Revelation and Science
Special Issue: Plant Science 2 (1446H/2024) 1-14

IN SILICO CHARACTERIZATION AND PHYLOGENETIC ANALYSIS OF *TRYPTOPHAN DECARBOXYLASE* (*TDC*) GENE IN *Mitragyna speciosa*

Nurul Izzah Mohammad Azlee[1], Nurul Izzahtul Liyana Azman[1], Sarah Hani Mohd Salleh[1], Zarina Zainuddin[1,2], Nurul Hidayah Samsulrizal[*1,2]

[1]Department of Plant Science, Kulliyyah of Science, International Islamic University of Malaysia (IIUM), Jalan Sultan Ahmad Shah, 25200 Kuantan, Pahang, Malaysia.

[2] Plant Productivity and Sustainable Resource Unit (PPaSRU), Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, 25200 Kuantan, Pahang, Malaysia.

ABSTRACT

Mitragyna speciosa also known as “kratom’ or ‘ketum’, it is a medicinal herb. It has been recognized for its medicinal benefits, but it contains a toxic compound known as Mitragynine. This compound is known for its capability to cause adverse reactions such as addiction, misuse and death. Due to its addictive properties, few countries have made the possession of *M. speciosa* illegal. The *tryptophan decarboxylase* (*TDC*) gene involved in the biosynthesis pathway to produce Mitragynine, which is known to contribute to the toxicity effects of *M. speciosa*. Thus, the aim of this study is to identify, characterize and analyze the *TDC* gene in *M. speciosa* using bioinformatic tools. In this study, the homologous search using Blastx revealed the highest percentage of identity, score, and E-value for Tryptophan decarboxylase *M. speciosa* protein with sequence ID AEQ01059.1. The presence of entry IPR002129 and PF00282 for the TDC *M. speciosa* protein sequence, which is available at positions 57-430 aa. Plus, 11 out of 100 homologous protein sequences from BLAST have been selected to construct the phylogenetic tree by using MEGA-X software, revealing that this protein belongs to the same family as *Coffea arabica*, *Coffea euginioides*, *Ophiorrhiza pumila* and *Ophiorriza prostrata* species, which is the Rubiaceae family. Moreover, physicochemical characteristics by Protparam revealed that TDC *M. speciosa* stable and exhibit hydrophilic nature. Thus, the findings of this work will contribute to better knowledge of TDC *M. speciosa* protein as well as related information regarding the biosynthesis pathway of Mitragynine alkaloid that contributes to the toxicity effect in *M. speciosa*.

Keywords: *Mitragyna speciosa*, Kratom, tryptophan decarboxylase (TDC), phylogenetic analysis

ABSTRAK

Mitragyna speciosa juga dikenali sebagai "kratom' atau 'ketum', adalah herba perubatan. Ia telah diiktiraf berdasarkan faedah perubatannya, tetapi mengandungi sebatian toksik yang dikenali sebagai Mitragynine. Kompaun ini terkenal dengan keupayaannya untuk menyebabkan reaksi buruk seperti ketagihan, penyalahgunaan dan kematian. Disebabkan sifat ketagihannya, beberapa negara telah menjadikan pemilikan *M. speciosa* sebagai haram. Gen *tryptophan decarboxylase* (*TDC*) yang terlibat dalam laluan biosintesis untuk menghasilkan Mitragynine, yang diketahui menyumbang kepada kesan ketoksikan *M. speciosa*. Oleh itu, matlamat kajian ini adalah untuk mengenal pasti, menciri dan menganalisis gen *TDC* dalam *M. speciosa* menggunakan alat bioinformatik. Dalam kajian ini, carian homolog menggunakan BlastX mendedahkan peratusan tertinggi identiti, skor dan nilai E untuk protein Tryptophan decarboxylase *M. speciosa* dengan ID jujukan AEQ01059.1. Kehadiran domain IPR002129 dan PF00282 untuk jujukan protein *TDC M. speciosa,* pada kedudukan 57-430 aa. Selain itu, 11 daripada 100 jujukan protein homolog daripada BLAST telah dipilih untuk membina pokok filogenetik dengan menggunakan perisian MEGA-X, telah mendedahkan bahawa protein ini tergolong dalam keluarga yang sama seperti *Coffea arabica*, *Coffea euginioides*, *Ophiorrhiza pumila* dan spesies *Ophiorriza prostrata*, iaitu keluarga Rubiaceae. Selain itu, ciri fizikokimia oleh ProtParam mendedahkan bahawa protein *TDC M. speciosa* stabil dan mempamerkan sifat hidrofilik. Oleh itu, penemuan kerja ini akan menyumbang kepada pengetahuan yang lebih baik tentang protein *TDC M. speciosa* serta maklumat berkaitan mengenai laluan biosintesis alkaloid Mitragynine yang menyumbang kepada kesan ketoksikan dalam *M. speciosa*.

Kata kunci: *Mitragyna speciosa*, Kratom, tryptophan decarboxylase (TDC), analysis filogenetik

**Corresponding author:*
Nurul Hidayah Samsulrizal
Kulliyyah of Science,
International Islamic University Malaysia
Email: ***hidayahsamsulrizal@iium.edu.my***

Introduction

M. speciosa Korth, also known as 'kratom' in Thailand or 'ketum' in Malaysia, is a medicinal herb from the Rubiaceae botanical family that exhibits a wide distribution within its native range of Southeast Asia, specifically encompassing Thailand, Malaysia, and Indonesia (Abdullah *et al.*, 2021). Moreover, *M. speciosa* predominantly thrives in tropical areas located in Southeast Asia and Africa (Hassan *et al.,* 2013). The leaves exhibit a dark green colour, possess an elliptical shape and have been conventionally used for its medicinal benefits. The leaves of kratom are known to relieve pain, fatigue, treat diarrhoea, enhance work productivity, and reduce opioid-related withdrawal symptoms (Zhang *et al.,* 2020).

