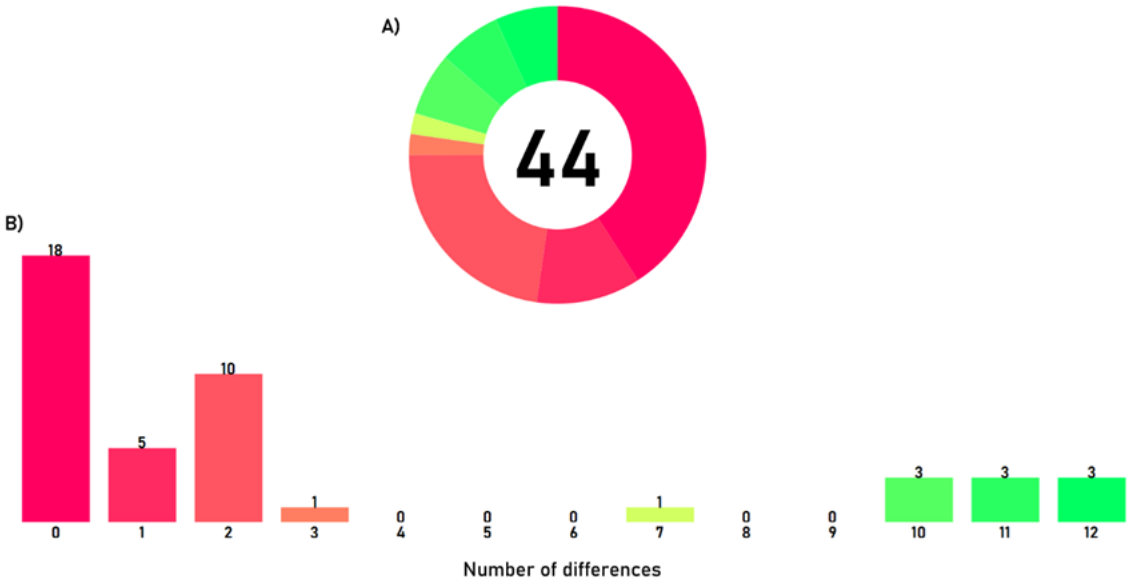



***Figure 3. Sequence identity and nucleotide difference distribution among 44 unique species matched to Coffea arabica using the trnL-UAA barcode****. The doughnut chart (A) displays the relative proportions of species grouped by sequence identity, with colour ranging from red (high identity), through yellow (lower identity), to green (lowest identity). The bar chart (B) below shows the number of species in each category based on nucleotide differences. Most species show 0–2 differences, while a few display 10–12, highlighting the varying levels of sequence divergence across matches. Species of nucleotide differences of more than 12 were excluded from this histogram. These outputs show how BLOOM transforms raw BLAST data into clear visual summaries, enabling rapid interpretation of sequence identity, species-level differences, and barcode identification accuracy*

## BLAST modes

Utilising the CSV file generated from BLOOM for *Coffea arabica* at the family taxonomy level, the relationship between BLAST modes for trnL-UAA and trnL-P6 were analysed to investigate whether

the number of species identified in BLOOM was impacted. Figure 4 displays bar charts for the different BLAST modes: MegaBLAST (blue) and Discontiguous BLAST (orange).

The function of MegaBLAST is for searching BLAST with a large genomic entry as it compares closely related sequences, whereas Discontiguous BLAST compares sequences which are less similar as it can ignore some bases, which can introduce mismatches (Altschul *et al.*, 1990). The X—axis in Figure 4 contains the identity percentage threshold (<98%, 98-100% and 100%), which is the NCBI score divided by the sequence length. Therefore, it can be concluded that a high identity percentage represents species that are more similar to the original barcode sequence. The Y-axis identifies the number of species identified with each identity percentage threshold.

Figure 4 shows the trnL-UAA barcode is not impacted by the different BLAST mode options selected, as the number of species per identity percentage threshold remains consistent. On the other hand, trnL-P6 is affected by the different BLAST modes, but only at low identity percentages of <98% (MegaBLAST: 872 and Discontiguous BLAST: 334). The identity percentage threshold for 100% and 98-100% is unaffected by the different BLAST modes, which can be argued as the most important identity percentage threshold to discriminate the identical species sequences from the similar species sequences.

Another observation Figure 4 shows is no identified species with an identity percentage between 98-100% for trnL-P6, but trnL-UAA does contain species at this identity percentage threshold.

## Taxonomy Rank

The BLOOM application offers users another parameter to change, which is the taxonomy ranks. CSV files were generated from BLOOM for *Coffea arabica* for trnL-P6 and trnL-UAA barcodes for different taxonomy ranks available within BLOOM. Figure 5 uses the taxonomy rank name for the X-axis and the number of species identified within each taxonomy rank is measured on the Y-axis. The bars are also colour-coded, representing the different identity percentage threshold within each taxonomy rank. The bar chart is divided into trnL-P6 (left) results and the trnL-UAA (right) results.

For both trnL-P6 and trnL-UAA, the identity percentage of <98% differs in the number of species discriminated throughout the different taxonomy ranks. For both barcodes, the number of species within the 98-100% and the 100% identity percentage threshold remain consistent. This identified the resolution power of species discrimination for trnL-P6 and trnL-UAA is not impacted by the different taxonomy ranks. As the taxonomy ranks go from genus to family, the number of species within each rank is expected to increase respectively.

Overall, trnL-P6 identifies more identical species to *Coffea arabica* compared to trnL-UAA. Figure 5 shows a similar pattern for trnL-P6, which the 98-100% threshold does not identify species with this identity percentage.

