RESEARCH PAPER *Asian Journal of Bio Science* (October, 2010) Vol. 5 Issue 2 : 200-205

Promoter analysis of an unknown gene in *Populus trichocarpa* x *Populus deltoides*, exposed to *Malacosoma disstria* Hubner

MAMTA SAGAR
Department of Bioinformatics, University Institute of Engineering and Technology, Chhatrapati Shahuji Maharaj University, KANPUR (U.P.) INDIA

(Accepted : June, 2010)

Several duplicated genes are present in populus genome and have multiple roles in transporting metabolite and synthesis of lignin and cellulose. Duplicated chs genes in *Populus* genome have been identified to be involved in defense mechanism due to production of flavonoid derived compound with pathogens interaction and sequences of the genes are available in the databases. In this work, one unknown gene (Accession No. EF148325,GI:118488982), expressed due to continuous feeding by *Malacosoma disstria* Hubner (forest tent caterpillar) in *Populus trichocarpa* x *Populus deltoids*, was analyzed by pattern searching and sequence comparison methods. Regulatory sequence analysis tool was used to identify regulatory site using sequence motifs of Arabidopsis *thaliana* and *Zea maize*. Using pattern matching approach the consensus pattern was predicted in Populus *hybrid*, to characterize the function of gene. DNA pattern search program was used to predict the patterns CNGTTR, CANNTG, AC like element, GCCTACC, GCCTACC, ACCTACA, ACCTAAC in upstream sequences of 7 chs genes of *P. trichocarpa* and one gene (EF148325) of *P. Trichocarpa* x *P. deltoides*. CNGTTR, CANNTG, known as b HLH pattern flanked by AC like element, have been predicted in the gene EF14325, which is the promoter for MYB134 regulatory protein which regulate phenylpropanoid metabolism and is involved in activating defenses in herbaceous plants. The gene sequence is compared with chs gene sequences of *Populus trichocarpa*, β-ketoacylcoa synthase from *Populus trichocarpa*; Acyltransferase from *Ricinus communis*; fiddlehead-like protein from *Pisum sativum*, were close homologs which share 97%, 83%, 77% identity, respectively with protein encoded by target gene and share same product binding site for coumaryl-CoA and malonyl-CoA, a critical step of flavonoid biosynthesis pathway.

Key words : *Populus trichocarpa*, *P. trichocarpa* x *P.deltoides*, Proanthocyanidine, Chalcone synthase, Flavonoid, *Malacosoma disstria*, EF148325

INTRODUCTION

Genome duplications followed by sequence divergence can create new gene in addition to gene with original function. *Populus* genome has duplicated genes that might be particularly useful for protection from pathogens. These genes are involved in synthesizing lignin and cellulose, transporting metabolites, and bringing about programmed cell death. Chalcone synthase (CHS), a key enzyme of plant flavonoid biosynthesis, is generally encoded by a multigene family. CHS plays multiple roles in the production of diverse flavonoid derived compounds under complex developmental and environmental regulation, and with pathogen interaction. *chs* genes have been cloned from a wide range of plant species. Studies have shown that specific patterns of expression of four *chalcone synthase* genes in a Hunnegem poplar clone have been identified *(Populus trichocarpa* x *Populus deltoides*) (Claire Lurin 1 and Lise Jouanin). Melampsora leaf rust, a destructive disease of poplar, causes premature leaf drop, loss of biomass and vigor, and even death in poplar plantations (Newcombe *et al.*, 1994). Some of the PR proteins have been detected in many species, which may be directly inhibitory to pathogens (Van Loon *et al.*, 2006).

In a comparative analysis with insect-induced defenses in leaves of the same hybrid poplar, Ralph *et al.* (2006) found that *M.medusae* infection down regulates many of the major herbivor-inducible defenses genes. *M.medusae* rust infection strongly induced genes for enzymes of flavonoid and proanthocyanidin (condensed tannin) biosynthesis late in the infection process. This study indicates that this pathway is closely related to the pathogen defense response in poplar**.** Myb and WRKY proteins are regulatory proteins involved in activating defenses in herbaceous plants (Eulgem, 2005). This study provided the fundamental basis to do the research analysis of expression pattern of an unknown gene in *P.*