M. speciosa consumption is thought to have "cocaine-like" and/or "opiate-like" effects depending on dosage and preparation method, with the latter leading users to consume products i.e., powder, capsules, or teas to alleviate pain or opiate addiction (Brown *et al.,* 2017). This has been a concern as overdosage of this plant can cause adverse reactions like addiction, misuse or even death. Due to its addictive properties, few countries have passed laws regulating the cultivation and possession of kratom. *M. speciosa* contains several pharmacologically active alkaloids including mitragynine, 7-hydroxy mitragynine, mitraphylline, speciociliatine, speciogynine, and paynantheine (Meireles *et al.*,2019) which have both stimulant and opiate-like effects (Feng *et al.*, 2017).

Mitragynine is the most abundant alkaloid found in *M. speciosa*, contributing to 66.2% of *M. speciosa* total alkaloid content (Takayama, 2004). Tryptophan decarboxylase (TDC) is a cytosolic enzyme that catalyzes an early step of the terpenoid indole alkaloid biosynthetic pathway by decarboxylation of L-tryptophan to produce the proto-alkaloid tryptamine (Di Fiore *et al.*, 2002). Tryptamine is a biogenic amine and serves as a precursor for the synthesis of various bioactive compounds, including alkaloids (McDonald *et al.*, 2019). Tryptophan decarboxylase is not directly involved in the production of mitragynine. Instead, tryptophan is one of the starting materials and its conversion into mitragynine is a multistep process. However, in general, the process includes the conversion of tryptophan to strictosidine, a key intermediate in the biosynthesis of various alkaloids. Tryptophan decarboxylase (TDC) catalyses the first committed step of indole alkaloid production by decarboxylating TDC into tryptamine (Glenn *et al*., 2011).

Thus, tryptophan decarboxylase has been studied in this research as it is part of regulatory sites of mitragynine biosynthesis. As a result, molecular characterization of mitragynine in *M. speciosa* would contribute to a better understanding of the activities of the toxicity genes involved in this plant. This research aims to characterise and analyse the *TDC* gene of *M. speciosa* using bioinformatics tools, identifying TDC homolog proteins and the domains present. Multiple sequence alignment (MSA) and phylogenetic analysis were conducted to examine the evolutionary relationships of these homolog proteins. We expect that this study will provide new insights into the functional characterization of *M. speciosa*, which could be valuable for future research on *M. speciosa* and its associated toxicity gene, and mitragynine.

Methodology

2.1 Homology Search of *TDC* Gene

The *in-silico* analysis of nucleotide sequence of *M. speciosa TDC* mRNA, complete coding sequence (CDS) was chosen from National Center for Biotechnology Information (NCBI) database (Accession number: JN643922.1). The nucleotide sequence of *M. speciosa* TDC mRNA, complete CDS was translated into protein sequence using Blast X program from NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). This Blast X program analyse DNA sequence and searches for homologous protein sequences. In order to analyse the results of the homology search, parameters like alignment scores, E-values and sequence similarity were considered. Eleven homologous protein sequences were selected in the process and the parameters were maintained as default i.e., Expect threshold= 0.05, Matrix= BLOSUM62 and Gap Extension=1. Additionally, the selection of homologous proteins were made based on differences of the species and three parameters: (i) percentage identity (100%), (ii) E-value (0.0), and (iii) total score (2659).

2.2 Protein Domains and Families Analysis

The protein sequence was analysed using a domain search through the InterPro database (https://www.ebi.ac.uk/interpro/), which helps predict and integrate sequences by representing protein domains, families, and functional sites. Additionally, InterPro Scan, which integrates 13 databases, including Pfam, was used. The Pfam database (https://www.ebi.ac.uk/interpro/entry/pfam/#table) specifically identified the location of protein domains within the sequence. Pfam is a curated database of protein families that utilizes the HMMER3 program. The HMM (Hidden Markov Model) was profiled with PSI-BLAST and searched against UniProt (Punta et al., 2012). For a protein to be included in a family, its domain and sequence bit scores must meet or exceed the Gathering Thresholds (GA). The results from the two databases were then compared and analysed.

2.3 Physicochemical Properties of TDC Protein

The translated protein sequence of TDC of *M. speciosa* was inserted in the query box of the ProtParam tool (https://web.expasy.org/protparam/). The output like molecular weight, theoretical pI and amino acid composition, Extinction Coefficient (EC), Instability Index (II), Aliphatic Index (AI) and Grand average of Hydrophobicity (GRAVY) were calculated using this program.

2.4 Phylogenetic Analysis of TDC protein

Constructing a phylogenetic tree involves two main steps: (1) Multiple Sequence Alignment (MSA) and (2) using the Molecular Evolutionary Genetics Analysis (MEGA X) software, available for download at (https://www.megasoftware.net/). A set of eleven selected protein sequences, including the TDC of *M. speciosa*, were input into the MUSCLE software (accessible at https://www.ebi.ac.uk/Tools/msa/muscle/) to generate a multiple alignment of protein sequences. These sequences were chosen based on their high and low consensus quality. The next step involved constructing a phylogenetic tree using MEGA X software, which allows for a comparative analysis of the evolutionary relationship between the query sequence and reference sequences. The phylogenetic tree was created using eleven protein sequences homologous to the *M. speciosa* TDC protein sequence. The multiple sequence alignment data were saved in a *.meg format file before proceeding with phylogeny analysis using the Maximum Likelihood method and the JTT matrix-based model (Tamura *et al.,* 2011). Additionally, a bootstrap consensus tree was generated with 1000 replicates.

