Plant Omics Journal POJ

POJ 9(5):295-303 (2016) ISSN:1836-3644
DOI: 10.21475/poj.09.05.16.pne50

Comparative genomics and expression profile of lipid biosynthesis pathway genes in *Camellia sinensis*

Madhurjya Gogoi[#]*, Afruza Zaman[#], Sangeeta Borchetia, Tanoy Bandyopadhyay

Department of Biotechnology, Tocklai Tea Research Institute, Tea Research Association, Jorhat- 785008, Assam, India

***Corresponding author: mdhjyagogoi@gmail.com**
[#]Authors contributed equally.

Abstract

The lipid biosynthesis pathway is directly associated with lipid content in plants. The genes involved in lipid biosynthesis of tea plant (*Camellia sinensis*) were identified through *in-silico* mining of available transcriptomic data in public domain. Seventy five homologous genes in tea were identified through comparative genomics and 56 sequences of them showed more than 50% sequence similarity with the *Arabidopsis* reference sequences. The expression pattern of five key genes biotin carboxylase (BC), Acyl CoA: diacylglycerol acyltransferase (DGAT), phosphatidylinositol synthase (PIS), monogalactosyldiacylglycerol synthase (MGDGS) and glycerol 3 phosphate dehydrogenase (G3PDH) were analysed through qRT-PCR in roots, leaves, flowers, and seeds of three Tocklai released tea varieties (TV1, TV17 and TV20). Relative expression analysis showed that all the five genes were highly up-regulated in the seeds compared to the other parts of the studied tea varieties. Except biotin carboxylase, other four genes showed highest expression in TV1 seeds compared to TV17 and TV20. The overall increased expression of four key lipid biosynthesis genes give an indication that TV1 seeds may contain higher lipid content than the others in this study. DGAT, PIS and G3PDH also showed upregulation in the leaves and roots of some of the studied varieties which is an indication of possible involvement of lipid biosynthesis genes in various growth and developmental processes.

Keywords: Lipid, oil, qRT-PCR, seed, tea.
Abbreviations: ACC_Acetyl CoA Carboxylase; BC_Biotin Carboxylase; BLAST_Basic Local Alignment Search Tool; DAG_Diacylglycerol; DGAT Acyl CoA_DiacylglycerolAcyltransferase; EST_Expressed sequence Tags; G3PDH_Glycerol 3 Phosphate Dehydrogenase; MEME_Multiple EM for Motif Elicitation; MGDGS_Monogalactosyldiacylglycerol synthase; MSA_ Multiple Sequence Alignment; NR_Non Redundant; PIS_Phosphatidylinositol Synthase; TAG_Triacylglycerol; TSA_Transcriptomics Shotgun Assembly; TTRI_Tocklai Tea Research Institute; TV_Tocklai Vegetative.

Introduction

Lipid, one of the most important macromolecule in the biological system, is required for various growth and developmental processes of the living beings. Lipids include fats, waxes, sterols, fat- soluble vitamins (vitamins A, D, E and K), monoglycerides, diglycerides, triglycerides, phospholipids, etc. A major fraction of the world's lipids are known to be produced by plants which play a vital role in most of the biological processes like seed germination, organ differentiation, pollination, signal transduction, membrane biosynthesis, etc (Bates et al., 2013; Welti and Wang 2004). Certain protective lipids are also found in plants which help in prevention of desiccation and infection (Schmid and Ohlrogge 2002). Plants store significant amount of lipids in seeds in the form of triglycerides (triacylglycerols) which are often commercially exploited for oil extraction. The extracted oil is being regularly used as vegetable oil for cooking and manufacturing of soap, perfumes, etc. Seed oil from certain plants like *Jatropha curcas* are used as biodiesel which have the potential to reduce burden on the conventional petroleum fuels (Bates et al., 2013; Sharma and Chauhan 2012; Xu et al., 2011). Apart from storage lipids present in seeds, the cells and organelles membranes possess a major proportion of lipids in the form of phospholipids and galactolipids, collectively known as glycerolipids (Andersson 2004). The non-phosphorus galactolipids and sulfolipids present in the plastid membranes constitute a major portion of the membrane lipid content (Riekhof and Benning, 2009). These lipids are widely distributed among photosynthetic organisms like higher plants, mosses, eukaryotic and prokaryotic algae but rarely found in non- photosynthetic organisms. Significant amount of galactolipid in flowers indicates its possible role other than photosynthesis (Nakamura 2013). The extraplastidial membranes are predominated by the phosphoglycerolipids (Riekhof and Benning, 2009). Phospholipids, one of the main structural components of plant cells, serve as novel second messengers in response to various stimuli helping in regulation of plant growth and developmental processes like root growth, pollen and vascular development, etc (Meijer and Munnik 2003; Xue et al., 2009). Other than the major oil producing crops like soybean, castor, olive, *Camellia oleifera*, very few crops have been exploited to carry out studies related to their lipid composition and oil content (Taieb et al., 2012; Abdelaziz et al., 2014; Hammond et al., 2005; Ma et al., 2011). Certain species of *Camellia* genus are known to have oil content comparable to the major oil producing crops like olive,

sunflower, groundnut, etc (Singh and Bhattacharjee, 1992). It has been reported that most of the households in southern China use *Camellia* oil for cooking purposes as well as in the industrial production of soap, margarine, hair oil, lubricants, etc (Ruter 2002).