Mamta Sagar (2010). Promoter analysis of an unknown gene in *Populus trichocarpa* x *Populus deltoides*, exposed to *Malacosoma disstria* Hubner, *Asian J. Bio. Sci.*, **5** (2) : 200-205

trichocarpa x *P. deltoides* during insect feeding of forest tent caterpillar. In this work, expression of one unknown gene is analyzed in poplar hybrid during the feeding of forest tent caterpillar, *Malacosoma disstria* Hubner. The gene is predicted using computational tool by identifying the transcription factor binding site. Sequence motifs have been taken from *Arabidopsis thaliana* and *Zea maize*. Using the comparative genomics approach and pattern recognition methods the consensus pattern is predicted in target gene of *Populus trichocarpa* x *P.deltoides*, known as b HLH pattern flanked by AC like element to characterize the function of gene. These patterns are also found in promoter of *Arabidopsis thaliana* indicating the biosynthesis of proanthocyanidine production in transpirant testa of seed mutants. Presence of proanthocyanin was observed to be associated with defense mechanism in herbaceous plants. The sequence of unknown gene of poplar hybrid has been compared with other sequences of chs genes of poplar genome to predict molecular phylogeny which aims to predict its role in flavonoid pathway.

MATERIALS AND METHODS

Nucleotide sequences of 7 chs genes of *T. trichocarpa* and one unknown gene (EF148325) of *T. trichocarpa* x *T. deltoides* were retrieved from NCBI ENTREZ Database.

Populus trichocarpa x *Populus deltoides* clone WS01313_D13 unknown mRNA, 1814 bp mRNA, ACCESSION EF148325, EF148325.1 GI:118488982. Source organism (1.1814) *Populus trichocarpa* x *Populus deltoides*, mol_type=" mRNA", cultivar= "H1111"db_xref="taxon:3695"clone= "WS01313_D13" tissue_type="Sapling trees one metre in height and grown under greenhouse conditions were exposed to continuous feeding by *Malacosoma disstria* Hubner (forest tent caterpillar) mid-instar larvae. Mature leaves from within the caged region were collected 4 hours, 8 hours and 24 hours after the onset of treatment. Mature leaves were also collected above the caged region (systemic response) 4 hours, 12 hours and 48 hours after the onset of treatment."(i) All the sequences were given as input in RSA (Regulatory analysis tool) tool http://rsat.ulb.ac.be/rsat/ (ii) DNA pattern search program was used to predict the motif in all 8 sequence, DNA patterns CNGTTR, CANNTG, AC like element, GCCTACC, GCCTACC, ACCTACA, ACCTAAC were given as a query pattern in the query box.(iii) All the motifs were searched on upstream sequence and direct strand of 8 genes (iv) CNNTG and AC like element pattern were searched separately also to predict the coexistence of both for the characterization of b HLH promoter.

Pattern analysis:

– Several patterns can be entered in the query box. All of these will be matched against the sequences. Patterns to be searched were entered in the query pattern(s) box. The first word of each line is the string description of the pattern; the second word is an identifier for this pattern. Patterns can contain letters from the degenerate nucleotide alphabet (*e.g*. N for "any nucleotide", W for "A or T".– All the sequences were entered in the sequence box; patterns were matched against all these. Leave all other parameters unchanged and click GO. Output appeared in the form of table. Positions of all matches with the patterns were observed within the upstream sequences of the selected genes. Each line showed a single match, and the different columns indicated, respectively. the pattern identifier; strand on which the match was found (D for direct, R for Reverse); pattern searched for (*i.e*. the query strings you provided); name of the sequence in which it was found; starting position of the match; end position of the match ;match sequence shown in Table 1. Query was processed to get the feature map (shown in Fig. 1-6)on the next page (i) Color boxes indicated the presence of pattern on a black line, accordingly to the strand of the match. (ii) Coordinates were provided with reference to the ORF starting position, negative values indicated an upstream position and positive coordinates were within the coding sequences. Complete information about pattern can be seen as on the status bar.

Sequence similarity search (i) Pairwise alignment was performed using program BLASTN 2.2.22 for target gene sequence.

Phylogenetic tree was generated by multiple sequence alignment using neighbour joining method of multiple sequence alignment.(ii)Program CLUSTAL 2.0.12 was used to generate multiple sequence alignment and phyolgenetic tree.