Results and Discussion

3.1 Homology Search Using BLASTX

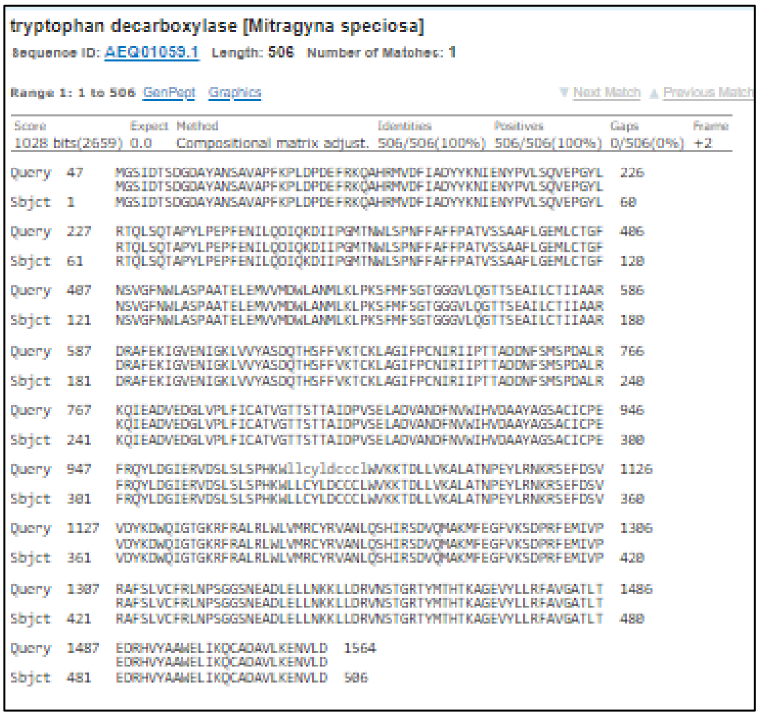
To discover homologous sequences, the obtained nucleotide sequence underwent analysis with NCBI tools, specifically BlastX. The BlastX search revealed over 100 sequences associated with the *TDC* gene in non-redundant database (*nr*). The top hit of protein sequence homolog to the *M. speciosa* TDC is shown in Table 3.1. As stated by Yim & Cushman (2017), BLAST is calculated from three parameters: the bit score, the length of query and the size of the database. This emphasizes the importance of the bit score in determining the statistical significance of sequence similarities identified by BLAST. The E-value and bit score aid in assessing the quality of each match (Shah *et al.*, 2018).

The results were sorted according to the Total Score or Bit score. The bit score gives an indication of how good the alignment is; the higher the score, the better the alignment and it is used to identify the presence of gaps between the query and the subject sequences. It is sorted from the highest Total Score to the lowest. Moreover, a lower E-value indicates a more significant similarity between sequences, with E-values below a certain threshold considered to be potentially meaningful matches. The lower the E-value, the more significant the hit (Madden, 2003). Based on the results, all eleven protein sequences obtained an E-value of 0.0. Out of 100 top-hit homologous proteins obtained via Blast X, only eleven homology protein sequences are selected based on desired parameters. The most similar alignment for the *M. speciosa* TDC is the TDC *M. speciosa* sequence with the accession ID AEQ01059.1. It is identified as the best protein sequence as it resulted in the most similar protein hit via Blast X. TDC *M. speciosa* also matched the parameters: (i) percentage identity (100%), (ii) E-value (0.0), and (iii) total score (2659).

The sequence identity is the number of identical bases between the query and the subject sequences (Figure 3.1). From the pairwise alignment, the sequence identity shows 506/506 (100%) with no gap (0%) and the positive values 506/506 (100%) where it represents the number of residues that either share the same chemical properties or are identical to each other (Khan *et al.*, 2021).

Table 3.1: Top Hit of Homology Search of *M. speciosa Tryptophan decarboxylase* gene (*TDC*) using BlastX.

ACCESSIONS	DESCRIPTION	IDENTITY (%)	TOTAL SCORE	E-VALUE
AEQ01059.1	Tryptophan decarboxylase (*Mitragyna speciosa)*	100	2659	0.0
XP_027063707.1	Tryptophan decarboxylase TDC2-like (*Coffea arabica)*	85.71	904	0.0
XP_027172386.1	Tryptophan decarboxylase TDC2-like (*Coffea eugenioides)*	85.91	902	0.0
QLQ34389.1	Tryptophan decarboxylase 2 (*Ophiorrhiza pumila)*	86.14	888	0.0
ABU40982.1	Tryptophan decarboxylase, partial (*Ophiorrhiza prostrata)*	79.92	833	0.0
XP_009622485.1	Tryptophan decarboxylase TDC1-like (*Nicotiana tomentosiformis)*	75.97	816	0.0
XP_059284124.1	Tryptophan decarboxylase TDC1-like (*Lycium ferocissimum)*	76.14	807	0.0
AON76721.1	Tryptophan decarboxylase (*Camptotheca acuminata)*	77.73	804	0.0
AXK92562.1	Tryptophan decarboxylase (*Gelsemium sempervirens)*	76.34	801	0.0
XP_015081169.1	Tryptophan decarboxylase TDC1-like (*Solanum pennellii)*	73.47	788	0.0
PHT43311.1	Tyrosine/ DOPA Decarboxylase 3 (*Capsiccum cabbatum*)	75.25	788	0.0