The tea plant (*Camellia sinensis*) is mainly cultivated as a monoculture crop in large areas for commercial production. It is mainly propagated vegetatively to maintain the uniformity in quality and yield of tea. As tea seeds are seldom used in cultivation, there is a possibility of alternate diversification of large seed stocks for the production of edible oil. The oil content in tea seeds was reported to be lower than other species of *Camellia* genus and limited studies are available to utilize tea seeds for oil production (Singh and Bhattacharjee, 1992). Tea seeds reported to possess a very high antioxidant capacity comparable to that of olive oil (Wang et al., 2011; George et al., 2013) which can be utilized for the commercial extraction inspite of its low oil content.

Selection of seeds is an important factor before going for extraction of seed oil. The oil content of seeds may vary in different planting materials and the reliability from different seed source is one of the major concerns with economic return. Biochemical and molecular strategies can be helpful for fast and reliable selection of high seed oil producing plant varieties. There is an urgent need to understand the genetic factors associated with biosynthesis of lipid and its accumulation in tea seeds as well as other parts. The information on gene expression patterns of lipid biosynthesis genes can be used to produce improved varieties of seeds with greater oil content by molecular breeding or genetic engineering approaches. Moreover, such studies will help to understand the role of lipids in growth and development of plants.

In this paper, a comparative genomics study was undertaken to identify the lipid biosynthesis genes in tea plant and expression patterns of few important genes were analyzed using quantitative real time-PCR to reveal its molecular regulation at genetic level.

Results

Comparative analysis of lipid biosynthesis genes in Arabidopsis and tea

The candidate gene mining approach using *Arabidopsis* Lipid Gene Database sequences as reference led to the identification of many corresponding homologous genes in tea (*C. sinensis*). Compared to 75 Arabidopsis lipid biosynthesis reference genes, a large number of *C. sinensis* transcript sequences were obtained from different databases. After filtering and processing, the retained sequences of *C. sinensis* are listed in Table 1. The homologous sequences showed varied resemblance to their reference counterparts in terms of percentage similarity. 56 sequences showed more than 50% similarity to the *Arabidopsis* reference sequences (Table 1).

Motif identification of the predicted sequences

The common motif identification analyses in the protein sequences of *C. sinensis*, *Arabidopsis* and other perennial plants showed that majority of the sequences have conserved common motifs. Of the identified 75 gene sequences, 5 most important genes (BC, DGAT, PIS, MGDGS and G3PDH) were comprehensively studied. The BC sequence of *C. sinensis* was found to have 100% motif similarity with *C. oleifera*. DGAT shared 9 common motifs with *Arabidopsis* out of 12 motifs identified in *C. sinensis*. PIS shared 6 common motifs with *Arabidopsis* out of 7 motifs identified in *C. sinensis*. MGDGS sequence of *C. sinensis* showed 12 common motifs with *Arabidopsis* out of 15 identified motifs. Similarly, G3PDH showed that 11 out of 12 identified motifs were common in *C. sinensis* and *Arabidopsis*. However, there were few minor motifs present in some sequences of *C. sinensis* but absent in the other plant sequences and vice versa [Fig 1 (A-E)].

Quantitative Real Time PCR and gene expression analysis

RNA isolated from the leaves, flowers, roots and seeds of TV1, TV17 and TV20 varieties were of good quality as revealed by the presence of intact 18S and 28S bands on 1% agarose gel. Real Time -PCR analysis of the five studied genes showed significant differences in expression pattern in different parts of the three varieties and showed consistency being highly up-regulated in the seeds. The fold change in the expression level of individual genes in different varieties was measured relative to TV1 leaves and their results are provided below.

Biotin carboxylase (BC)

Biotin carboxylase, one of the most conserved subunit among the four subunits of heteromeric Acetyl CoA carboxylase involved in fatty acid biosynthesis (Nikolau et al., 2003), was found to be highly up-regulated in the seeds of TV1, TV17 and TV20 varieties showing 25- fold, 51.8- fold and 36.2-fold increase in expression level respectively. The roots and flowers of the three varieties showed down regulation of BC with least expression being seen in TV17 flowers. Leaves of TV17 and TV20 also showed lower levels of BC expression (Fig 2).

