ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 12, Issue, 01(A), pp. 40445-40450, January, 2021

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

INSILICO ANALYSIS OF CYP82E GENE INVOLVED IN DEMETHYLATION OF NICOTINE IN TOBACCO

Saroja T[1]., T. Samuel David Raj[1]., Prabhakara Rao K[2] and Balaji Chandra Mouli J[3]*

[1]Dr.V.S. Krishna College, Andhra University, Visakhapatnam
[2]Division of Crop improvement, ICAR- Central Tobacco Research Institute, Rajahmundry
[3]Department of Biotechnology, Adikavi Nannaya University, Rajahmundry

DOI: http://dx.doi.org/10.24327/ijrsr.2021.1201.5697

ARTICLE INFO

Article History:

Received 6th October, 2020
Received in revised form 15th November, 2020
Accepted 12th December, 2020
Published online 28th January, 2021

Key Words:

CYP genes, TSNA regulation, bioinformatics tools, *in silico* analysis and content manipulation.

ABSTRACT

To identify CYP gene variants involved in the regulation of TSNA, members of CYP family encoding nucleotide sequence of demethylase gene were used as query to search the online databases. Based on the number of hits, percentile of similarity and E value, candidate CYP genes were identified. Further, authentication was done from Solgenome network database harbouring sequenced genomes of Solanaceae members. The identified sequences were scrutinized using online open source tools viz., prosite, pfam and Expassy. The 29 CYP genes/ CYP variants related to TSNA formation from five sequenced genomes of *Nicotiana* species were identified in this study. These sequences can be utilized for *in silico/ in vivo* expression analysis to understand the underlying molecular mechanism of TSNA regulation and subsequent manipulation to reduce their content in tobacco.

INTRODUCTION

TSNAs are pyridine alkaloids formed during secondary metabolite production. These nitrosamines are formed from nicotine and related compounds by nitrosation reaction that occurs during curing and processing of tobacco. The major alkaloid in tobacco plant is nicotine and it constitutes 90% of total plant alkaloid production. It is a precursor for tobacco specific nitrosamines by converting into secondary alkaloid – "nor nicotine" by the action of an enzyme N-demethylase (NND). Based on the rate of conversation of nicotine to nor nicotine, plants are discriminated as converters and non-converters. The varieties of tobacco convert a high portion of their nicotine content to nor nicotine are referred as "converters" and the process of conversion mainly occurs during senescence and curing. Those that remain nicotine as a predominant alkaloid are considered to be "non converters".

Nitrosamines in tobacco are Nitrosonornicotine (NNN), nicotine derived nitrosamine ketone (NNK), N-nitrosoanatabine (NAT), N-nitosoanabasine (NAB) and 4-(methylnitosamino-1-(3-pyridyl)-1-butanol (NNAL). In addition to these compounds some more TSNAs are identified and assessed in tobacco and tobacco smoke. They are 4-(methylenitrosamino)-4-(3pyridyl)-butyric acid (iso-NNAC), 4-(methylnirosamino)-4-(3-pyridyl)-1-butanol (iso-NNAL) and 4 (methylnirosamino)-1-(3-pyridyl)-1-butanol (NNAL). Among all of these nitrosoamines N-nitrasonornicotine (NNN) is more carcinogenic in nature proven experimentally. It causes several types of cancer in animals (Hetch and Haffman 1988, 1989).

Nornicotine is synthesized from nicotine by the action of an enzyme. NND has been studied through the characterization of nornicotine. The gene which codes for NND enzyme belongs to cytochrome P450 family and super family monooxygenase which catalyzes demethylation of nicotine (Siminsky *et. al*., (2006); Xu al (2007).

