IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.7, Special Issue.3, Feb. 2019 E-ISSN: 2347-2693

EST Based Studies For Identification Of Curcumin Synthase Gene In *Curcuma caesia* Roxb.

Neha Behar[1*], K. L. Tiwari[2], S. K. Jadhav [3]

[1]Dept. of Biotechnology, D.L.S. P.G. College, Bilaspur (C.G.) India
[2,3]S.o.S. in Biotechnology, Pt. Ravi Shankar Shukla University Raipur (C.G.) India

[*]*Corresponding Author: neha1_biotech@yahoo.com, Tel.: +91-9039-857167*

Available online at: www.ijcseonline.org

***Abstract*:** *Curcuma caesia* Roxb. (Kali Haldi) is commonly used as folklore medicine in Chhattisgarh state. *Curcuma* species contains curcumin, which is responsible for its pharmacological activities. This compound is subject of interest due to its potential use in prevention and treatment of cancer, HIV Infection and Alzheimer's disease. This plant-specific research has always been hampered due to limited resources but with the development of turmeric EST database by David Gang's group, the molecular and functional analysis is now possible in different curcuma species. In this study, the EST regions of the major gene CURS (Curcumin synthase) from previously reported expressed sequence tag sequences (NCBI accession numbers DY384950 and DY386934) in *C. longa* were used, for studies in *C. caesia*, the amplicon generated of CURS EST sequence in *C. caesia* (DY384950) was sequenced using cycle sequencer. The sequence was then analysed for similarity by using BLASTn in NCBI. The expressed sequence showed 98% similarity with the CURS1 of *C. longa*. Identification of candidate genes using functional genomics and bioinformatics tools can significantly contribute to understanding the biosynthesis of curcumin in *Curcuma caesia*.

Keywords—Curcumin, Curcumin synthase, database, Expressed Sequence Tags, gene

I. INTRODUCTION

Plants have always been the valuable source of pharmacologically important compounds and are used for treatment of many diseases. However, plant-specific research has been hampered due to limited resources. The conventional research involved manual screening for identifying and studying medicinal plants but this is not sufficient to keep the pace with the pharmaceutical industry needs. Bioinformatics approaches may provide an essential set of tools for designing efficient and targeted searches for plant-based remedies [1]. For effective and applied research, modern and innovative practices and tools should be introduced from time to time. The genome based studies of medicinal plants lag behind those of model plants and important crop plants. This type of study is important as genome sequences contains many essential information from its origin, to evolution, development, physiology, inheritable traits and secondary metabolites content and the gene involved for its biosynthesis and diversity at the molecular level [2]. But, limited number of plants has whole-genome sequence data. The majority of genomics resources for plants have come from ESTs (Expressed Sequence Tags). ESTs are partial sequences from cDNA, representing the genes expressed in a particular tissue of a particular organism**.** ESTs may be used to identify gene transcripts, and are thus, instrumental in gene discovery, gene sequence determination and in identifying putative genes and pathway involved in secondary metabolite biosynthesis in medicinal plants. One such important family of medicinal plant is Zingiberaceae, it accumulates pharmacologically important curcuminoids in their rhizomes. Curcumin demethoxycurcumin, bisdemethoxycurcumin are chemically bis-α-β unsaturated β-diketones and are collectively called as curcuminoids. The most important of these compounds and the most intensively studied by far is curcumin, which has been shown to also possess the remarkable activities of preventing or treating Alzheimer's disease, immunomodulation, and correcting cystic fibrosis defects. Turmeric has been used as major food additive and as a drug in various traditional system of medicine from ages, but the plant-specific research has always been hampered due to limited resources, the development of turmeric EST database by David Gang's group will help for molecular and functional analysis in curcuma species and for studying the curcumin biosynthetic pathway in *Curcuma longa* and other lesser known species of the genus. *Curcuma caesia* Roxb. commonly known as "Kali Haldi" is an important medicinal plant used as traditional folklore medicine in Chhattisgarh state. In Ayurveda system of medicine, it is commonly used for anti-inflammation and anti-asthma. The difference in the