RESULTS AND DISCUSSION

Sequence motifs known as b HLH pattern flanked by AC like element to characterize the function of gene have been taken from *Arabidopsis thaliana* and *Zea mays*. These patterns were associated with the biosynthesis of proanthocyanidine production, observed in transpirant testa of seed mutants *Arabidopsis thaliana*.

Consensus pattern in the target gene of populus hybrid:

Using the comparative genomics approach and pattern recognition methods, the consensus pattern was predicted in unknown gene of *Populus trichocarpa* x *P.deltoides* Conserved motifs associated with proanthocyanin production were searched in all seven chs genes of *Populus trichocarpa* CNGTTR,CANNTG, AC like element, GCCTACC, GCCTACC, ACCTACA, ACCTAAC. Debeaujon *et al*. (2003) characterised the minimal Arabidopsis BAN promoter that matches the animal c-myb consensus (CNGTTR) and bHLH domain protein-binding consensus (CANNTG) as well as two AC element-like motifs. These are the crucial consensus patterns in sequences which play significant roles in falavonoid derived pathway by activating the genes involved in insect induced defences, leading to production of proanthocynadin. It further predicted CANNTG, 3 CNGTTR motifs with numerous AC elements like motif have been predicted by RSA tool in EF148325 gene of *Populus trichocarpa* x *Populus deltoides.*

BAN promoter is present in target gene:

Presence of CNGTTR, CANNTG and AC element showed that all the chs genes and one unknown gene taken for analysis contained BAN promoter and b HLH with AC motif (Fig. 1, 2 and 3). Sequence motifs for BAN promoter have been predicted in EF148325 sequence of *P.trichocarpa* x *P. deltoides* with the score value of one, position of CNGTTR and CANNTG are shown where regulatory protein b HLH can bind and regulate the expression of gene involved in defense mechanism (Table 1).

An important observation was the presence of various consensus patterns in the target gene sequence of *P. Trichocarpa* x *P. deltoides,* that were associated with the proanthocyanin biosynthesis in other plants (Arabidopsis and maize). It was hypothesized that the unknown gene was associated with proanthocyanin production.

Target gene was compared with all the seven chalcone synthase genes in order to observe if there was any chs like activity in the uncharacterized gene in poplar hybrid during the feeding of *Malacosoma disstria* Hubner. Similarity to this gene was observed in chs 1 gene only. ACCTAAC motif was phenylpropanoid biosynthetic promoter for MYB134 transcription factor, which was predicted only in A sequence (chs 1 of *Populus trichocarpa),* this motif was not found in all other gene sequences (chs 2-chs7) of *Populus trichocarpa* and EF148325 of *Populus trichocarpa* x *P. deltoides***.** Numerous AC like elements have been predicted in EF148325, although motif ACCTAAC was absent (Fig. 4).

All 8 sequences were searched for various patterns and the four motifs present were shown in different color; CNGTTR (blue), CANNTG (red), AC element (green) and ACCTAAC (pink). Seven chs genes of *Populus trichocarpa* and EF148325 of *Populus trichocarpa* x *P. deltoides* having the probability of one for these four motifs are in Fig. 6.

Homology search:

Homologous proteins of predicted gene are shown in (Table 2), beta-ketoacylcoa synthase (protein id-XM_002314624.1) family protein from *Populus trichocarpa* has 97% identity and acyltransferase (Protein id- XM_002516989.1) from *Ricinus communis*

Table 1: Sequence motif for BAN promoter have been detected inEF148325 sequence of *P.trichocarpa* X *P. deltoides* with the score value of one, position CNGTTR and CANNTG have been given in the third and fourth column, these promoter sequence have been taken from *Arabidopsis thaliana* (positions of AC element have not been shown here)