CYP82E4 (NND) comes under CYP82E2 family, plays a key role in production of nornicotine through demethylation process of nicotine. CYP82E2 family is very small and has minimum number of genes which shows 90% nucleotide identity among them. Tobacco has totally 5 types of demethylase genes, which are inherited from *N.tomentosiformis* and *N.sylvestris* to modern day tobacco. In which three of them CYP82E3 and CYP82E4 (senescence specific) and CYP82E5V2 (weakly expressed) gene originated from *N.tomentosiformis*. The other two genes, were CYP82E2

Corresponding author:* **Balaji Chandra Mouli J
Department of Biotechnology, Adikavi Nannaya University, Rajahmundry

(strongly expressed senescence - specific) and CYP82E10 (weakly expressed) originated from *N.sylvestris*. In modern day tobacco, CYP82E4 accounts for the majority of nicotine conversion observed in senescent and cured leaves (Gavilano *et al* 2006 Siminisky *et al* 2005 Xu *et al* 2007, Chakrabarthi *et al* 2008) whereas CYP82E5 mediates a low level of nicotine conversion primarily in the green leaves of tobacco (Gavilano and Siminsky 2007*)* and CYP82E10 catalyzes low level of nicotine at root level (Lewis *et al* 2010)

RESULT AND DISCUSSION

In plants, the cytochrome P450 (CYP) super family is the largest enzymatic protein family and has wide presence across the kingdom. They play vital role in promoting plants growth and development and involved in multiple metabolic pathways (McKinnon *et al*., 2008). Among the various members of the CYP family CYP82 family genes are proven to be involved in coding the demethylase enzyme (Gavilano and Siminszky, 2007).