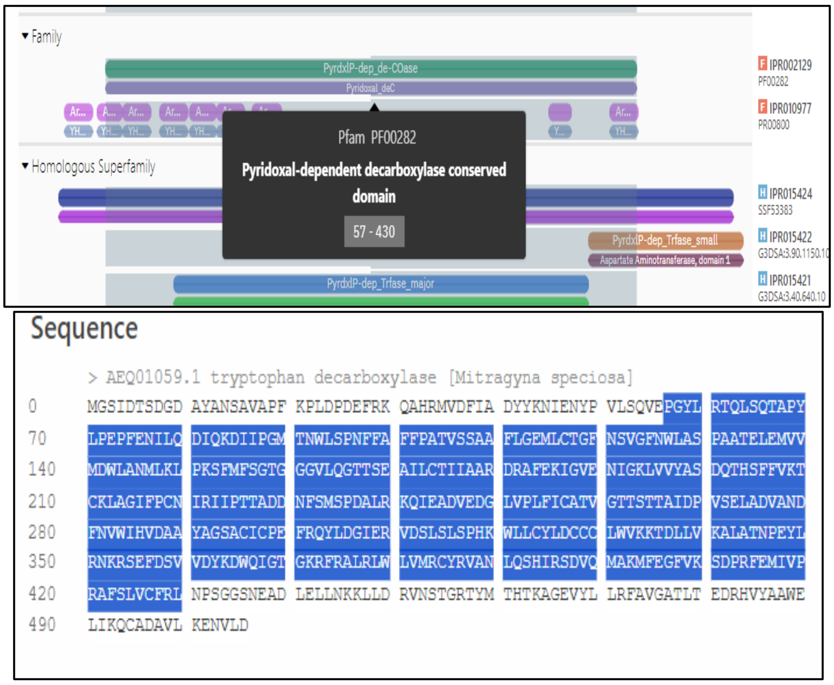
```
tryptophan decarboxylase [Mitragyna speciosa]
Sequence ID: AEQ01059.1  Length: 506  Number of Matches: 1

Range 1: 1 to 506  GenPept  Graphics                          ▼ Next Match  ▲ Previous Match

Score             Expect  Method                        Identities     Positives      Gaps       Frame
1028 bits(2659)   0.0     Compositional matrix adjust.  506/506(100%)  506/506(100%)  0/506(0%)  +2

Query  47    MGSIDTSDGDAYANSAVAPFKPLDPDEFRKQAHRMVDFIADYYKNIENYPVLSQVEPGYL  226
             MGSIDTSDGDAYANSAVAPFKPLDPDEFRKQAHRMVDFIADYYKNIENYPVLSQVEPGYL
Sbjct  1     MGSIDTSDGDAYANSAVAPFKPLDPDEFRKQAHRMVDFIADYYKNIENYPVLSQVEPGYL  60

Query  227   RTQLSQTAPYLPEPFENILQDIQKDIIPGMTNWLSPNFFAFFPATVSSAAFLGEMLCTGF  406
             RTQLSQTAPYLPEPFENILQDIQKDIIPGMTNWLSPNFFAFFPATVSSAAFLGEMLCTGF
Sbjct  61    RTQLSQTAPYLPEPFENILQDIQKDIIPGMTNWLSPNFFAFFPATVSSAAFLGEMLCTGF  120

Query  407   NSVGFNWLASPAATELEMVVMDWLANMLKLPKSFMFSGTGGGVLQGTTSEAILCTIIAAR  586
             NSVGFNWLASPAATELEMVVMDWLANMLKLPKSFMFSGTGGGVLQGTTSEAILCTIIAAR
Sbjct  121   NSVGFNWLASPAATELEMVVMDWLANMLKLPKSFMFSGTGGGVLQGTTSEAILCTIIAAR  180

Query  587   DRAFEKIGVENIGKLVVYASDQTHSFFVKTCKLAGIFPCNIRIIPTTADDNFSMSPDALR  766
             DRAFEKIGVENIGKLVVYASDQTHSFFVKTCKLAGIFPCNIRIIPTTADDNFSMSPDALR
Sbjct  181   DRAFEKIGVENIGKLVVYASDQTHSFFVKTCKLAGIFPCNIRIIPTTADDNFSMSPDALR  240

Query  767   KQIEADVEDGLVPLFICATVGTTSTTAIDPVSELADVANDFNVWIHVDAAYAGSACICPE  946
             KQIEADVEDGLVPLFICATVGTTSTTAIDPVSELADVANDFNVWIHVDAAYAGSACICPE
Sbjct  241   KQIEADVEDGLVPLFICATVGTTSTTAIDPVSELADVANDFNVWIHVDAAYAGSACICPE  300

Query  947   FRQYLDGIERVDSLSLSPHKWllcyldccclWVKKTDLLVKALATNPEYLRNKRSEFDSV  1126
             FRQYLDGIERVDSLSLSPHKWLLCYLDCCCLWVKKTDLLVKALATNPEYLRNKRSEFDSV
Sbjct  301   FRQYLDGIERVDSLSLSPHKWLLCYLDCCCLWVKKTDLLVKALATNPEYLRNKRSEFDSV  360

Query  1127  VDYKDWQIGTGKRFRALRLWLVMRCYRVANLQSHIRSDVQMAKMFEGFVKSDPRFEMIVP  1306
             VDYKDWQIGTGKRFRALRLWLVMRCYRVANLQSHIRSDVQMAKMFEGFVKSDPRFEMIVP
Sbjct  361   VDYKDWQIGTGKRFRALRLWLVMRCYRVANLQSHIRSDVQMAKMFEGFVKSDPRFEMIVP  420

Query  1307  RAFSLVCFRLNPSGGSNEADLELLNKKLLDRVNSTGRTYMTHTKAGEVYLLRFAVGATLT  1486
             RAFSLVCFRLNPSGGSNEADLELLNKKLLDRVNSTGRTYMTHTKAGEVYLLRFAVGATLT
Sbjct  421   RAFSLVCFRLNPSGGSNEADLELLNKKLLDRVNSTGRTYMTHTKAGEVYLLRFAVGATLT  480

Query  1487  EDRHVYAAWELIKQCADAVLKENVLD  1564
             EDRHVYAAWELIKQCADAVLKENVLD
Sbjct  481   EDRHVYAAWELIKQCADAVLKENVLD  506
```

Figure 3.1: Top hit of Blast X of Tryptophan decarboxylase *M. speciosa* with 2659-bit score,

