**Evolution of Caffeine Biosynthetic Enzymes**

Caffeine (1,3,7-trimethylxanthine) is an allelopathic chemical, synthesized by ~ 100 plant species (Ashihara, 2006). It is synthesized from xanthosine by the action of a series of methyl transferases. Functionally it forms part of the producing plant’s defense system (it suppresses pathogens) and also inhibits germination/growth of seeds (this allows the plant to get first access to water and nutrients). Common plant sources of caffeine for human consumption include the coffee, tea, cacao, and mate. However, grapes, citrus, rose and other plants also make caffeine. In this exercise we will explore if the enzymes involved in biosynthesis on caffeine are the same in all these plants.

We will begin by learning about the biosynthesis on caffeine, identify some of the enzymes involved.

A. Start by learning about the metabolic steps in caffeine synthesis by exploring the KEGG pathways (<https://www.genome.jp/kegg/pathway.html>). Type caffeine in the top search box and find the diagram showing caffeine metabolism (https://www.genome.jp/dbget-bin/www\_bget?pathway:map00232)

Q1. What are the key steps in caffeine biosynthesis?

* Xanthosine converted to 7-methylxanthosine by the enzyme 7-methylxanthosine synthase
* Convert 7-methylxanthosine to 7-methylxanthine by the enzyme 7-methylxanthosine nucleosidase
* Convert 7-methylxanthine to theobromine and caffeine by the enzyme caffeine synthase with 1N- and 3N- methylation activity.

2. Let’s start with the amino acid sequence of the Xanthosine methyl transferase UniProt ID XMT1\_COFCA. Get the FASTA sequence and run protein BLAST as follows:

a. In one window – go to the blastp suite (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=>). Copy and paste the FASTA sequence of the protein and start BLAST search with default options.

Q2. What are the top 10 matches to the query? Take a screenshot of these from the results page and paste it here. What do you notice about these results?

A screenshot of a social media post

Description automatically generated

All these matches are coffee enzymes and the % identity is also very high (>90%).

b. In another window – again open the blastp suite (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=>). Copy and paste the sequence of the protein and start BLAST search with an option to exclude all Coffea sequences.

Q3. What are the top 10 matches to the query? Take a screenshot of these from the results page and paste it here. What do you notice about these results?

A screenshot of a cell phone

Description automatically generated

These are all proteins from different types of plants but their % identity of sequences is in the ~48-50%

Q4. Scan through the list of non-coffee plants that have the query or a related enzyme. List three different plants that show up here.

Tobacco, Eucalyptus, habanero peppers, hemp, grapes, and valley oak.

Note: Tea and Cacao enzymes do not show up in these sequence alignments. Does this mean that the enzymes that make caffeine in tea, coffee, and cacao plants share little or no sequence homology?

B. Select and save the FASTA sequences of 5 coffee proteins from the first BLAST results and any 5 sequences from the second BLAST run (each from a different plant source). Save the FASTA sequences of these proteins in the file and generate a tree. Take a screenshot of the tree and paste it here. What does the tree tell you about these sequences?

Open Mega-X >> Click on Align >> Edit/Build Alignment >> Retrieve sequences from file >> OK >> Provide file name sequences.fasta

Once the sequences are loaded click on the MUSCLE icon to align the sequences or click on Alignment >> Align by MUSCLE.

Data >> Export alignment >> MEGA format >> save file as sequences.meg

Models >> Find best models >> sequences.meg >> OK >> run the calculations and save output file

Click on the Phylogeny button to generate trees and take a screenshot of the trees generated.

A screenshot of a cell phone

Description automatically generated

Here are some observations from the tree:

1. All the 5 Coffee enzyme sequences are all closely aligned. Of these the arabica, and canephora sequences are more similar compared to the liberica and pseudoxanguebariae sequences.

2. The Capsicum and Nicotiana enzyme sequences belong to the same family and are closely related to each other and closer in evolutionary relationship to the coffee enzymes.

3. The other 3 non-coffee enzyme sequences are more different from the coffee enzymes.

C. Tea, coffee, and cocoa plants all make caffeine. Here we will explore how the caffeine synthesizing enzymes in these plants compare with each other. Are the genes synthesizing these proteins related or have they evolved convergently?

To get started we have identified two coffee, one tea, and one cocoa enzyme for this exercise. You will need the protein and gene sequences of these enzymes for analysis.

*Coffee enzymes*:

7-methylxanthosine synthase 1 🡪 UNP - A4GE69 (XMT1\_COFCA)

3,7-dimethylxanthine N-methyltransferase 🡪 UNP - A4GE70 (DXMT1\_COFCA)

*Tea enzyme*:

Caffeine synthase 1 – UNP - Q9FZN8 (TCS1\_CAMSI)

*Cocoa enzyme*:

S-adenosyl-L-methionine-dependent methyltransferases superfamily protein 🡪 UNP - A0A061FMF5 (A0A061FMF5\_THECC)

Q6. Download the FASTA sequences of the above proteins from UniProt and align them - you can run the sequence alignment using the NCBI/UniProt alignment tools or the MEGA-X software.

The sequence alignment and phylogeny tree for these sequences were computed following the same steps as in the above question to yield the following:

A screenshot of a social media post

Description automatically generated

Note that the Tea and cocoa enzymes are closer in sequence relationship and distantly related to the coffee enzymes.

Q7. Download the FASTA sequences of the genes of these enzymes and compare them to see if you can identify any rearrangements in specific regions of the genes to indicate evolutionary relationships.

Go to the UniProt pages for each of these protein sequences and from the “Sequences databases” section (under Cross-references) click on the GenBank radio button and then the link on the gene sequence. This should open the NGBI GenBank page with the relevant gene sequence. Copy the sequence and paste it in a text file. Name this file seq1.faa. Repeat this process to download the fast sequences of genes all 4 proteins.

Open Mauve and click on File >> Align with Progressive Mauve. In the new window that opens click on Add sequences and upload each of the 4 gene sequences that you have saved. Now run the alignment by clicking on the Align button in this window. Give the output file name when asked and let the software run the alignment. A new will open and list the alignment results. Once the run is complete an alignment viewer box will open and display the alignment. Select the options as follows:

‘View-> Color Scheme’ and select ‘LCB Color’.

Again, click on ‘View -> Style’ and select the following:

LCB Outlines

Similarity Plots

LCB Strikethrough Lines

LCB Connecting Lines

Show Mouse Highlighting

Draw Histograms

Save the alignment that you see below. What does this tell you about the caffeine synthesis genes from Coffee, Tea, and Cocoa?

A screenshot of a social media post

Description automatically generated

Only the 2 coffee enzymes are aligned. The tea and cocoa enzymes do not align with the coffee genes nor do they align with each other.

The coffee genes have some conserved regions (shown in the darker colors), but different gaps in the sequences.

Closing: Read about the convergent evolution of caffeine synthesis proteins in plants in the review article Huang et al., 2016.

References:

* Ashihara H. Metabolism of alkaloids in coffee plants. Braz J Plant Physiol. 2006; 18: 1-8.
* Huang, R., O’Donnell, A.J., Barboline, J.J., and Barkman, T.J. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymesPNAS September 20, 2016 113 (38) 10613-10618; <https://doi.org/10.1073/pnas.1602575113>