bioactive compounds in plants and different species of same genus is suggested not only by the availability of the substrates for the enzymes, but also by the expression levels of the genes encoding these enzymes [3]. There are reports that chalcone synthase (CHS) superfamily of plant type III polyketide synthases (PKSs) catalyze iterative decarboxylative condensations of malonyl unit with a CoA-linked starter molecule to produce structurally diverse pharmaceutically important plant secondary metabolites [4]. Katsuyama *et al*., (2009a) [5] revealed the curcumin biosynthetic route in turmeric, in which DCS (Dikektide co enzyme A) synthesizes feruloyldiketide-CoA, and CURS (Curcumin synthase) then converts the diketide-CoA esters into a curcuminoid scaffold. DCS catalyzed the formation of feruloyldiketide-CoA by condensing feruloyl-CoA and malonyl-CoA. CURS [4] and CURS2 [3] showed similar substrate specificity for feruloyl-CoA as a starter substrate while CURS3 [3] preferred both p-coumaroyl-CoA and feruloyl-CoA as a starter substrate. Thus, this study involved The EST regions of the major gene CURS (Curcumin synthase) from previously reported expressed sequence tag sequences (NCBI accession numbers DY384950 and DY386934) in *C. longa*, the amplicon generated of CURS EST sequence (DY384950) in *C. caesia* was sequenced using cycle sequencer. The sequence was then analysed for similarity by using BLASTn in NCBI.

II. METHODOLOGY

The CURS gene sequences were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/). Structural and functional characteristics of gene were studied. EST (Expressed Sequence Tags) of CURS (Curcumin Synthase) from previously reported expressed sequence tag sequences (NCBI accession numbers DY384950 and DY386934) were used. The amplicon generated of CURS EST sequence (DY384950) in *C. caesia* was sequenced using cycle sequencer, and analysed for sequence similarity by using BLASTn in NCBI. BLASTN programs searched nucleotide databases using a nucleotide query, we have searched database nucleotide collection (nr/nt) using mega blast which optimize for highly similar sequences.

III. RESULTS AND DISCUSSION

The Aromatic Rhizome Expressed Sequence Tags (ArREST) database is produced from 8 cDNA libraries from high quality RNA extracted from the rhizome, leaf, and root from both ginger and turmeric species. After sequencing random clones from these libraries, a total of 50,408 ESTs have been deposited in the database with over 12,500 ESTs, 6382 from rhizomes in *Curcuma longa* (http://ag.arizona.edu/research/ganglab/ArREST.htm) [6]. Thus *in silico* survey was done to mine ESTs sequences for CURS and NCBI accession numbers DY384950 and DY386934 EST sequences were used for studies. It is already reported that primer designing of the genes involved in curcuminoids synthesis, RNA extraction from rhizome and leaf of the plant, cDNA preparation and semi-quantitative expression studies of genes involved in different tissue of *Curcuma caesia* Roxb. were done [7]. The amplicon thus generated of CURS EST sequence (DY384950) was sequenced using cycle sequencer, and analysed for sequence similarity by using BLASTn in NCBI. The sequence showed 98% similarity with the CURS1 gene of *C. longa* and 96% with CIPKS10 of *C. longa*, suggesting difference in the sequences of the gene of *C. longa* and *C. caesia*. **Expressed Sequence Tag (EST)** is a short stretch of DNA sequence that is used to identify an expressed gene, this is generally sufficient to identify the full-length complementary DNA (cDNA). ESTs are generated by sequencing a single segment of random clones from a cDNA library. Now, the majority of the sequences in sequence databases are ESTs. The sequencing and analysis of expressed sequence tags (ESTs) has been a primary tool for the discovery of novel genes in plants, especially in non-model plants for which full genome sequences are not currently available [8]. EST sequencing represents a rapid and cost-effective method for analyzing the transcribed regions of genomes. EST analysis is also a powerful tool for the discovery of genes involved in plant secondary metabolism [9]. EST sequence analysis can be done by using sequence comparison algorithms to search protein and DNA. One implementation that is widely used to speed up similarity searching is the related sequences or "neighbour" service of the National Centre for Biotechnology Information's (NCBI) Entrez system [http://www.ncbi.nlm.nih.gov]. The neighbour service is dependent on the fast, heuristic protein and DNA similarity algorithm in the BLAST program [10]. Mining and characterization of EST derived microsatellites in *Curcuma longa* L. has been reported [11]. And authors stated that recently EST-derived SSRs (Simple sequence repeats) are a free by-product of the currently expanding EST (Expressed Sequence Tag) databases and these detected SSRs can be greatly used for designing primers that can be used as markers for constructing saturated genetic maps and conducting comparative genomic studies in different *Curcuma* species. Sheeja *et al*., 2015, [12] investigated the genes involved in curcuminoid biosynthesis by using Illumina sequencing platform and generated a substantial amount of expressed sequence tag (EST) dataset from two species viz., *C. longa* and its wild relative *C. aromatica* Salisb. contrasting in curcumin content. The transcripts were assembled by BLAST similarity searches and annotated with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology identifiers and concluded that these EST sequences can be valuable public information platform for functional studies in turmeric and form a rich resource for studies on marker development and turmeric breeding. In our study, expressed sequence showed 98% similarity with the CURS1 and 96%

with CIPKS10 gene of *C. longa*. Identification of candidate genes involved in the curcumin biosynthetic pathway will significantly contribute to understanding the biosynthesis of active compound curcumin in *Curcuma caesia*.