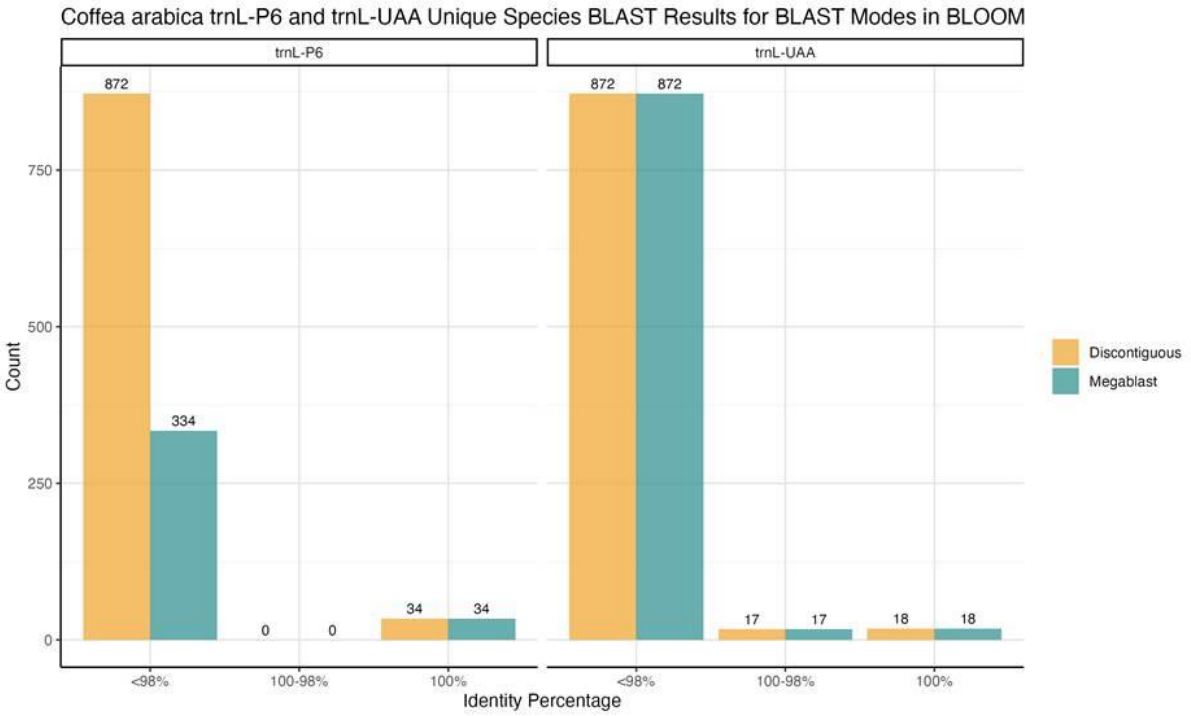



***Figure 4. Bar plots showing the number of species identified at different identity percentage thresholds using two BLAST modes (MegaBLAST and Discontiguous BLAST) for Coffea arabica trnL-P6 (left) and trnL-UAA (right) markers.*** *Identity threshold are grouped into three categories: <98%, 98-100% and 100%. Each threshold includes a bar for MegaBLAST and an orange bar for Discontiguous BLAST. The number of species identified using trnL-UAA remains consistent across BLAST modes, indicating no effect of BLAST mode. For trnL-P6 the number of species identified is also consistent between BLAST modes at 98-100% and 100% thresholds but differ at the <98% threshold.*

The total number of unique species identified for *Coffea arabica* for trnL-P6 and trnL-UAA using the BLOOM application are displayed in Table 3. Evidently, there is no difference on species discrimination between the DNA barcodes at Genus and Tribe taxonomy level. However, as the taxonomy rank expands to Subfamily and Family, trnL-UAA identifies more species compared to trnL-P6.

***Table 3 Total number of unique species* identified *per taxonomy level for* Coffea arabica using *trnL-UAA and trnL-P6* barcodes via *the BLOOM application***. The table below contains the total number of unique species based on the DNA barcode selected and the taxonomy rank for Coffea arabica. The number of species discriminated for at Tribe and Genus level are consistent with both barcodes. However, at Subfamily and Family taxonomy level, trnL-UAA identifies more species compared to trnL-P6.

| Taxonomy rank | trnL-UAA | trnL-P6 |
|---|---|---|
| Genus | 42 | 42 |
| Tribe | 73 | 73 |
| Subfamily | 504 | 318 |
| Family | 907 | 368 |