100% identity and E-value of 0.0

3.2 Protein Domains and Families

Protein domains are the fundamental building blocks of proteins, capable of folding, functioning, and evolving independently. Knowledge of protein domains is crucial for protein classification, understanding their biological roles, annotation of their evolutionary mechanisms, and protein design (Wang *et al*., 2021). In this study, the Interpro domain search revealed the presence of entry IPR002129 and PF00282 for the TDC *M. speciosa* protein sequence, which is available at positions 57-430 aa (Figure 3.2). The IPR002129 is known as pyridoxal phosphate-dependent decarboxylase (Figure 3.3) That belongs to group II pyridoxal-dependent carboxylases, which are the family of enzymes includes aromatic-L-amino acid decarboxylase, which catalyzes the decarboxylation of tryptophan to tryptamine.



```
> AEQ01059.1 tryptophan decarboxylase [Mitragyna speciosa]
0     MGSIDTSDGD AYANSAVAPF KPLDPDEFRK QAHRMVDFIA DYYKNIENYP VLSQVEPGYL RTQLSQTAPY
70    LPEPFENILQ DIQKDIIPGM TNWLSPNFFA FFPATVSSAA FLGEMLCTGF NSVGFNWLAS PAATELEMVV
140   MDWLANMLKL PKSFMFSGTG GGVLQGTTSE AILCTIIAAR DRAFEKIGVE NIGKLVVYAS DQTHSFFVKT
210   CKLAGIFPCN IRIIPTTADD NFSMSPDALR KQIEADVEDG LVPLFICATV GTTSTTAIDP VSELADVAND
280   FNVWIHVDAA YAGSACICPE FRQYLDGIER VDSLSLSPHK WLLCYLDCCC LWVKKTDLLV KALATNPEYL
350   RNKRSEFDSV VDYKDWQIGT GKRFRALRLW LVMRCYRVAN LQSHIRSDVQ MAKMFEGFVK SDPRFEMIVP
420   RAFSLVCFRL NPSGGSNEAD LELLNKKLLD RVNSTGRTYM THTKAGEVYL LRFAVGATLT EDRHVYAAWE
490   LIKQCADAVL KENVLD
```

Figure 3.2: The sequence position (57-430 aa) that is responsible for the decarboxylation of tryptophan to tryptamine.

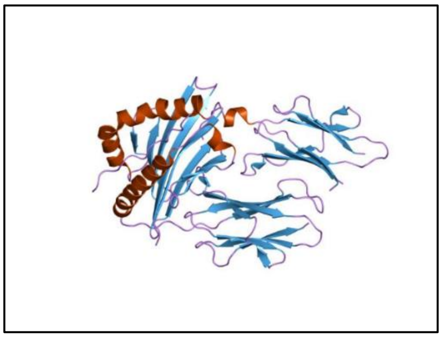


Figure 3.3: Pyridoxal-Dependent Decarboxylase conserved domain (IPR 002129/ PF00282)

Pyridoxal phosphate is the active form of vitamin B6 (pyridoxine or pyridoxal). Pyridoxal 5'-phosphate (PLP) is a versatile catalyst that acts as a coenzyme in a variety of processes, including decarboxylation, deamination, and transamination (Eliot and Kirsch, 2004). PLP-dependent enzymes are primarily involved in the biosynthesis of amino acids and amino acid-derived metabolites, but they are also found in the biosynthesis pathways of amino sugars and in the synthesis or catabolism of neurotransmitters; pyridoxal phosphate can also inhibit DNA polymerases and several steroid receptors (Mozzarelli and Bettati, 2006). Among the enzymes that are PLP-dependent is the aromatic L-amino-acid decarboxylase (L-dopa decarboxylase or tryptophan decarboxylase), which catalyzes the decarboxylation of tryptophan to tryptamine (Han and Shin, 2022). This process is one of the components in the multistep reactions involved in the production of mitragynine alkaloid. Another enzyme dependent on pyridoxal phosphate (PLP) is histidine decarboxylase, responsible for the decarboxylation of histidine to histamine (Hirasawa, 2019).

3.3 Physicochemical Properties of TDC Protein

As shown in Table 3.2, the TDC *M. speciosa* protein has 506 amino acids and a molecular weight of 56,582.97 Daltons. The isoelectric point (pI) is the pH at which the protein carries no net electrical charge, with its positive and negative charges balancing out, making the protein electrically neutral. This is a crucial analytical and molecular parameter (Mohanta *et al.*, 2022). For this protein, the isoelectric point is 5.22, indicating its acidic nature. According to Brenner (2022), the isoelectric point is significant in the selective extraction, separation, and identification of proteins, emphasizing its importance in protein analysis.

The instability index (II) of TDC *M. speciosa* (Sequence ID: AEQ01059.1) is 43.17. According to Gamage *et al.* (2019), a protein is considered stable if its instability index is below 40. Thus, the TDC *M. speciosa* protein is deemed stable. The instability index is an essential parameter as it provides insights into a protein's stability and potential for denaturation. Additionally, the aliphatic index and GRAVY were found to be 88.87 and -0.007, respectively. The aliphatic index measures the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine) and is considered a beneficial factor for increasing the thermostability of globular proteins. According to Osorio (2015), a higher aliphatic index indicates more aliphatic side chains and greater stability. GRAVY represents the average hydropathy value of a peptide or protein, with negative scores indicating hydrophilicity and positive scores indicating hydrophobicity (Wang *et al.*, 2021). The negative GRAVY score for TDC *M. speciosa* suggests it is hydrophilic. This information can help to predict protein properties and assist in empirical research.

Table 3.2: Physicochemical properties analysis of TDC protein *M. speciosa*.

Type	Value
Number of amino acids	506
Molecular weight	56582.97
Theoretical pI	5.22
Instability index (II)	43.17
Aliphatic index (AI)	88.87
Grand average of hydropathicity (GRAVY)	-0.007

3.4 Phylogenetic Tree Analysis Using MEGA-X

Multiple Sequence Alignment (MSA) of the 11 candidate sequences was performed using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/). This process helps illustrate the conserved regions of TDC within *M. speciosa*. Figure 3.4 displays the eleven selected proteins along with their accession numbers, with results obtained from MSA using MUSCLE software. In the results, the blue box highlights highly conserved proteins, indicated by colons (:), which signify strong conservation or similarity among sequences at those positions. Semi-conserved substitutions are shown in the green box with a period (.), indicating some degree of conservation but lower similarity compared to positions marked with an asterisk or colon. The red box with an asterisk (*) marks positions where all sequences share an identical amino acid or nucleotide, demonstrating a high level of conservation and complete identity among the sequences at that position.

These symbols visually represent the degree of similarity or difference between sequences at each position in the alignment (Pais *et al.,* 2014). Multiple sequence alignments are essential for many sequence analysis methods and are a routine step in analyzing homologous genes or proteins (Boyce *et al.,* 2015). With the rapid increase in new sequence data, developing efficient and accurate automatic algorithms for multiple alignment is crucial (Pais *et al.,* 2014). The biological significance of the alignments produced by current methods for multiple sequence alignment determines their quality (Prokić *et al.,* 2009).

