In silico Survey of Genes involved in Curcuminoid Synthesis for Expression Studies in *Curcuma caesia* Roxb.

Neha Behar
PhD Scholar
SOS in Biotechnology
Pt Ravi Shankar Shukla University
Raipur (C.G.)

K.L Tiwari
Professor and Head
Department of Biotechnology
Guru GhasiDas Vishwavidhyalaya
Bilaspur (C.G.)

S.K Jadhav
Professor and Head
SOS in Biotechnology
Pt Ravi Shankar Shukla University
Raipur (C.G.)
Corresponding author

ABSTRACT

Curcuma caesia Roxb. (Black turmeric) is an endangered medicinal plant of the state. The rhizome of this plant has been traditionally used for centuries as a folklore remedy. The main bioactive substances in the rhizomes are curcuminoids, despite the tremendous importance of this compound, molecular and functional analyses of its medicinal value were hampered by lack of tools such as ESTs and ordered genomic contigs. But with the development of turmeric EST database by David Gang's group (ArREST), it has provided a platform for elucidating the curcuminoid biosynthetic pathway in *Curcuma longa*. All the gene sequences involved in curcuminoid synthesis were retrieved from NCBI. EST and CDS based primers were designed by using online Batch Primer3 software. Designed primers would be further validated in wet lab, by amplifying these primers with cDNA of *Curcuma caesia* for putative expression analysis.

General Terms

Medicinal plant research, Database, Primer designing

Keywords

Curcuma caesia Roxb., curcuminoid, Genes, Polyketide synthases

1. INTRODUCTION

India has rich repository of medicinal plants, which have been the source of medicine for centuries. Our state Chhattisgarh is known as "Herbal state" due to presence of wide varieties of medicinal plants. But the genome information of medicinal plants in databases is far behind than other crop plants. In medicinal plants, for identification of functional genes and enzymes that govern the biosynthetic pathway of secondary metabolites, improvement and authentication of traditional knowledge and development of new methods based on genomics, proteomics and metabolomics, many informations are needed at a common podium. Here bioinformatics plays a major role, as one of the main tasks of bioinformatics is the management and analysis of large volume of biological data. Curcuma caesia Roxb. commonly known as "Black turmeric" is an endangered medicinal plant of the state. The rhizome of this plant has been traditionally used for centuries as a folklore remedy. The main bioactive substances in the rhizomes are curcuminoids, which are widely distributed in plants. Curcumin, demethoxycurcumin and bisdemethoxycurcumin are collectively called as curcuminoids. Curumin is chemically bis-α-β unsaturated β-diketones and has been widely used for its anti-inflammatory, antiangiogenic, antioxidant, wound- healing and anti-cancer effects (Jayprakash *et al*., 2006).Turmeric or Haldi has been worldwide used as a major spice, food and as a drug, but despite the tremendous importance of this compound, molecular and functional analyses of its medicinal value are hampered by lack of tools such as ESTs and ordered genomic contigs. However things have been changed recently due to the development of turmeric EST database by David Gang's group (ArREST) which has provided a platform for elucidating the curcuminoid biosynthetic pathway in Curcuma longa (Katsuyama *et al*., 2009a) and points toward the role of Type III Polyketide synthases (PKSs) in the biosynthesis of curcumin. The chalcone synthase (CHS) superfamily of type III polyketide synthases (PKSs) generates backbones of a variety of plant secondary metabolites including chalcones, stilbenes, phloroglucinols, resorcinols, benzophenones, biphenyls, bibenzyls, chromones, acridones, pyrones, and curcuminoids (Abe and Morita, 2010). Since the first isolation of CHS gene from parsley (Petroselium hortense) in 1983, more than 900 type III PKS genes have been reported in public databases (NCBI: http://www.ncbi.nlm.gov/), and more than 20 functionally different plant and bacterial CHS-superfamilies of type III PKSs have been cloned and characterized (Abe and Morita, 2010). Thus present study involved *in silico* survey of genes reported in different plants involved in curcuminoids synthesis and designing of primers by on-line primer designing tool for studying putative expression in different tissues in Curcuma caesia Roxb.