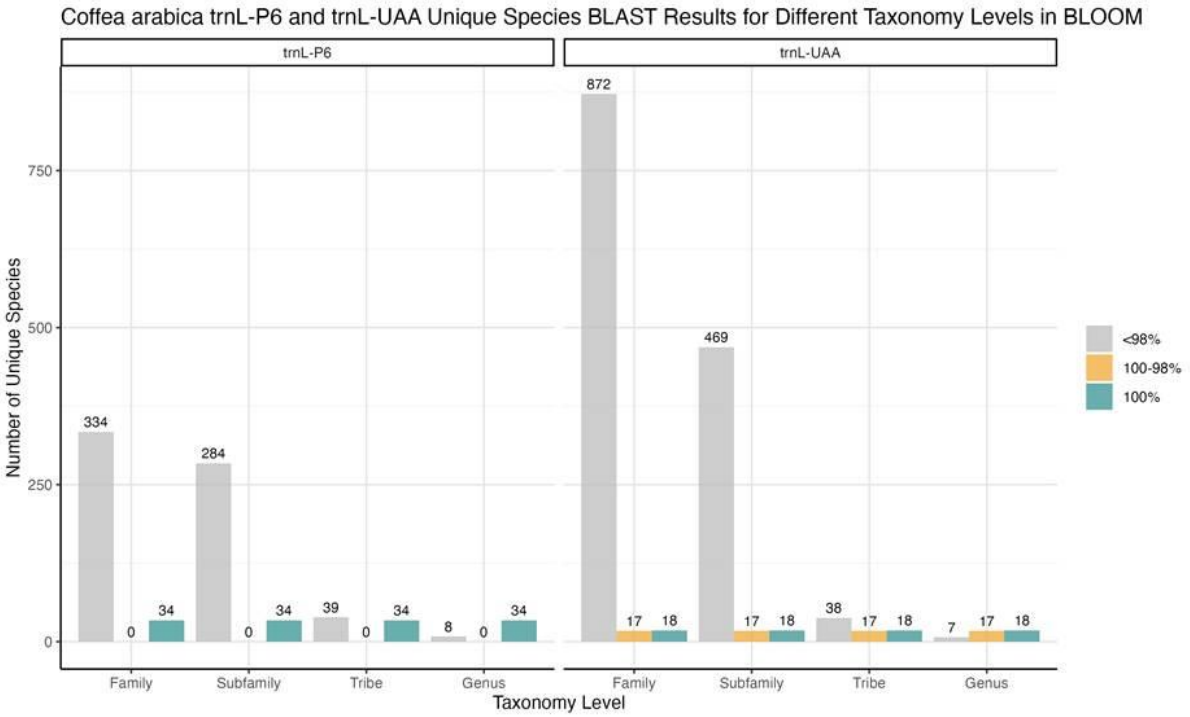



***Figure 5. Bar plots showing the number of species identified at different taxonomy ranks within different identity percentage thresholds for Coffea arabica trnL-P6 and trnL-UAA barcodes.*** *This figure identifies the taxonomy rank on the X-axis, and the number of species on the Y-axis. Bars are colour-coded to represent the different identity percentage thresholds (<98%: grey, 98-100%: orange and 100%: blue). The results for trnL-P6 (left) and trnL-UAA (right) barcodes show the 98-100% and 100% identity thresholds remain consistent throughout the different taxonomy ranks. They also show the <98% identity threshold increases in the number of species as the taxonomy rank goes from Genus to Family*

## Species Identity Percentage

As previously mentioned, *Coffea arabica* is commonly adulterated with cheaper and inferior *Coffea* species, so food producers can avoid the expensive prices of *Coffea arabica*. The most common species that are adulterated with Coffea arabica include coffee husks, twigs, roots, legumes and other roasted grains (Flores-Valdez *et al.,* 2020). Some known adulterants were investigated further within the BLOOM application, and to identify the different hidden clusters with the species sequences BLOOM won't explicitly display to the user. The histogram in Figure 6 is generated using data exported from BLOOM. The BLAST function of BLOOM was used for *Coffea arabica* using barcode trnL-UAA and sequence PP708565.1 where rank was genus. Subsequently, the data was exported to CSV for downstream analysis in R. *Coffea liberica* trnL-UAA sequences were aggregated to show identity percentage metrics for this species compared to the *Coffea arabica* reference blasted. The histogram shows values ranging from 71.1% to 98.8%.

BLOOM assigns the identity of the top BLAST hit (i.e., the sequence with the highest identity percentage) to determine whether a barcode can reliably distinguish a species. This conservative approach prioritizes minimizing false positives but may increase the likelihood of false negatives—that is, instances where BLOOM reports that a species cannot be distinguished using a given barcode, even though post-sequencing analysis reveals that discrimination is indeed possible.

| Scientific.name | Identity.percentage | Score | Accession.Number |
|---|---|---|---|
| Coffea liberica | 98.77193 | 563 | OQ946698 |
| Coffea liberica | 98.77193 | 563 | MW970411 |
| Coffea liberica | 88.07018 | 502 | MN715168 |
| Coffea liberica | 88.07018 | 502 | MN715164 |
| Coffea liberica | 87.71930 | 500 | MN715166 |
| Coffea liberica | 87.36842 | 498 | MN715165 |
| Coffea liberica | 87.01754 | 496 | MN715161 |
| Coffea liberica | 85.26316 | 486 | MN715162 |
| Coffea liberica | 84.73684 | 483 | MN715163 |
| Coffea liberica | 71.05263 | 405 | MN715167 |