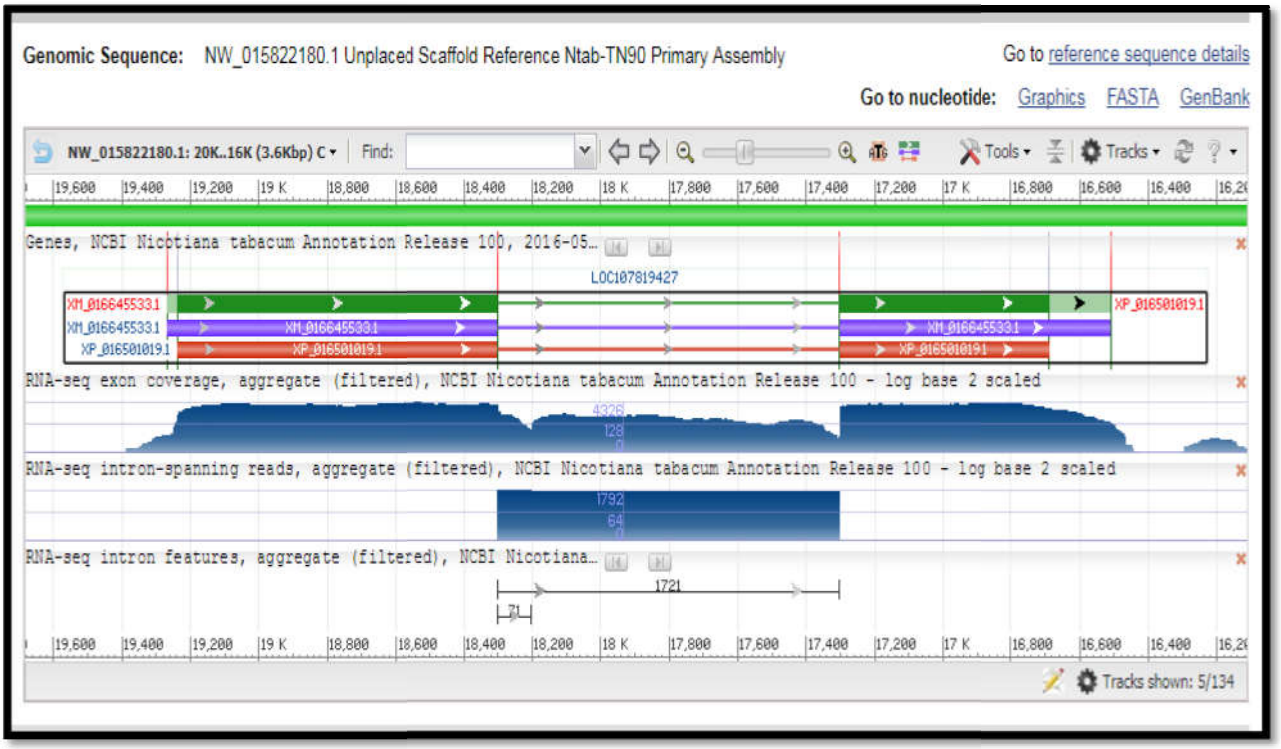

Figure 1 NCBI genome data viewer – structure of CYP82E4 gene –length of gene: 2769bps; Exon-1- 615bps; Exon-2 -933bps; length of mRNA: 1761bps; length of CDS -1554bps; length of protein sequence – 517a.a

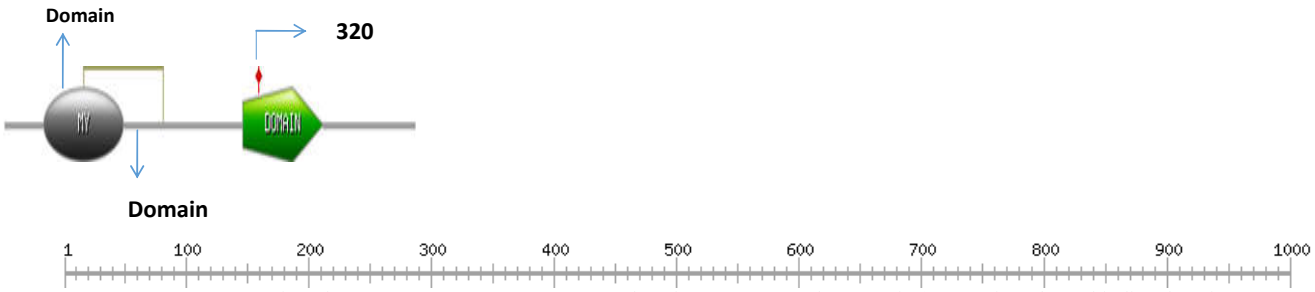

Figure 2 Domains of CYP82E4; ACTIVE SITE -320 – domain -1 (Transmembrane region), Domain 3- heme binding domain

There are numerous computational strategies that can be employed for *insilico* gene analysis. Many of these are based on utilization of available online sequences which are submitted to the gene databases by biologist. The present study reviews the usage of different online sources and tools like Pfam, multialine etc., to investigate and identify the different domains among the sequences. To understand the gene structure and assess exon and intron length GENE VIEWER OF NCBI was employed.

In an attempt to retrieve the CYP82 genes from the sequenced Nicotiana genomes a total of twenty nine 29 genes were identified after vigilant curation. Maximum number of CYP variants is from N. tabacum because, it is a natural hybrid of N.tomentosiformis and N.sylvestris and its amphidiploid nature. Majority of the sequence retrieved have direct or indirect role in regulation of TSNA.CYPs are considered as hemoproteins due to heme cofactors. The heme group plays crucial role in the pathway of catalysis and there exists the signature sequence FxxGxxxCxG in the heme-binding domain, functioning as the fifth ligand to the heme iron. The identified sequences possess heme binding domain at specified positions

along with the transmembrane region, coiled coil region which supports them as members of CYP family.