Matching positions	Pattern	Start	End	Matching sequence	score
START-END	-	-1816	-1	-	0.0
CNGTTR	CNGTTR	-1588	-1583	ctatCAGTTAaact	1.00
CNGTTR	CNGTTR	-613	-608	atggCTGTTGctgg	1.00
CNGTTR	CNGTTR	-569	-564	agggCCGTTAgtcc	1.00
CANNTG	CANNTG	-1630	-1625	gaacCAGCTGaaac	1.00
CANNTG	CANNTG	-899	-894	actgCAGGTGcatc	1.00
CANNTG	CANNTG	-807	-802	tctcCAACTGctta	1.00
CANNTG	CANNTG	-681	-676	gttaCAATTGtgtg	1.00
CANNTG	CANNTG	-383	-378	ctggCACATGgagc	1.00
CANNTG	CANNTG	-240	-235	gattCAAGTGtaac	1.00
CANNTG	CANNTG	-146	-141	tgttCATGTGccta	1.00

Fig. 1, 2, 3 : Feature map depicts the presence of CNGTTR, CANNTG and AC element like motif that shows the presence of BAN promoter and b HLH with AC motif, respectively

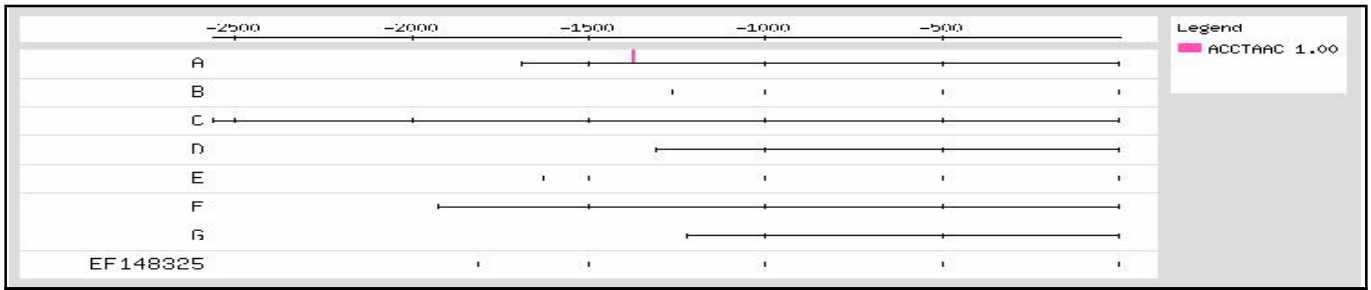

Fig. 4 : Feature map depicts the presence of ACCTAAC motif (phenylpropanoid biosynthetic promoter for MYB134 transcription factor), which is present only in A sequence (chs 1 of *Populus trichocarpa*), this motif is not found in all other genes (chs 2-chs7 of *Populus trichocarpa* and EF148325 of *Poplus trichocarpa* X *P. deltoides* having the probability of one.

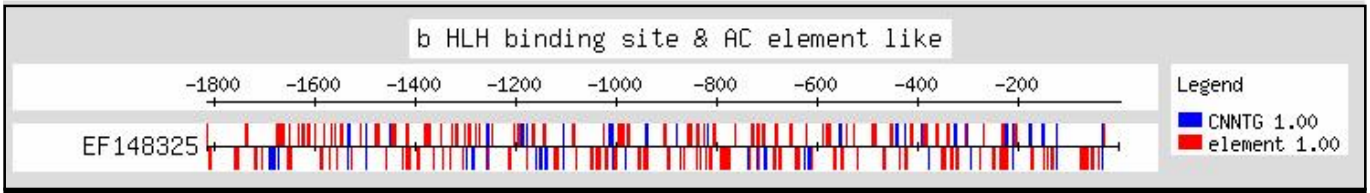

Fig. 5: Feature map shows b HLH binding site and AC element like site of predicted gene *Populus trichocarpa* x *P. deltoides*