The examination of TDC Multiple Sequence Alignment (MSA) reveals richer information compared to pairwise alignments, as it highlights conserved regions with other protein families. Additionally, the MSA results indicate the presence of the IPR002129 domain (position: 57–430 aa), which is depicted in Figure 3.4 and predominantly situated within the marked red box, blue box and green box.

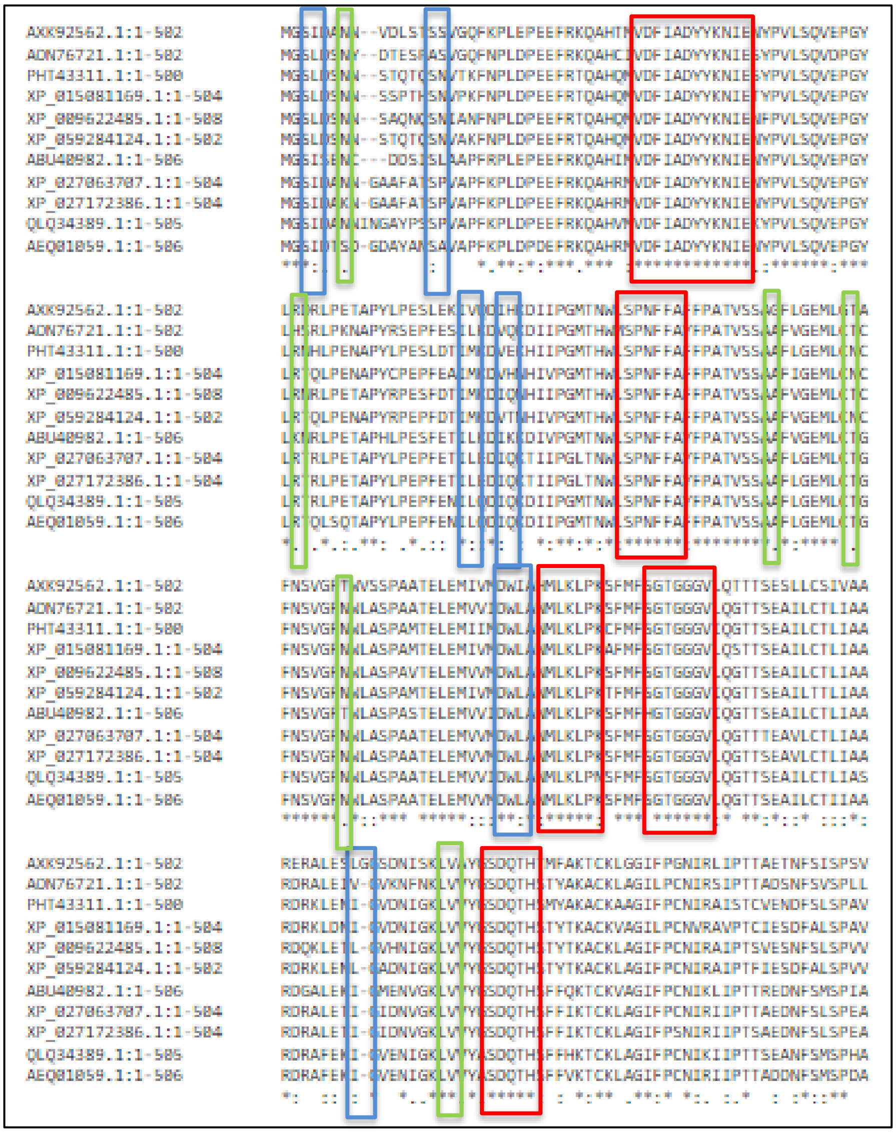


Figure 3.4: Multiple Alignment Sequence of *M. speciosa* TDC protein using MUSCLE.

The phylogenetic tree visually represents the evolutionary relationships among various biological species (Cavalier-Smith, 2010). In this study, we used the Maximum Likelihood method along with the JTT matrix-based model to provide a robust framework for conducting phylogenetic analyses and examining the evolutionary relationships among the protein sequences. The evolutionary history of the taxa studied is depicted by the bootstrap consensus tree derived from 1,000 replications. The bootstrap percentages shown after the test indicate the reliability of each node in the tree (Hall, 2013). The numbers near the nodes represent the Bootstrap (BS) values, which measure the statistical support for a specific branch or clade in the phylogenetic tree and estimate the reliability of the inferred evolutionary relationships. Higher bootstrap values indicate strong support for the clustering of certain species (Zhou *et al.*, 2018). According to Soltis & Soltis (2003), a BS value greater than 95% strongly supports the ancestor clade, while a BS value less than 5% does not support it. Therefore, using bootstrap values is crucial for assessing the robustness of the phylogenetic tree and the confidence in the inferred evolutionary relationships (Gabriel et al., 2014). This analysis involves eleven protein sequences, categorized into two groups based on plant families: Group 1 and Group 2. All proteins in Group 1 have a significant relationship with TDC *M. speciosa* (Figure 3.5). The TDC proteins from *Coffea arabica*, *Coffea eugenioides*, *Ophiorrhiza pumila*, and *Ophiorrhiza prostrata* belong to the Rubiaceae family, the same family as *M. speciosa*. In Group 2, the TDC proteins *from Nicotiana tomentosiformis*, *Lycium ferocissium*, *Solanum pennelli*, and *Capsicum baccatum* are identified as members of the Solanaceae family. The BS value is influenced by genetic variation and speciation due to evolutionary processes within these species.

The phylogenetic analysis results show that TDC *M. speciosa* is in the same clade as Tryptophan decarboxylase TDC-2 from *Coffea arabica* and *Coffea eugenioides*. This is because these organisms belong to the Rubiaceae family. The BS value for *M. speciosa*, *Coffea arabica*, and *Coffea eugenioides* indicates a strong relationship within this clade. Phylogenetic studies, such as those by Bremer and Eriksson (2009), have provided valuable insights into the evolutionary connections within the Rubiaceae family, supporting the classification of *M. speciosa* and *Coffea species* within the same taxonomic group. Additionally, genomic analyses, comparisons of chloroplast genomes, and molecular phylogenetic research have all contributed to understanding the evolutionary relationships among species within the Rubiaceae family. This body of evidence consistently reinforces the grouping of *M. speciosa* and Coffea species in the same taxonomic category (Zhang *et al.*, 2019; Chen *et al.*, 2022).