BLAST reports showing Sequences producing significant alignments with CURS gene of *Curcuma caesia.*

Sequences producing significant alignments:						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**% Identity**	**Accession**
***Curcuma longa* chalcone synthase-like protein (ClPKS10) mRNA, complete cds**	156	156	83%	4e-35	96%	JN017185.1
***Curcuma longa* CURS1 mRNA for curcumin synthase, complete cds**	148	148	74%	7e-33	98%	AB495007.1

IV CONCLUSION AND FUTURE SCOPE

Identification of candidate genes using functional genomics and bioinformatics tools can significantly contribute to understanding the biosynthesis of curcumin in *Curcuma caesia*. In our study, expressed sequence of *C. caesia* showed 98% similarity with the CURS1 and 96% with CIPKS10 gene of *C. longa*. Identification of candidate genes involved in the curcumin biosynthetic pathway will significantly contribute to understanding the biosynthesis of active compound curcumin in *Curcuma caesia*.

REFERENCES

[1] V. Sharma, I. N. Sarkar, "*Bioinformatics Opportunities for Identification And Study Of Medicinal Plants"*, Briefings in Bioinformatics, vol. 14. pp. 238 -250, 2012.

[2] D. C. Hao, P. G. Xiao (2015), "*Genomics and Evolution in Traditional Medicinal Plants: Road to a Healthier Life*", Evolutionary Bioinformatics, vol. 11, pp. 197–212, 2015.

[3] Y. Katsuyama, T. Kita, S. Horinouchi, "*Identification and Characterization of Multiple Curcumin Synthases from the Herb Curcuma Longa"*, FEBS Letters, vol. 583, pp. 2799–2803, 2009b.

[4] I. Abe, I. H. Morita, "*Structure And Function Of The Chalcone Synthase Superfamily Of Plant Type III Polyketide Synthases"*, Natural Product Report, vol. 27, pp. 809–838, 2010.

[5] Y. Katsuyama, T. Kita, N. Funa, S. Horinouchi, "*Curcuminoid Biosynthesis by Two Type III Polyketide Synthases in the Herb Curcuma longa"*, Biological Chemistry, vol. 284, pp.11160–11170, 2009a.

[6] H. J. Koo *et al*., "*Ginger and Turmeric Expressed Sequence Tags Identify Signature Genes for Rhizome Identity and Development and the Biosynthesis of Curcuminoids, Gingerols And Terpenoids"*. BMC Plant Biology, vol. 13, pp. 1471-2229, 2013.

[7] Neha Behar, K.L. Tiwari, S.K. Jadhav, "*Semi Quanitative Expression Studies Of Gene Involved In Biosynthesis Of Curcuminoids In Curcuma Caesia Roxb*. Indian journal of biotechnology", vol. 15, pp. 491-494, 2016.

[8] J. Parkinson, "*Expressed Sequence Tags (ESTs) Generation and Analysis"*, Humana Press, 2009.

[9] Y. Li *et al*., "*EST Analysis Reveals Putative Genes Involved In Glycyrrhizin Biosynthesis"*. BMC Genomics, vol. 11, pp.1471-2164, 2010.

[10] S. F. Altschul, T.L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J., "*Gapped BLAST and PSI-BLAST: a new generation of protein database search programs"*, Nucleic Acids Research, vol. 25, pp. 389-402, 1997.

[11] R. K. Joshi, A. Kuanar, S Mohanty, E. Subudhi, S. Nayak, "*Mining and characterization of EST derived microsatellites in Curcuma longa L". Bioinformation*, vol. 5, issue, 3, pp. 128-131, 2010.

[12] T.E. Sheeja, K. Deepa, R. Santhi, B. Sasikumar, "*Comparative Transcriptome Analysis of Two Species of Curcuma Contrasting in a High-Value Compound Curcumin: Insights into Genetic Basis and Regulation of Biosynthesis*", Plant Molecular Biology Reports, vol.33, issue 6, pp. 1825–1836 2015.