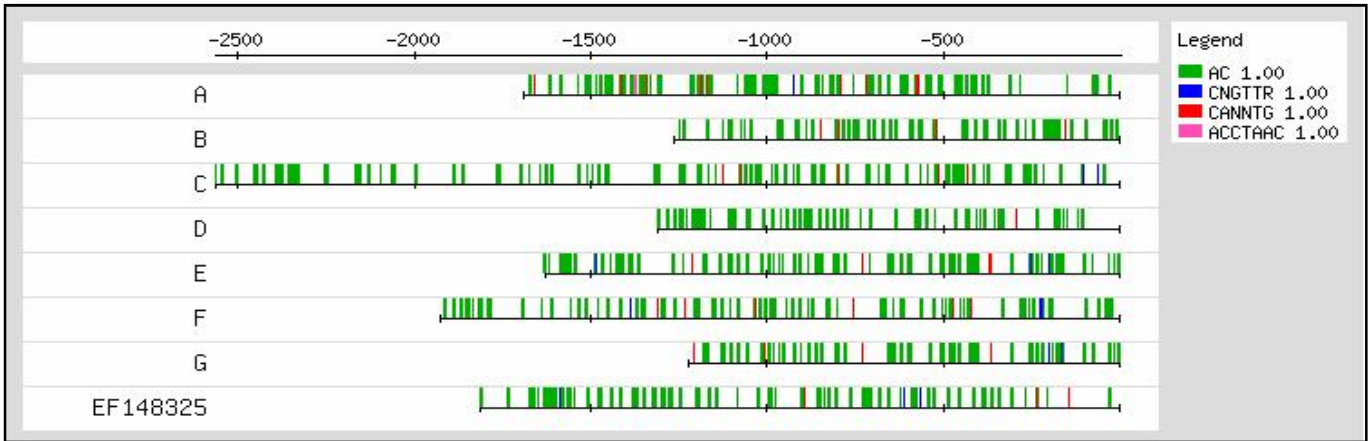

Fig. 6 : Shows presence of four motifs in different color ; blue, red, green, pink(CNGTTR, CANNTG,AC element and ACCTAAC) in all 8 sequences (seven chs gene of Populus trichocarpa and EF148325 of *Populus trichocarpa* x *P. Deltoides* having the probability of one

Table 2: Homologous proteins of EF148325 sequence (mRNA) were predicted using BLAST, closely related protein is beta-ketoacylcoa synthase family protein from *Populus trichocarpa*. Acyltransferase (*Ricinus communis*) and fiddlehead-like protein(*Pisum sativum*) are other homologs

Accession No	Description	Total score	Query coverage	E value	Max ident
EF148325.1	*Populus trichocarpa x Populus deltoides* clone WS01313_D13 unknown mRNA	3350	100%	0.0	100
XM_002314624.1	*Populus trichocarpa* beta-ketoacyl-coa synthase family protein, mRNA	2859	91%	0.0	97%
XM_002312526.1	*Populus trichocarpa* beta-ketoacyl-coa synthase family protein, mRNA	2025	87%	0.0	89%
AC217632.1	*Populus trichocarpa* clone POP013-F03, complete sequence	6814	98%	0.0	98%
XM_002516989.1	*Ricinus communis* acyltransferase, putative, mRNA	1415	82%	0.0	83%
XM_002332589.1	*Populus trichocarpa* beta-ketoacyl-coa synthase family protein, mRNA	809	68%	0.0	79%
XM_002516991.1	*Ricinus communis* acyltransferase, putative, mRNA	278	48%	1e-70	73%
EU041722.2	*Pisum sativum* fiddlehead-like protein mRNA, complete cds	110	10%	4e-20	77%

has 83% identity, fiddlehead-like protein (protein id-EU041722.2) *Pisum sativum* has 77% identity. Homology search for these three proteins showed that Chalcone and stilbene synthases; plant-specific polyketide synthases (PKS) and related enzymes, also called type III PKSs. PKS generates an array of different products, dependent on the nature of the starter molecule. These proteins are the similar proteins with common function and have the same product binding site indicating possible role of target gene (protein product) in chalcone synthase pathway. Acyl transferase from *Ricinus communis* (protein id-_002516991.1 condensing enzymes has identity 73%. Members shared strong structural similarity, which are involved in the synthesis and degradation of fatty acids.