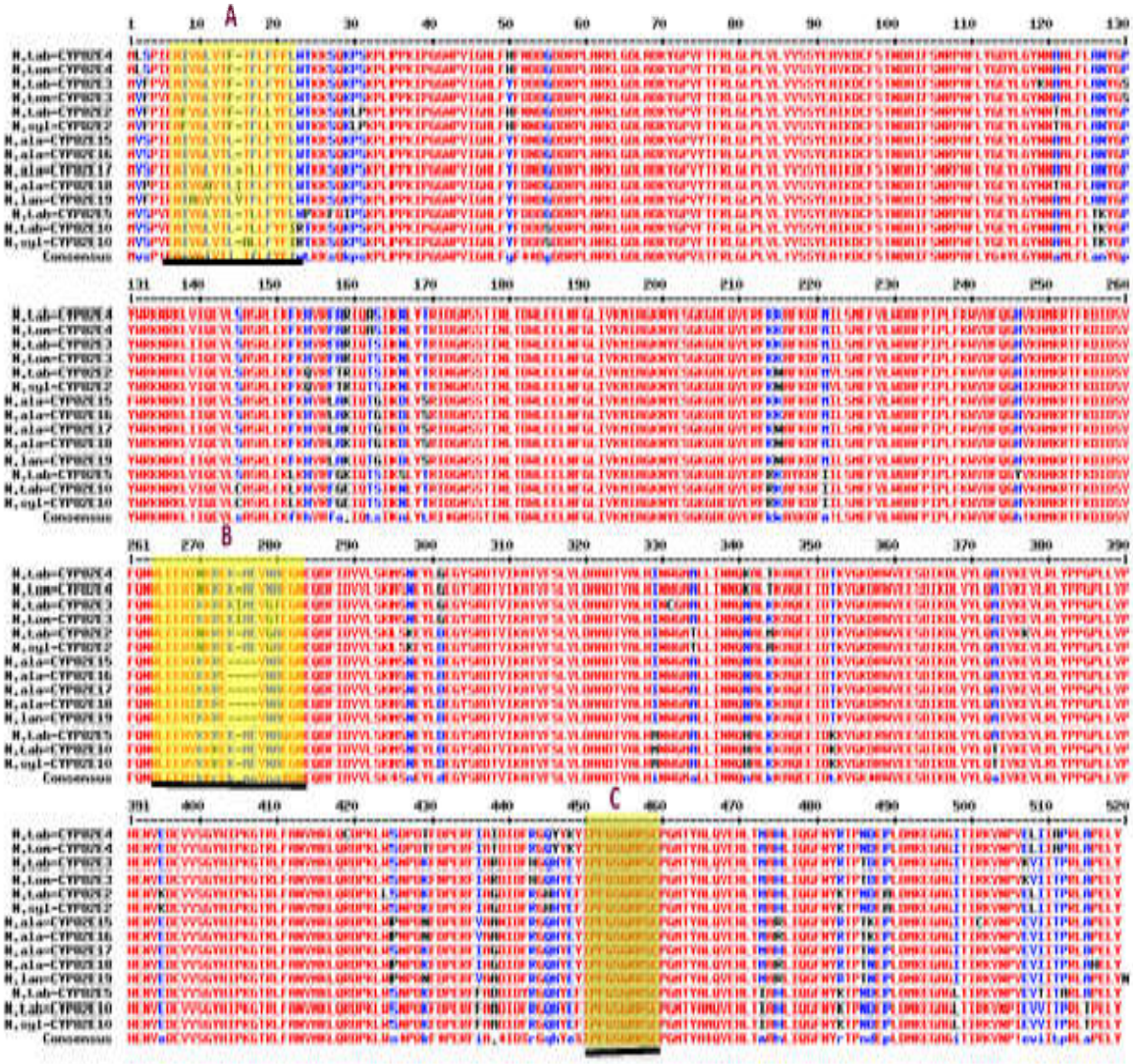

Figure 3 Multiple protein sequence alignment of CP82E variants of Nicotiana species and domain identification. A Domain:- Transmembrane region (6-23aa), B Domain: Coiled coil region (264-284 aa),C Domain: Heme binding domain (450-459 aa).Heme-binding domain (FGSGRRSCPG)

Sequence analysis revealed that all the selective sequences have two exons in each sequence scattered through the length. The length of the first exon is around 930 bp and the length of the second exon is around 630 bp. These were separated by introns of varied length in each sequence.CYP82E2 and CYP82E3 genes have mutations at key positions leading to their inactiveness. In the phylogenic analysis the branching pattern in all the five clusters is in accordance with the respective subfamily of CYP82E genes regardless of the species. The genes from the CYP82E2 and CYP82E4 subfamily shared a common branching point.