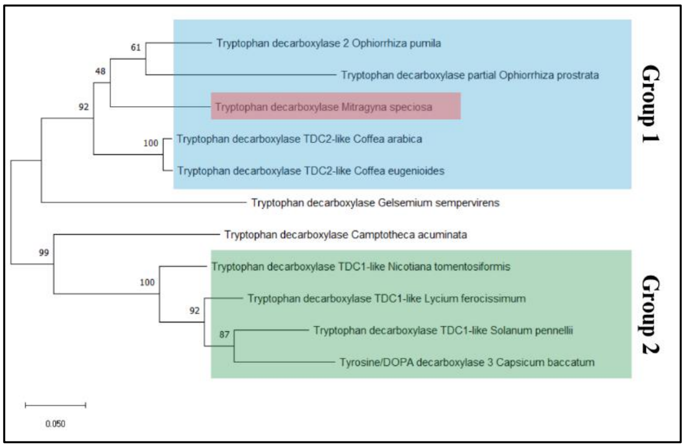


Figure 3.5: Phylogenetic analysis of Tryptophan Decarboxylase *M. speciosa* using MEGA X.

Conclusion

The purpose of this study is to use bioinformatic methods to characterise and analyse the *TDC* gene in *M. speciosa*. The nucleotide sequence of *TDC M. speciosa*, mRNA, complete CDS has been retrieved, and the closest hit protein sequence is the TDC *M. speciosa* protein AEQ01059.1. Supported by a domain or family revealed the presence and entry of IPR002129 and PF00282. The phylogenetic analysis of TDC *M. speciosa* and other similar proteins revealed that TDC *M. speciosa*, *C. arabica*, *C. euginioides*, *O. pumila*, and *O. prostrasta* all belong to the Rubiaceae family with the bootstrap values were 92%, indicating that the clade is well supported. This study has successfully addressed a knowledge gap in the *TDC* gene of *M. speciosa* using in silico sequence-based techniques. It would be very helpful to use this information to manipulate the structure to characterize the mitragynine component, which is known to contribute to the toxicity effect of *M. speciosa*.

Acknowledgement

We would like to express our gratitude to the staff and students of the Kulliyyah of Science, IIUM, who contributed to this work. We also acknowledge the financial support from Regal Malay Capital Berhad [SPP22-137-0137].

References

Abdullah, L., M. F. I., & Singh, D. (2021). The adverse cardiovascular effects and cardiotoxicity of kratom (*Mitragyna speciosa* Korth.): A Comprehensive Review. *Frontiers in pharmacology*, *12,* 726003. https://doi.org/10.3389/fphar.2021.726003

Boyce, K., Sievers, F., & Higgins, D. G. (2015). Instability in progressive multiple sequence alignment algorithms. *Algorithms for Molecular Biology: AMB*, *10*(1). https://doi.org/10.1186/s13015-015-0057-1

Bremer, B., & Eriksson, T. (2009). Time tree of Rubiaceae: Phylogeny and dating the family, subfamilies, and tribes. *International Journal of Plant Sciences*, *170*(6), 766–793. https://doi.org/10.1086/599077

Brenner, T. (2022). Graph neural networks for prediction of protein isoelectric points. In *ChemRxiv*. https://doi.org/10.26434/chemrxiv-2022-3vq76

Brown, P. N., Lund, J. A., & Murch, S. J. (2017). A botanical, phytochemical and ethnomedicinal review of the genus *Mitragyna korth*: Implications for products sold as kratom. *Journal of Ethnopharmacology*, *202*, 302–325. https://doi.org/10.1016/j.jep.2017.03.020

Cavalier-Smith, T. (2010). Deep phylogeny, ancestral groups and the four ages of life. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1537), 111–132. https://doi.org/10.1098/rstb.2009.0161

Chen, M. M., Zhang, M., Liang, Z. S., & He, Q. L. (2022). Characterization and comparative analysis of chloroplast genomes in five *Uncaria* species endemic to China. *International Journal of Molecular Sciences*, *23*(19), 11617. https://doi.org/10.3390/ijms231911617

Di Fiore, S., Li, Q., Leech, M. J., Schuster, F., Emans, N., Fischer, R., & Schillberg, S. (2002). Targeting tryptophan decarboxylase to selected subcellular compartments of tobacco plants affects enzyme stability and in vivo function and leads to a lesion-mimic phenotype. *Plant Physiology*, *129*(3), 1160–1169. https://doi.org/10.1104/pp.010889

Eliot, A. C., & Kirsch, J. F. (2004). Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations. *Annual Review of Biochemistry*, *73*(1), 383–415. https://doi.org/10.1146/annurev.biochem.73.011303.074021

Feng, L. Y., Battulga, A., Han, E., Chung, H., & Li, J. H. (2017). New psychoactive substances of natural origin: A brief review. *Journal of Food and Drug Analysis*, *25*(3), 461–471. https://doi.org/10.1016/j.jfda.2017.04.001

Gabriel, M. W., Matsui, G. Y., Friedman, R., & Lovell, C. R. (2014). Optimization of Multilocus sequence analysis for identification of species in the genus Vibrio. *Applied and Environmental*

Microbiology, *80*(17), 5359–5365. https://doi.org/10.1128/aem.01206-14

Gamage, D. G., Gunaratne, A., Periyannan, G. R., & Russell, T. G. (2019). Applicability of Instability Index for in vitro protein stability prediction. *Protein and Peptide Letters*, *26*(5), 339–347. https://doi.org/10.2174/0929866526666190228144219

Glenn, W. S., Nims, E., & O'Connor, S. E. (2011). Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *Journal of the American Chemical Society*, *133*(48), 19346–19349. https://doi.org/10.1021/ja2089348

Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*, *30*(5), 1229–1235. https://doi.org/10.1093/molbev/mst012