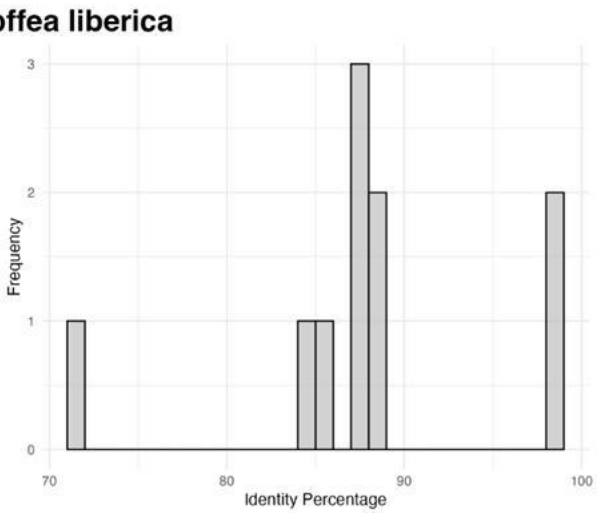
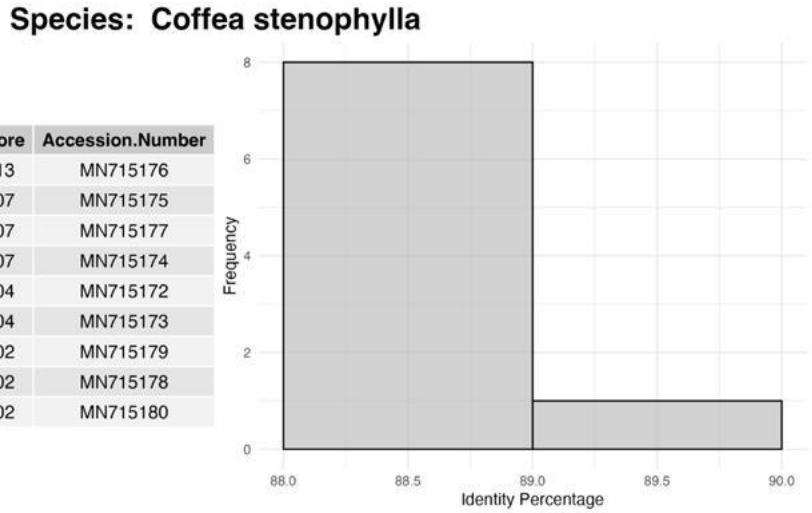



***Figure 6. Coffea liberica trnL-UAA hits compared against Coffea arabica trnL-UAA: PP708565.1.*** *The table displays Coffea liberica trnL-UAA identity percentages, BLOOM score, and accession number against a reference: Coffea arabica trnL-UAA: PP708565.1 (rank = genus). The histogram on the right visualizes the distribution of identity percentages among hits; binwidth = 1.*

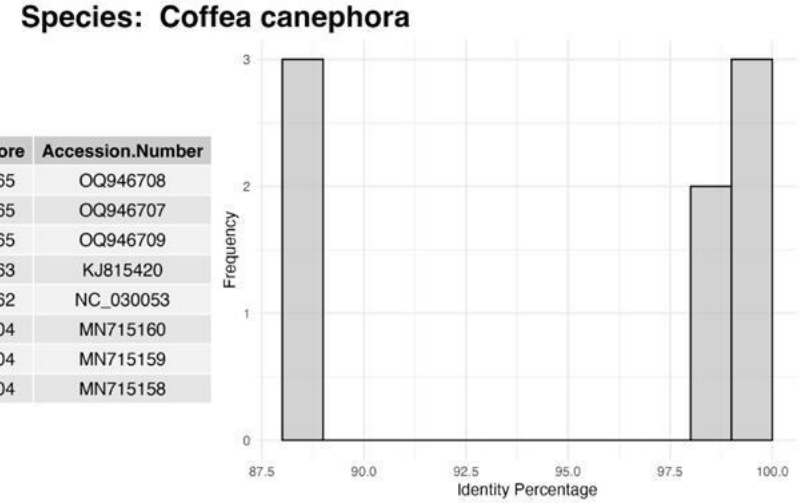
Figure 7 presents a histogram of *Coffea stenophylla* trnL-UAA barcode identity values resulting from a BLAST search against *Coffea arabica* using the trnL-UAA region of sequence PP708565.1 where rank was genus. Identity values range from 88.0% to 90.0%. The narrow range of variation suggests that the trnL-UAA barcode may reliably distinguish *Coffea stenophylla* from *Coffea arabica*.

One of the most adulterated *Coffea* species for *Coffea arabica*, is *Coffea canephora*. Figure 8 presents a histogram of *Coffea canephora* trnL-UAA barcode identity values resulting from a BLAST search against *Coffea arabica* using the trnL-UAA region of sequence PP708565.1 where rank was genus. Identity values form two clusters ranging from 88.0% in the first group to 99.0% in the second. BLOOM uses the highest value (99.12281%) to decide if the barcode can be distinguished. However, the observed clustering could be misleading if the differences between barcode length have not been truly represented. Sequences with lower identity percentage tend to be shorter, which likely contributed to a reduced alignment score. Overall, the BLOOM results suggests that the trnL-UAA barcode may distinguish *Coffea canephora* from *Coffea arabica*.




| Scientific.name | Identity.percentage | Score | Accession.Number |
|---|---|---|---|
| Coffea stenophylla | 90.00000 | 513 | MN715176 |
| Coffea stenophylla | 88.94737 | 507 | MN715175 |
| Coffea stenophylla | 88.94737 | 507 | MN715177 |
| Coffea stenophylla | 88.94737 | 507 | MN715174 |
| Coffea stenophylla | 88.42105 | 504 | MN715172 |
| Coffea stenophylla | 88.42105 | 504 | MN715173 |
| Coffea stenophylla | 88.07018 | 502 | MN715179 |
| Coffea stenophylla | 88.07018 | 502 | MN715178 |
| Coffea stenophylla | 88.07018 | 502 | MN715180 |

***Figure 7 Coffea stenophylla trnL-UAA hits compared against Coffea arabica trnL-UAA: PP708565.1.*** *The table displays Coffea stenophylla trnL-UAA identity percentages, BLOOM score, and accession number against a reference: Coffea arabica trnL-UAA: PP708565.1(rank = genus) . The histogram on the right visualizes the distribution of identity percentages among hits; binwidth =* 1*.*




| Scientific.name | Identity.percentage | Score | Accession.Number |
|---|---|---|---|
| Coffea canephora | 99.12281 | 565 | OQ946708 |
| Coffea canephora | 99.12281 | 565 | OQ946707 |
| Coffea canephora | 99.12281 | 565 | OQ946709 |
| Coffea canephora | 98.77193 | 563 | KJ815420 |
| Coffea canephora | 98.59649 | 562 | NC_030053 |
| Coffea canephora | 88.42105 | 504 | MN715160 |
| Coffea canephora | 88.42105 | 504 | MN715159 |
| Coffea canephora | 88.42105 | 504 | MN715158 |