In conclusion, 29 CYP NND genes were identified from five species of Nicotiana and they were classified in to five families of CYP genes. With the help of Scan prosite, pfam, GENOME DATA VIEWER genes were analyzed very efficiently and found two exons and one intron among the gene sequences. The three important domains Transmembrane region (6-23aa), Coiled coil region (264-284 aa), Heme binding domain (450-459 aa) at protein level of CYP genes.

The insilico analysis of these sequences may be used to manipulate the invivo expression of TSNA, to regulate their levels in tobacco which reduces its carcinogenic nature to overcome several types of cancers.

MATERIALS AND METHODS

All the selected CYP82E (Demethylase) sequences of tobacco were collected from National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov) by using the putative CYP82E4 sequence as a query for basic alignment search tool (BLAST) with expected E 0.0 value. All the sequences which showed 92-100% similarity were collected and examined using hidden Markov models. All the pseudo and partial sequences which showed more than E- 0.0 Value were exempted for the analysis.

Domain Identification

All the putative CYP82E protein sequences of tobacco were collected and aligned using Multialin (http://multalin.toulouse.

inra.fr/) online tool which aligns the sequences with hierarchical clustering & default parameters.

allows visualizing trees in ASN (text and binary), Newick and Nexus formats.

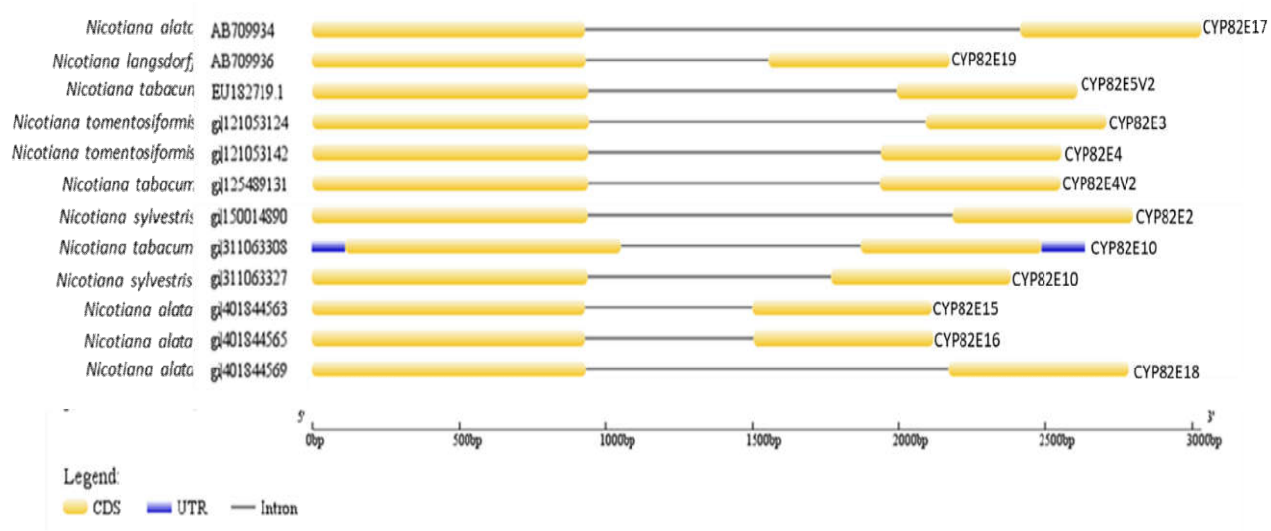

Figure 4 Nucleotide sequence Organization of CYP82E genes.