2 MATERIALS AND METHODS

All the gene sequences were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/). Structural and functional characteristics of genes were compiled and studied. Protein Knowledgebase (UniProt KB) was used to know gene name, origin, protein attributes, general annotations and molecular functions. EST (Expressed Sequence Tags) and CDS (Coding DNA Sequences) based primers were designed by using Batch Primer3, a high-throughput web tool for picking PCR primers (http://probes.pw.usda.gov/cgibin/batchprimer3/batchprimer3.cgi). Primer designing was done by putting the nucleotide sequences in FASTA format and the specifications for primer designing included 18-22 bp of primer length, Tm range of 50-60 °C and 100-300 bp of product size. All other options were left on default value.

3. RESULTS AND DISCUSSIONS

The Chalcone synthase superfamily of Type III PKS is a conserved homodimeric protein (monomer size, 40–45 kDa) responsible for the biosynthesis of a variety of plant polyketides (Austin and Noel, 2003). Katsuyama *et al*., (2007) reported the first type III polyketide synthase encoded by os07g17010 from a cDNA library of Oryza sativa named CUS (Curcuminoid Synthase). Then katsuyama *et al*., (2009a), proposed a pathway for curcuminoid biosynthesis in the herb Curcuma longa, which includes two novel types III polyketide synthases - named Diketide-CoA Synthase (DCS) and Curcumin synthase (CURS). Multiple curcumin synthases from the *Curcuma longa* were identified and characterized by katsuyama *et al*., 2009b and demonstrated that the herb contains at least two other type III PKSs, CURS2 and CURS3 that are also capable of curcuminoid biosynthesis, in addition to the previously identified CURS. The existence of three different type III PKSs in curcuma. longa suggests that the composition of the curcuminoids in the rhizomes of turmeric is affected not only by the availability of the substrates but also by the expression levels of the genes encoding these enzymes. Resmi and Soniya (2012) reported two new Type III PKS, ClPKS9 and ClPKS10 using homology based RT-PCR and data mining. ClPKS9 sequence showed similarity to chalcone synthase whereas ClPKS10 to curcuminoid synthases. Recently CHSI (Chalcone synthase like gene) and CHSII have been found in curcuma caesia specifically and their partial sequence has been deposited by Remakanthan *et al*., (2010) to the databases, thus gene annotations are available in databases. In silico analysis of the genes and characterizing them structurally and functionally will help in having better knowledge of genes and so metabolite biosynthesis. All the genes are Chalcone/ stilbene synthase family and selected on the basis of its role in biosynthetic pathway and chances of it being express in curcuma caesia. Curcuma caesia is non-conventional turmeric and less work is done in it due to limited availability while Curcuma longa the common haldi, used as spice in food has been greatly worked upon thus maximum genes are taken from it. Primers were designed by online tool. The sequences of the primers can have a major effect on the PCR amplification. Primers were selected by following parameters shown in Table 1.Primers with long runs GC were avoided. The Tm difference between Forward and reverse primer were about 2°C and the primers with The 3' T-overhangs were avoided. Keeping these parameters in mind primer was picked and their further validation would be done in wet lab, by amplifying these primers in cDNA of *Curcuma caesia* for putative expression analysis.

Table 1: List of genes and their sequence characteristics

Gene name	Locus ID	Protein name	Organism	GO ID	Gene size (bP)	Protein sequence status	Protein sequence length	Protein Mass (Da)	Pfam hit
CUS	LOC_Os07g 17010	Curcuminoid synthase	Oryza sativa	GO:0016747	1598	complete	402 AA	43,212	PF02797 Chal_sti_synt_C. 1hit PF00195 Chal_sti_synt_N. 1 hit.
DCS	BAH56225	Diketide CoA synthase	*Curcuma longa*	GO:0016747	1170	complete	389 AA	42,047	PF02797 PF00195
CURS	BAH56226	Curcumin synthase	*Curcuma longa*	GO:0016747	1170	complete	389AA	43,034	PF02797 PF00195
CURS 2	BAH85780	Curcumin synthase	*Curcuma longa*	GO:0016747	1176	complete	391AA	43,146	PF02797 PF00195
CURS 3	BAH85781	Curcumin synthase	*Curcuma longa*	GO:0016747	1173	complete	390 AA	43,099	PF02797 PF00195
ClPKS9	JN017186	Chalcone synthase-like protein	*Curcuma longa*	GO:0016210	1392	complete	396 AA	43,333	PF02797 PF00195
ClPKS10	JN017185	Chalcone synthase-like protein	*Curcuma longa*	GO:0016210	1452	complete	389 AA	43,313	PF02797 PF00195
CHS	HM161842	Chalcone synthase	*Curcuma longa*	GO:0016746	568	fragment	189 AA	20,283	PF02797 PF00195
CHS 1	HM161811	Chalcone synthase	*Curcuma caesia*	GO:0016746	568	fragment	189 AA	20,335	PF02797 PF00195
CHS 2	HM161812	Chalcone synthase	*Curcuma caesia*	GO:0016746	568	fragment	189 AA	20,154	PF02797 PF00195