Hassan, Z., Muzaimi, M., Navaratnam, V., Yusoff, N. H. M., Suhaimi, F. W., Vadivelu, R., Vicknasingam, B. K., Amato, D., von Hörsten, S., Ismail, N. I. W., Jayabalan, N., Hazim, A. I., Mansor, S. M., & Müller, C. P. (2013). From Kratom to mitragynine and its derivatives: Physiological and behavioural effects related to use, abuse, and addiction. *Neuroscience and Biobehavioral Reviews*, *37*(2), 138–151. https://doi.org/10.1016/j.neubiorev.2012.11.012

Han, S. W., & Shin, J. S. (2022). Aromatic L-amino acid decarboxylases: mechanistic features and microbial applications. *Applied Microbiology and Biotechnology*, *106*(12), 4445-4458. https://doi.org/10.1007/s00253-022-12028-4

Hirasawa, N. (2019). Expression of histidine decarboxylase and its roles in inflammation. *International journal of molecular sciences*, *20*(2), 376. https://doi.org/10.3390/ijms20020376

Khan, A. R., Mokthar, N. I., Zainuddin, Z., & Samsulrizal, N. H. (2021). In silico Characterization of UGT74G1 Protein in *Stevia rebaudiana* Bertoni Accession MS007. *Journal of Tropical Life Science*, *11*(3), 323–330. https://doi.org/10.11594/jtls.11.03.09

Madden, T. (2003). The BLAST sequence analysis tool. The NCBI handbook.

McDonald, A. D., Perkins, L. J., & Buller, A. R. (2019). Facile in vitro biocatalytic production of diverse tryptamines. *Chembiochem: A European Journal of Chemical Biology*, *20*(15), 1939–1944. https://doi.org/10.1002/cbic.201900069

Meireles, V., Rosado, T., Barroso, M., Soares, S., Gonçalves, J., Luís, Â., Caramelo, D., Simão, A. Y., Fernández, N., Duarte, A. P., & Gallardo, E. (2019). Mitragyna speciosa: Clinical, toxicological aspects and analysis in biological and non-biological samples. *Medicines (Basel, Switzerland)*, *6*(1), 35. https://doi.org/10.3390/medicines6010035

Mohanta, T. K., Kamran, M. S., Omar, M., Anwar, W., & Choi, G. S. (2022). PlantMWpIDB: a database for the molecular weight and isoelectric points of the plant proteomes. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-11077-z

Mozzarelli, A., & Bettati, S. (2006). Exploring the pyridoxal 5’-phosphate-dependent enzymes. *Chemical Record (New York, N.Y.)*, *6*(5), 275–287. https://doi.org/10.1002/tcr.20094

Osorio, D., Rondón-Villarreal, P., & Torres, R. (2015). Peptides: a package for data mining of antimicrobial peptides. *Small*, *12*, 44-444.

Pais, F. S. M., Ruy, P. D. C., Oliveira, G., & Coimbra, R. S. (2014). Assessing the efficiency of multiple sequence alignment programs. Algorithms for molecular biology, 9, 1-8. https://doi.org/10.1186/1748-7188-9-4

Prokić, J., Wieling, M., & Nerbonne, J. (2009). Multiple sequence alignments in linguistics. *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education - LaTeCH-SHELT&R ’09*, 18–25. https://doi.org/10.3115/1642049.1642052

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L. L., Eddy, S. R., Bateman, A., & Finn, R. D. (2012). The Pfam protein families database. *Nucleic Acids Research*, *40*(D1), D290–D301. https://doi.org/10.1093/nar/gkr1065

Shah, N., Altschul, S. F., & Pop, M. (2018). Outlier detection in BLAST hits. *Algorithms for Molecular Biology: AMB*, *13*(1). https://doi.org/10.1186/s13015-018-0126-3

Soltis, P. S., & Soltis, D. E. (2003). Applying the bootstrap in phylogeny reconstruction. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *18*(2), 256–267. https://doi.org/10.1214/ss/1063994980

Takayama, H. (2004). Chemistry and pharmacology of analgesic indole alkaloids from the rubiaceous plant, Mitragyna speciosa. *Chemical & Pharmaceutical Bulletin*, *52*(8), 916–928. https://doi.org/10.1248/cpb.52.916

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, *28*(10), https://doi.org/2731-2739. 10.1093/molbev/msr121

Wang, H., Zhong, H., Gao, C., Zang, J., & Yang, D. (2021). The distinct properties of the consecutive disordered regions inside or outside protein domains and their functional significance. *International Journal of Molecular Sciences*, *22*(19), 10677. https://doi.org/10.3390/ijms221910677

Wang, Y., Zhang, H., Zhong, H., & Xue, Z. (2021). Protein domain identification methods and online resources. *Computational and*

Structural Biotechnology Journal, *19*, 1145–1153. https://doi.org/10.1016/j.csbj.2021.01.041

Yim, W. C., & Cushman, J. C. (2017). Divide and Conquer (DC) BLAST: fast and easy BLAST execution within HPC environments. *PeerJ*, *5*(e3486), e3486. https://doi.org/10.7717/peerj.3486

Zhang, Y., Zhang, J. W., Yang, Y., & Li, X. N. (2019). Structural and comparative analysis of the complete chloroplast genome of a mangrove plant: *Scyphiphora hydrophyllacea* Gaertn. F. And related Rubiaceae species. *Forests*, *10*(11), 1000. https://doi.org/10.3390/f10111000

Zhang, M., Sharma, A., León, F., Avery, B., Kjelgren, R., McCurdy, C. R., & Pearson, B. J. (2020). Effects of Nutrient Fertility on Growth and Alkaloidal Content in *Mitragyna speciosa* (Kratom). *Frontiers in plant science*, *11*, 597696. https://doi.org/10.3389/fpls.2020.597696

Zhou, Y., Wang, Y., Shi, X., & Mao, S. (2018). The complete chloroplast genome sequence of *Campylandra chinensis* (Liliaceae). *Mitochondrial DNA. Part B, Resources*, *3*(2), 780–781. https://doi.org/10.1080/23802359.2018.1491346

Received: 18 September 2024
Accepted: 27 November 2024