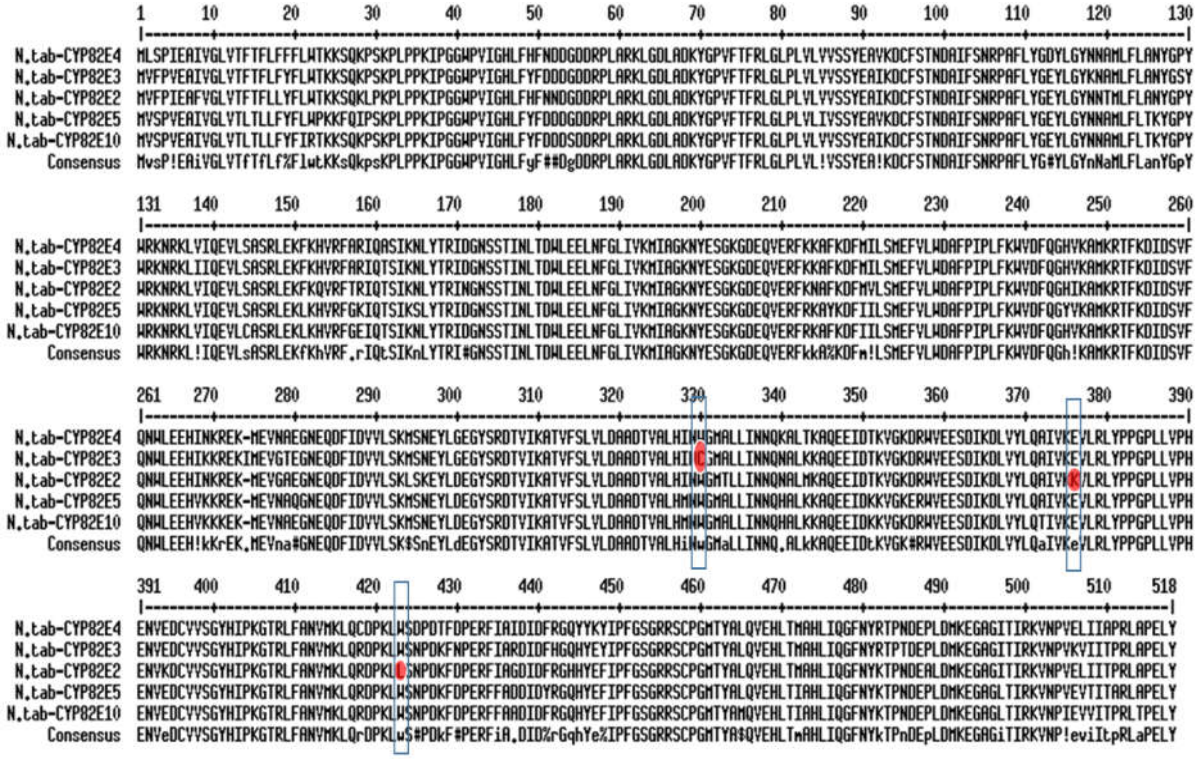

Figure 5 Amino acid sequence alignment of CYP82E genes from *N.tabaccum.*

Important domains of $CYP8_2E4$ were identified by using online free tool scan prosite (https://prosite.expasy.org).

Phylogenic analysis

The assumed nucleotide sequences of CYP82E were aligned in Newick format using phylogeny.fr format converter (http://phylogeny.lirmm.fr) and the format was subjected to NCBI Tree viewer NIH (http://www.ncbi.nlm.nih.gov/projects/treeview/) to create circular phylogenetic tree of $CYP8_2E$ sequences to visualize the evolutionary relationship between them. NCBI Tree viewer NIH is the graphical display for phylogenetic trees. NCBI TV

Structure of $CYP8_2E4$

NCBI GENOME DATA VIEWER (https://www.ncbi.nlm.nih) was used to understand the structure of CYP82E4 sequence. NCBI Genome data viewer is the online tool which clearly visualizes the exon and intron stretches of gene.

Structure of $CYP8_2E4$ Gene

To comprehend the structure of CYP82E4, online available source of NCBI GENOME DATA VIEWER was used, in analysis part two exons and one intron were identified, 2769bps complete length of the gene sequence was recognized and this whole sequence was parted by 615bps and 933 length bits of exons and these two were departed by 998bps length of intron.