***Figure 8 Coffea canephora trnL-UAA hits compared against Coffea arabica trnL-UAA: PP708565.1.*** *The table displays Coffea canephora trnL-UAA identity percentages, BLOOM score, and accession number against a reference: Coffea arabica trnL-UAA: PP708565.1 (rank = genus). The histogram on the right visualizes the distribution of identity percentages among hits; binwidth =* 1*.*

Another case where species discrimination can be ambiguous is between *Avena sativa* (common oat) and its common adulterant *Avena barbata* (slender oat). The high degree of sequence similarity between *Avena sativa* and *Avena barbata* can stem from them deriving from the same genus, *Avena*. Figure 9 presents a frequency plot generated for an *Avena barbata* trnL-UAA barcode with reference *Avena sativa* trnL-UAA where rank was genus. The top BLAST hit (NC_044173) shows 100% identity with *Avena sativa*, indicating that BLOOM would classify *Avena barbata* as indistinguishable from *Avena* sativa [MK336398.1]. However, BLOOM also reports ten additional *Avena barbata* accessions with identity values ranging from 75.3% to 85.7% (mode: 85.7%), suggesting that the trnL-UAA region may offer sufficient resolution to differentiate *Avena barbata* from *Avena sativa* in some cases.

Notably, the presence of only a single 100% identity hit raises the possibility that this sequence may be misannotated in the NCBI database.

| Scientific.name | Identity.percentage | Score | Accession.Number |
|---|---|---|---|
| Avena barbata | 100.00000 | 405 | NC_044173 |
| Avena barbata | 85.67901 | 347 | KF996114 |
| Avena barbata | 85.67901 | 347 | KF996089 |
| Avena barbata | 85.67901 | 347 | KF996115 |
| Avena barbata | 85.67901 | 347 | KF996101 |
| Avena barbata | 85.67901 | 347 | KF996091 |
| Avena barbata | 85.67901 | 347 | KF996117 |
| Avena barbata | 85.67901 | 347 | KF996094 |
| Avena barbata | 85.67901 | 347 | KF996106 |
| Avena barbata | 78.76543 | 319 | KF996088 |
| Avena barbata | 75.30864 | 305 | KF996104 |

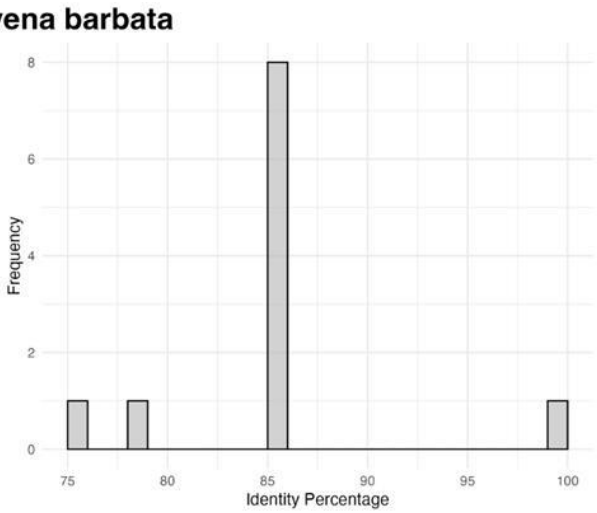



***Figure 9 Avena barbata trnL-UAA hits compared against Avena sativa trnL-UAA: MK336398.1.*** *The table displays Avena barbata trnL-UAA identity percentages, BLOOM score, and accession number against a reference: Avena sativa trnL-UAA: MK336398.1 (rank = genus). The histogram on the right visualizes the distribution of identity percentages among hits ; binwidth* = 1

# Discussion

This report has revealed insight into the functionality and results generated from the BLOOM application. Fundamentally, BLOOM was designed as a tool for biologists to determine the most accurate DNA barcode for taxa identification, before starting the metabarcoding protocol. The results demonstrate the impact on selecting the correct barcode, different parameters and interpreting the species identity percentage within the BLOOM application.

Focusing on the accuracy of taxa identification between trnL-UAA and trnL-P6, BLOOM was utilised to generate results for *Coffea arabica* to determine the barcode with the highest number of taxa discriminated. As a result, for family and subfamily taxonomy rank trnL-UAA discriminated more species compared to trnL-P6, and so greater resolution for these taxonomy levels (Figure 5). Nevertheless, at the tribe and genus level, trnL-UAA and trnL-P6 barcodes both had the same resolution for species discrimination (Table 3). Taberlet *et al*., (2007) evaluated the level of resolution for the plant DNA barcodes, trnL-UAA and trnL-P6, for species identification. They utilised available sequences from GenBank, and discovered the whole chloroplast trnL-UAA intron has a low-resolution percentage of 67.3%, and the trnL-P6 returned an even lower resolution of 19.5%. Whereas Valentini *et al.,* (2009) identified a 50% resolution for the trnL-P6 DNA barcode at the species level, and 90% resolution at the genus level. The differences of the resolution power between DNA barcodes could be due to lower intraspecific variation of the trnL-UAA region, compared to other non-coding regions of the chloroplast DNA, as proposed by Taberlet *et al*., (2007). As trnL-P6 is a structural loop within the whole trnL-UAA chloroplast DNA, this explanation can be applicable to the trnL-P6 DNA barcode too. However, Valentini *et al.,* (2009) and Taberlet *et al*., (2007) both suggested that resolution of trnL-UAA and trnL-P6 DNA barcodes is greater when species originate from a single ecosystem, or they are common eaten plants. This is due to closely related species exhibiting similar DNA barcode sequences, especially in plants, as the chloroplast DNA is evolving slowly.