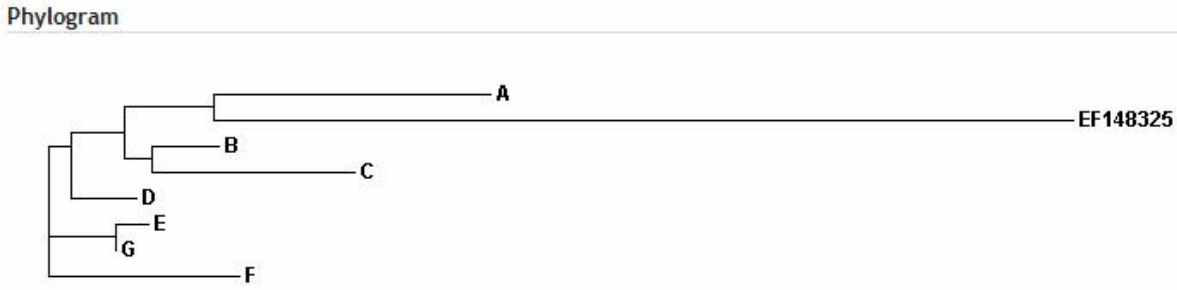

Fig. 7 : Phylogenetic tree depicts the 8 sequences including EF148325 which shows more similarity to sequence A (chs1), others sequences (chs2-chs7) show sequence divergence, suggesting that EF148325 would have been duplicated and diverged from chs1

Phylogenetic tree was generated by multiple sequence alignment using neighbour joining method, depicting the phylogenetic relationship among seven chs gene and one EF148325 (Fig. 7). Predicted gene showed more similarity to sequence A (chs1) than sequences of others (chs2-chs7).This shows that the target gene has been diverged earlier, suggesting that EF148325 is a paralogous gene, which would have been duplicated and diverged from chs 1.

Conclusion:

CNGTTR, CANNTG, AC like element have been successfully predicted in the unknown gene EF14325 of Poplar hybrid. Although GCCTACC(DFR-1), ACCTACA (ANR-2), ACCTAAC motifs are not present, yet presence of AC element may provide basis to engineer the ACCTAAC like motif which is phenylpropanoid biosynthetic promoter for MYB134 transcription factor. Promoter region of the predicted gene contains motifs similar to adenosine and cytosine rich AC element found in the region of biosynthetic gene of different branches of phenylpropanoid metabolism including both flavonoid and lignin biosynthesis. Motifs can be engineered using genetics engineering approach to produce the disease resistant variety of *Populus* hybrids by inducing the plant defence mechanism. Beta-ketoacylcoa synthase (protein id-XM_002314624.1) from *Populus trichocarpa*, acyltransferase (Protein id-XM_002516989.1) from *Ricinus communis* has fiddlehead-like protein (protein id-EU041722.2) *Pisum sativum*, share 97%;83% ; 77% identity, respectively with protein encoded by target gene EF148325 and share same product binding site for coumaryl-CoA and malonyl-CoA, a critical step of flavonoid biosynthesis pathway which produced chalcone followed by flavonone. This is the necessary reaction to synthesize the anthocyanine and proanthocyanidine. Phylogenetic tree showed that the gene has more similarity to sequence A (chs1), which suggests that EF148325 is a paralogous gene, and it would have been duplicated and diverged from chs 1.

Acknowledgement

Author is extremely grateful to Professor Dayal Saran, former Director, UIET, CSJM University, Kanpur for constant encouragement for publication and support as well as providing good computational facility at the Institute.

REFERENCES

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M. and Lepiniec, L. (2003) Proanthocyanidin -accumulating cells in Arabidopsistesta : regulation of differentiation and role in seed development. *Plant Cell.*, **15**: 2514–2531

Eulgem, T. (2005). Regulation of the *Arabidopsis* defense transcriptome. *Trends Plant Sci.*, **10**:71-78

Newcombe, G., Chastagner, G. A., Schuette, W. and Stanton, B. J. (1994). Mortality among hybrid poplar clones in a stool bed following leaf rust caused by *Melampsora medusae* f. sp. *deltoidae*. *Canadian. J. For. Res.*, **24** : 1984-1987.

Newcombe, G., Bradshaw, H. D., Chastagner, G. A. and Stettler, R. F. (1996). A major gene for resistance to *Melampsora medusae* f. sp. *deltoidae* in a hybrid poplar pedigree. *Phytopathology*, **86**:87-94

Ralph, S., Oddy, C., Cooper, D. and Yueh, H. (2006). Genomics of hybrid poplar (*Populus trichocarpa* × *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): Normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar. *Mol. Ecol.*, **15**:1275-1297.

Van Loon, L.C., Rep, M. and Pieterse, C.M.J. (2006). Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol*., **44**:135-162.