By using available online data expression profiles 1761 bps length of the mRNA sequence was acknowledged and 1554bps length coding sequence was identified.

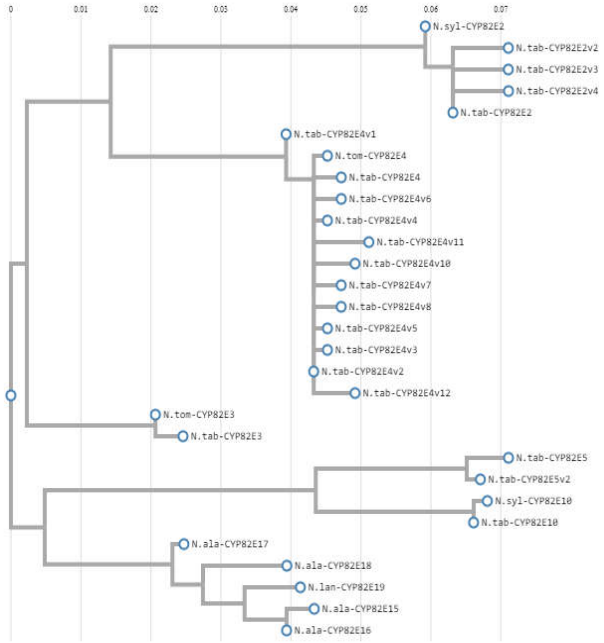

Figure 6 Cladogram and clustering of CYP82E genes from *Nicotiana species.*

Identification of CYP82E homologues

The genome sequencing of *Nicotiana* species unraveled the sequence information and the raw sequences along with partly annotated regions, available in public domain. Regulation of TSNA is crucial in identification of CYP gene variants. CYP82E family members encodes demethylase gene were used as query to search the online databases. It was resulted in more than 100 sequences from different databases. Further, these sequences were manually curetted based on the hit score, percentile similarity and E value in case of BLAST. Candidate CYP genes were identified and further authenticated from Solgenome network database to eliminate the redundancy. After thorough manual curation, it was finally narrowed to twenty nine (29) CYP82E genes presented in the table.

Domain organization of CYP82E homologues

To become a member of CYP82 family, a protein sequence must qualify with the presence of specific domain organization. To check the presence of these domains, all the identified CYP82E genes were translated to respective protein sequences using translate tool. The respective ORFs are aligned by multialign program for detecting the physical presence of domain. Cythochrome p450 dependent monooxygenases are large group of heme containing enzymes which consists of Transmembrane region (6-23aa), coiled coil region (264-284 aa), heme binding domain (450-459 aa) with specific sequence (FGSGRRSCPG). The same has been analyzed by scanprosite as well. As the result reveals that (Figure 2) the identified sequences possess transmembrane region, coiled coil region and heme binding domain at specified positions. In the process of curation the sequences with incomplete/absence of part of domains were rejected.

Sequence characterization of CYP82E homologues

The nucleotide sequence of retrieved CYP82E genes were vary in length from aprox.1500 bp to 3200 bp. In order to characterize the sequences, a total of 12 selective sequentially numbered CPY82E genes from CYP82E2 to CYP82E19 were analyzed using NCBI tools. It reveals that all the selective sequences contain two exons which are separated by varied length of intronic regions. The respective lengths of the two exons are 930 bp and 630 bp (Figure 4). These exon lengths are conserved across the sequences of various CYP82E genes. The intron length is more in CYP2E17 and smaller in CYP82E15 & CYP82E16 with respect to CYP82E10. There is a UTR region at 5' and 3' ends in CYP82E15. These introns aid in general absorption of mutations. The CDS region in many of the CYP82E genes stands around 1554 bps.

Further, the five CYP82E translated sequences of *N tabaccum* were aligned for amino acid changes. In the sequence alignment the key amino acid changes of glutamic acid (E) mutated to lysine (K) at 375 position and tryptophan (W) mutated and changed as leucine (L) at 422 position was observed in CYP82E2 and one amino acid replacement at 330 position was noticed in CYP82E3 apart from all other replacements.