Additionally, the power of species discrimination is impacted for each barcode when the different BLAST options are selected. Figure 4 displays the trnL-UAA barcode is not impacted for species discrimination for *Coffea arabica* at the family taxonomy rank level. However, trnL-P6 discrimination is impacted, when comparing against the different BLAST modes due to the differences . The 98-100% identity percentage threshold could not be discriminated against trnL-P6 whereas for trnL-UAA it was. This identifies the trnL-P6 barcode cannot discriminate species at the family level, however trnL-UAA can discriminate. Uncu and Uncu (2020) demonstrated the use of the trnL-UAA and trnL-P6 in conjunction with each other, allowing for species discrimination and detection sensitivity, thus increasing resolution power. Incorporating multiple barcodes use to the BLOOM app will provide users insight to more accurate and reliable taxa identification by merging the good qualities of barcodes so unique identification of taxa is met.

Focusing on the level of species discrimination between trnL-UAA and trnL-P6 for *Coffea arabica*, Figure 5 demonstrated the impact of the BLAST taxonomy rank had on species discrimination for the DNA barcodes. As previously stated, trnL-P6 struggled to provide the species discrimination for the 98-100% identity percentage threshold, and it is evident in Figure 5 the same issue occurs from genus to family. This highlights the importance of DNA barcode sequence length, as trnL-P6 has a length of ~70 bp and trnL-UAA has a length of ~500 bp (sequence length derived from the BLOOM application). If there a single base difference in trnL-P6, the percentage identity threshold might not identify it as a <1% difference, indicating the reason for trnL-P6 not having species with a 98-100% identity

percentage. As trnL-UAA is a longer sequence, 1 base change in the barcode sequence will be registered as <1%, resulting in the population of species for the identity percentage 98-100%. Botha et al., (2023) created two datasets for the trnL-UAA and rbcL DNA barcodes to assess the success rate of taxonomical coverage from online databases. They found the rbcL had a better success rate at taxa identification at 85.86% compared to trnL-UAA at 73.72%. Previously, it appeared that for trnL-P6 and trnL-UAA the greater the sequence length the better the species discrimination, however for trnL-UAA the sequence length is ~570 and for rbcL it is between 550-600 bp (Kress & Erickson, 2007). Therefore, we cannot rely on longer sequence length providing more accurate species identity, as during practice longer sequences are prone to amplification error. Botha et al., (2023) prove the separate rbcL and trnL-UAA datasets created need to be used together for effective species discrimination. Botha et al., (2023) and Uncu and Uncu (2020) both agree with the emerging concept of utilizing multiple barcodes to perform DNA metabarcoding in uniform to generate more accurate species identification. Additionally, perhaps a future improvement of the BLOOM application can integrate other databases, such as the Barcode of Life Data Systems (BOLD), to provide a wider scope of DNA barcode sequences available in the application (Ratnasingham & Hebert, 2007).

The BLOOM tool was developed to support biologists in evaluating whether investing time and resources into sequencing a particular DNA barcode would be worthwhile. While BLOOM addresses this specific need, it is not the only tool available for assessing taxonomy performance—other tools and pipelines exist that offer similar functionality.

ecoPCR is a Unix-based tool that simulates PCR by identifying sequences in a reference database. It calculates a metric for taxonomic coverage of a primer pair by calculating the ratio of the number of taxa successfully amplified to the total number of taxa present in the database at a specified taxonomic rank (e.g., species, genus, or family) (Ficetola et al., 2010).

An analogous measure can be estimated from BLOOM results, though it is not directly provided. In BLOOM, taxonomic coverage could be defined as the number of species with identity scores <100% divided by the total number of unique species returned. While BLOOM does not currently report this value explicitly, it provides output in CSV format from which the coverage index can be derived.

For example, using BLOOM to perform a MegaBLAST search with the trnL-UAA barcode and Avena sativa as the reference sequence at the genus level returns 27 unique species, of which 6 exhibit identity scores below 100%. Based on this, the estimated taxonomic coverage of the barcode at the genus level is 6/27, or approximately 0.222. ecoPCR can be used in combination with OBITools (Boyer et al., 2016) for use in matabarcoding *in-silico* analysis. The main drawback of these tools is they use Python 2.7* which is now depreciated. The tools themselves appear no longer supported.

ExTaxsI (Exploring Taxonomy Information) is a bioinformatics tool designed to streamline access to taxonomic data from NCBI, facilitating data download and visualization (Agostinetto et al., 2022). It provides a user-friendly interface aimed at biologists, although initial setup requires familiarity with Python environments. Unlike BLOOM, which offers a one-click executable for Windows users, ExTaxsI focuses on enabling complex NCBI queries—such as txid13443[ORGN] AND TrnL UAA[Gene]—to retrieve FASTA files and generate informative plots. Importantly, ExTaxsI does not support downstream BLAST analysis or result aggregation; its primary utility lies in data acquisition and exploratory visualization.

Genius Prime is a bioinformatics platform designed to simplify research via an intuitive user interface. It automates complex pipelines for high-throughput analysis and scalable workflows, delivering actionable biological insights (*Geneious* | *Bioinformatics Software for Sequence Data Analysis*, n.d.).