Table 2: **criteria followed for primer designing**

Criteria	Optimum	Range
Length of target sequence to be amplified	250bp	100-300bp
Tm	60^0C	55-64^0C
GC content	55%	50-60%
Length of primer	20bp	18-20bp

Table 3: List of primers designed from EST and CDS regions

Name of gene	EST /CDS Sequences ID	Primers	Product size (bP)
CURS	DY384950	F- ATAACACCCCTCCTCCTCTC	155
		R- AAGTGCTCTGCTCGTAAAGG	
	DY386934	F- ATCACCCACCTGGTATTCTG	118
		R- CCTGGCTGTAGATCATGAGG	
DCS	BAH56225	F- AGGTCACCGTGCTCTCCTAC	232
		R- CTCTTGAGGTGGAAGGTCAG	
CURS2	DY393763	F- CCAATCTCTACGAGCAGGAC	129
		R- AAAGATACCGCCTCTTCACC	
	DY387045	F- CTCGTCCATCACGAAGTAGA	164
		R- ACTGGAACGACATCTTCTGG	
CURS3	DY394591	F- TCTGTGAGAAGACGAAGGTG	136
		R- TTTGGTATCTCCTCCACCAC	
CIPKS9	JN017186	F- CACATGTACTTGACGGAGGA	245
		R- AGCCCGAGAAGCTTAGTGAG	
CIPKS10	JN017185	F-TACTTCCTGAACCAGCTTCC	248
		R-TCGTCCATCACGAAGTACAC	
CHS1	HM161811	F-AGGTCAACACGCTCATCTTC	269
		R-CGTCTCCAAGTTGTTGGCTA	

4. References

[1] Jayaprakasha, G. K., Rao, L. J., Sakariah, K.K. 2006 Antioxidant activities of curcumin, demethoxycurcumin and bisdemethoxy curcumin. Food Chem 98, 720-724.

[2] Katsuyama, Y., Kita, T., Funa, N. and Horinouchi, S. 2009a. Curcuminoid Biosynthesis by Two Type III Polyketide Synthases in the Herb Curcuma longa. Biol Chem. 284, 11160–11170.

[3] Abe, I. and Morita, H. 2010. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. Nat. Prod. Rep, 27, 809–838.

[4] Austin, M.B. and Noel, J.P. 2003. The chalcone synthase superfamily of type IIIpolyketide synthases. Nat. Prod. Rep. 20, 79–110.

[5] Katsuyama, Y., Matsuzawa, M., Funa, N. and Horinouchi, S.2007. In vitro synthesis of curcuminoids by type III polyketide synthase from Oryza sativa. J Biol Chem. 282, 37702-37709.

[6] Katsuyama, Y.,Kita, T. and Horinouchi, S. 2009b. Identificationand characterization of multiple curcumin synthases from the herb Curcuma longa. FEBS Letters, 583, 2799–2803.

[7] Resmi, M.S. and Soniya, E.V. 2012.Molecular cloning and differential expressions of two cDNA encoding Type III polyketide synthase in different tissues of Curcuma longa.491, 278-283.

[8] Remakanthan, A., Gincy, P.T. and Soniya, E.V. 2010 *Curcuma caesia* chalcone synthase like (CHS 1) gene, partial sequence Submitted (APR-2010) to theEMBL/GenBank/DDBJdatabases.Unpublished