Phylogenic Analysis

Based on the alignment of full length sequences of CYP gene, phylogenetic tree (Figure 5.) was constructed to determine the evolutionary relationship of CYP families in detail. The phylogenetic tree illustrated that CYP genes were classified into five families named as CYP82E4, CYP82E3, CYP82E2, CYP82E10 and CYP82E 15-19, which was consistent with alignment results of whole collected CYP sequences from different species of *Nicotiana*. According to the phylogenetic tree, the families of CYP82E4, CYP82E3 and CYP82E2 are in same branching point from CYP ancestors and remaining CYP82E5 and CYP82E10 are in different branching point, whereas CYP82E 11-19 was found on different branching point from the ancestors. These CYP 15-19 sequences are from *Nicotiana alata* and *Nicotiana langsdorffi.*

Funding

We the Authors individually declaring that there is no sources of funding received for the research work submitted to this journal**.**

Acknowledgement

Authors express their sincere gratitude to Dr. D. Damodara Reddy, Director, CTRI, Rajamundry, for giving support to work in CTRI, Rajahmundry, East Godavari District, Andhra Pradesh, India.

Disclosures

Authors declares that there is no other disclosers for the research work submitted to the journal

Conflict of Interest:- Non declared

References

Gavilano, Lily B., and Balazs Siminszky. 2007. "Isolation and characterization of the cytochrome P450 gene CYP82E5v2 that mediates nicotine to nornicotine

conversion in the green leaves of tobacco."*Plant and Cell Physiology* 48 (11): 1567-1574.

Gavilano, Lily B., Nicholas P. Coleman, Leigh Emma Burnley, Melissa L. Bowman, Newton E. Kalengamaliro, Alec Hayes, Lowell Bush, and Balazs Siminszky. 2006. "Genetic engineering of Nicotiana tabacum for reduced nornicotine content."*Journal of Agricultural and Food Chemistry*.

Gavilano, Lily B., Nicholas P. Coleman, Steven W. Bowen, and Balazs Siminszky. 2007. "Functional analysis of nicotine demethylase genes reveals insights into the evolution of modern tobacco."*Journal of Biological Chemistry* 282 (1): 249-256.

Lewis, Ramsey S., Steven W. Bowen, Matthew R. Keogh, and Ralph E. Dewey. 2010. "Three nicotine demethylase genes mediate nornicotine biosynthesis in Nicotiana tabacum L.: Functional characterization of the CYP82E10 gene."*Phytochemistry* (Elsevier Ltd) 71 (17-18): 1988-1998.

Ross A McKinnon, Michael J Sorich, Michael B Ward.2008. "Cytochrome P450 Part 1: Multiplicity and Function".Journal of Pharmacy Practice and Research ,Volume 38, No. 1:55-7.

Saitoh, Fumiyo, Masana Noma, and Nobumaro Kawashima. 1985. "The alkaloid contents of sixty Nicotiana species."*Phytochemistry* (Pergamon) 24 (3): 477-480.

Siminszky, B, L Gavilano, S W Bowen, and R E Dewey. 2005. "Conversion of nicotine to nornicotine in Nicotiana tabacum is mediated by CYP82E4, a cytochrome P450 monooxygenase."*Proceedings of the National Academy of Sciences of the United States of America* 102 (41): 14919-14924.

Siminszky, B., L. Gavilano, S. W. Bowen, and R. E. Dewey. 2005. "Conversion of nicotine to nornicotine in Nicotiana tabacum is mediated by CYP82E4, a cytochrome P450 monooxygenase."*Proceedings of the National Academy of Sciences.*

How to cite this article:

Saroja T *et al*.2021, Insilico Analysis of CYP82E Gene Involved in Demethylation of Nicotine In Tobacco. *Int J Recent Sci Res.* 12(01), pp. 40445-40450. DOI: http://dx.doi.org/10.24327/ijrsr.2021.1201.5697

****