The platform includes a BLAST tool that enables users to query NCBI (e.g., txid13443[ORGN] AND TrnL UAA[Gene] for *Coffea arabica* and *trnL-UAA*) and view the results directly. Users can download the data and generate phylogenetic trees within the tool. Unlike BLOOM, Genius Prime does not support direct DNA barcode comparisons and is instead functionally similar to platforms like Illumina BaseSpace and Connected Analytics, which focus on large-scale genomic data analysis pipelines. Furthermore, BLOOM provides a user-friendly interface and a more appealing visualisation for the taxonomy tree, compared to Genius Prime.

Among the tools discussed, BLOOM remains the only one specifically designed for direct barcode comparison. While other platforms offer broader functionalities in data retrieval, taxonomic analysis, or high-throughput processing, BLOOM uniquely addresses the targeted evaluation of barcode effectiveness, making it a valuable resource for researchers focused on marker selection and resolution.

The potential of the BLOOM application, and its capacity to be applied to a wide variety of use cases, ranging from environmental to forensics science, is represented by the analysis of the trnL-UAA barcode of *Coffea arabica*. Its major advantages are its easy-to-use visual results, interactive taxonomy tree and straightforward interpretation of the BLAST results, rendering it a user-friendly resource even for those without extensive experience in bioinformatics.

BLOOM's strengths can be observed in environmental research and in DNA-based monitoring. DNA from complex samples derived from soil, water, or air, is often highly degraded, which would make short barcodes like trnL-UAA and trnL-P6 useful due to their higher amplification even from low-quality samples (Taberlet et al., 2007; Jiang et al., 2017). Considering the impressive speed with which BLOOM can create a taxonomy tree and compare distributions of identity from such sparse sequence data, it can accelerate the detection of species and support the assessment of biodiversity. For example, ecologists studying invasive species or the effects of climate change on plant communities can utilize BLOOM for initial observations without needing a complex bioinformatic pipeline.

In addition to its biodiversity assessment aspects, the visualization of nucleotide mismatches by BLOOM (for example the histogram outputs discussed) could be adapted for pollen barcoding purposes. Even minor sequence variations are enough to differentiate between plant origins, as demonstrated for studies in pollination ecology or honey identification (Bell et al., 2016; Prudnikow et al., 2023). For example, ecologists might use BLOOM to track the origin of pollen or assess floral diversity using samples obtained from bees or the air. Moreover, its barcode identity plots improve readability in situations such as allergen monitoring and environmental control.

The usefulness of BLOOM also applies to forensics sciences. Botanical DNA evidence – like seeds, leaves, water, or pollen – collected from clothing, automobiles and crime scenes can provide important information (Fang et al., 2019). Since trnL-UAA and trnL-P6 markers are efficient in barcoding degraded materials, BLOOM may be used as a useful tool for forensic or botanical analysis.

The taxonomy trees generated by this tool may support verifying whether sequenced contribute to plant species likely to be encountered from the place for a geographic area of interest, and therefore, strengthening the support of DNA-based plant identification.

# Conclusion

Overall, BLOOM presented that it can be used to both evaluate and visualise barcode performance. Comparison between trnL-UAA and trnL-P6 for *Coffea arabica* indicates that trnL-UAA achieves a better resolution for the identification of identical species, which is also supported by previous literature. Furthermore, BLOOM has proven to provide taxonomy information as it displayed identical species are only found within the Genus rank for *Coffea Arabica*. Overall, BLOOM addresses issues that arise with the DNA metabarcoding approach.

BLOOM in its current design also has room for improvements. The barcode selection is limited only to trnL-UAA and trnL-P6 and does not support more variable markers (e.g., matK, rbcL, COI). Adding more markers to BLOOM would be necessary to discriminate at the species level more accurately or to incorporate non-plant barcodes. In addition, BLOOM is currently limited to BLAST searches through NCBI and does not have the compatibility to access other databases. BOLD or region-specific repositories could increase the accuracy in BLOOM and support local biodiversity studies. The application of universal markers such as trnL-UAA and trnL-P6 may also be limited for discrimination of closely related species, especially in herbal product identification or species substitution studies (Swetha et al., 2017; Zhang et al., 2021).

Despite these limitations, the BLOOM tool is a useful extension to DNA metabarcoding. It occupies a complementary space, which is fast, easy-to-use, and provides intuitive visualisations. For applications demanding rapid deployment, accessibility, and visual results for quick interpretation — such as field-based ecological surveillance, or forensic screening—the design of BLOOM is useful in a very practical sense. Future updates with increased barcode coverage and background metadata are likely to extend its use to other domains.

# References


Agostinetto, G., Brusati, A., Sandionigi, A., Chahed, A., Parladori, E., Balech, B., Bruno, A., Pescini, D., & Casiraghi, M. (2022). ExTaxsI: an exploration tool of biodiversity molecular data. *GigaScience*, *11*, 1–10. https://doi.org/10.1093/GIGASCIENCE/GIAB092

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Botha, D., du Plessis, M., Siebert, F., & Barnard, S. (2023). Introducing an rbcL and a trnL reference library to aid in the metabarcoding analysis of foraged plants from two semi-arid eastern South African savanna bioregions. *PLOS ONE*, *18*(5), e0286144. https://doi.org/10.1371/JOURNAL.PONE.0286144

Das, P., Altemimi, A. B., Nath, P. C., Katyal, M., Kesavan, R. krishnan, Rustagi, S., Panda, J., Avula, S. K., Nayak, P. K., & Mohanta, Y. K. (2025). Recent advances on artificial intelligence-based approaches for food adulteration and fraud detection in the food industry: Challenges and opportunities. *Food Chemistry*, *468*. https://doi.org/10.1016/j.foodchem.2024.142439

*Python Software Foundation*. (n.d.). Retrieved May 9, 2025, from https://www.python.org/downloads/

Federhen, S. (2015). Type material in the NCBI Taxonomy Database. *Nucleic Acids Research*, *43*(D1), D1086–D1098. https://doi.org/10.1093/NAR/GKU1127

Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*(1), 434. https://doi.org/10.1186/1471-2164-11-434

Flores-Valdez, M., Meza-Márquez, O. G., Osorio-Revilla, G., & Gallardo-Velázquez, T. (2020). Identification and Quantification of Adulterants in Coffee (Coffea arabica L.) Using FT-MIR Spectroscopy Coupled with Chemometrics. *Foods 2020, Vol. 9, Page 851*, *9*(7), 851. https://doi.org/10.3390/FOODS9070851

Francesco Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F. (2010). *An In silico approach for the evaluation of DNA barcodes*. http://www.grenoble.prabi.fr/

*Geneious | Bioinformatics Software for Sequence Data Analysis*. (n.d.). Retrieved May 7, 2025, from https://www.geneious.com/

Kaasalainen, U., Olsson, S., & Rikkinen, J. (2015). Evolution of the tRNALeu (UAA) Intron and Congruence of Genetic Markers in Lichen-Symbiotic Nostoc. *PLOS ONE*, *10*(6), e0131223. https://doi.org/10.1371/JOURNAL.PONE.0131223

Kress, W. J., & Erickson, D. L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLoS ONE*, *2*(6), e508. https://doi.org/10.1371/journal.pone.0000508

Lanubile, A., Stagnati, L., Marocco, A., & Busconi, M. (2024). DNA-based techniques to check quality and authenticity of food, feed and medicinal products of plant origin: A review. In *Trends in Food Science and Technology* (Vol. 149). Elsevier Ltd. https://doi.org/10.1016/j.tifs.2024.104568

Monterisi, S., Zuluaga, M. Y. A., Porceddu, A., Cesco, S., & Pii, Y. (2023). The Application of High-Resolution Melting Analysis to trnL (UAA) Intron Allowed a Qualitative Identification of Apple Juice Adulterations. *Foods*, *12*(7), 1437. https://doi.org/10.3390/FOODS12071437/S1

National Center for Biotechnology Information. (2025, May). *National Center for Biotechnology Information (NCBI)*. Available from: Https://Www.Ncbi.Nlm.Nih.Gov/.

Olsson, S., Kaasalainen, U., & Rikkinen, J. (2012). Reconstruction of structural evolution in the trnL intron P6b loop of symbiotic Nostoc (Cyanobacteria). *Current Genetics*, *58*(1), 49–58. https://doi.org/10.1007/S00294-011-0364-0/FIGURES/4

*PyInstaller Manual — PyInstaller 6.13.0 documentation*. (n.d.). Retrieved May 9, 2025, from https://pyinstaller.org/en/stable/

*R: The R Project for Statistical Computing*. (n.d.). Retrieved May 9, 2025, from https://www.r-project.org/

RATNASINGHAM, S., & HEBERT, P. D. N. (2007). <scp>bold</scp> : The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Connor, R., Feldgarden, M., Fine, A. M., Funk, K., Hoffman, J., Kannan, S., Kelly, C., Klimke, W., Kim, S., Lathrop, S., Marchler-Bauer, A., Murphy, T. D., O'Sullivan, C., Schmieder, E., … Pruitt, K. D. (2025). Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*, *53*(D1), D20–D29. https://doi.org/10.1093/nar/gkae979

Sugita, M., Luo, L., Ohta, M., Itadani, H., Matsubayashi, T., & Sugiura, M. (1995). Genes Encoding the Group I Intron-containing tRNA Leu and Subunit L of NADH Dehydrogenase from the Cyanobacterium Synechococcus PCC 6301. *DNA Research*, *2*(2), 71–76. https://doi.org/10.1093/DNARES/2.2.71

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., & Willerslev, E. (2007). Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3), e14–e14. https://doi.org/10.1093/NAR/GKL938

Uncu, A. O., & Uncu, A. T. (2020). A barcode-DNA analysis method for the identification of plant oil adulteration in milk and dairy products. *Food Chemistry*, *326*, 126986. https://doi.org/10.1016/J.FOODCHEM.2020.126986

Uncu, A. T., Uncu, A. O., Frary, A., & Doganlar, S. (2017). Barcode DNA length polymorphisms vs fatty acid profiling for adulteration detection in olive oil. *Food Chemistry*, *221*, 1026–1033. https://doi.org/10.1016/J.FOODCHEM.2016.11.059

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E., Coissac, E., Pompanon, F., Gielly, L., Cruaud, C., Nascetti, G., Wincker, P., Swenson, J. E., & Taberlet, P. (2009). New

perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Resources*, *9*(1), 51–60. https://doi.org/10.1111/J.1755-0998.2008.02352.X

Yoshiyama, K. O., Kobayashi, J., Ogita, N., Ueda, M., Kimura, S., Maki, H., & Umeda, M. (2013). ATM-mediated phosphorylation of SOG1 is essential for the DNA damage response in Arabidopsis. *EMBO Reports*, *14*(9), 817–822. https://doi.org/10.1038/EMBOR.2013.112/SUPPL_FILE/EMBR2013112.REVIEWER_COMMENTS.PDF