Acyl CoA: Diacylglycerolacyltransferase (DGAT) DGAT is known to be the rate-limiting enzyme for accumulation of triacylglycerol (TAG), the storage lipid, present in seeds (Turchetto-Zolet et al., 2011). Overall expression of DGAT was found to be higher in the seeds of TV1 and TV17 showing 2.6- fold and 2.4- fold increase and showed lower expression in TV20 seeds. On the other hand, leaves of TV20 showed 2.5-fold higher DGAT expression whereas, it was down regulated in the flowers and roots of all the three varieties (Fig 2).

Phosphatidylinositol synthase (PIS)

Phosphatidylinositol synthase gene showed 119.4- fold increased expression in TV1 seeds. It is known to regulate the accumulation of phospholipids in plant cell membranes and the oil bodies present in seeds (Salas et al., 2006a, b). Seeds of the other two varieties, TV17 and TV20 showed gene up-regulation but with a low fold increase of 19.15 and 13.59 respectively. PIS gene was also minimally up-regulated in the roots of all the studied varieties. Flowers of all the three varieties and leaves of TV17 and TV20 showed a down regulated expression trend with TV1 flowers exhibiting the minimal expression (Fig 2).

Monogalactosyldiacylglycerol synthase (MGDGS)

MGDG synthase, an important enzyme involved in the synthesis of the galactolipid Monogalactosyldiacylglycerol (MGDG) (Nakamura et al., 2010), showed constant decline in gene expression in all the samples except TV1 seeds. TV1 seed, with 1.5-fold increase in expression of MGDGS gene,

Table 1. Percentage similarity of *Camellia sinensis* protein sequences with their corresponding sequences in *Arabidopsis thaliana*. C: Contig, S: Singleton.

Reference *Arabidopsis* Sequence		*Camellia sinensis*	
Gene Name	Gene ID	Highest similarity sequence	Highest Similarity (%)
Plastidial pyruvate dehydrogenase E1a	At1g24180	C3	73.1
Mitochondrial glycerol phosphate	At4g01950	C2	41.3
*Mitochondrial LPAAT	At4g30580	C1	47.2
*1-Acylglycerophosphorylcholine acyltransferase	At3g57650	C1	60.4
3-methylcrotonyl carboxylase biotinylated subunit	At1g03090	C1	9.7
Mitochondrial lipoyltransferase	At1g04640	C1	36.1
Acyl ACP thioesterase FatB	At1g08510	C4	64.1
Mitochondrial diacylglycerol cholinephosphotransferase	At1g13560	C2	79.9
Ceramide sphingobase C4 hydroxylase	At1g69640	C1	81.3
ER Phosphatidate Phosphatase	At1g15080	C2	55.5
Phospholipid base exchange (putative phosphatidyl serine synthase)	At1g15110	S1	0.9
a-ketoacid decarboxylase E1a subunit	At1g21400	C3	60.5
Caleosin	At4g26740	S1	53.0
Plastidial Ketoacyl ACP Reductase	At3g55290	C3	67.8
Plastidial pyruvate dehydrogenase E1b	At1g30120	C1	71.6
Plastidial glycerol phosphate acyltransferase	At1g32200	S1	31.5
Choline kinase	At1g74320	C1	64.8
Plastidial dihydrolipoamide acetyltransferase (Plastidial PD complex)	At1g34430	C4	48.1
Plastidial homomeric acetyl CoA carboxylase	At1g36180	C1	9.0
Stearoyl ACP desaturase	At3g02620	C3	64.1
Mitochondrial dihydrolipoamide dehydrogenase	At1g48030	C1	84.3
Phosphoethanolamine N methyltransferase	At3g18000	C1	79.8
ER 1-acylglycerol phosphate acyltransferase	At3g57650	C1	60.4
Plastidial Acyl carrier protein	At1g54580	C1	50.7
a-ketoacid decarboxylase E1b subunit	At1g55510	C1	78.7
ER CDP-diacylglycerolsynthetase	At1g62430	C4	74.2
Ketoacyl ACP synthase III	At1g62640_a	C2	55.8
Mitochondrial ketoacyl ACP reductase	At1g63380	C3	61.6
Phosphatidyl inositol synthase	At1g68000	C1	77.0
Ketoacyl ACP synthase II	At1g74960_a	C1	72.1
ER 1-acylglycerophosphorylcholine acyltransferase, ER 2-acylglycerophosphorylcholine acyltransferase, ER 2- acylglycerol- phosphate acyltransferase	At3g05510	C1	50.9
Plastidial phosphatidate phosphatase	At2g01180	C1	53.8
Plastdial enoyl ACP reuctase	At2g05990	C1	67.4
Monogalactosyldiacylglycerol synthase	At4g31780	C2	58.2
Acyl-CoA : Diacylglycerol Acyltransferase	At3g51520	S1	64.9
Oil body oleosin	At5g51210	S1	52.3
Ethanolamine kinase (putative choline kinase)	At2g26830	C1	26.3
ER Linoleate Desaturase (Omega 3 fatty acid desaturase)	At2g29980	C3	63.3
Plastidial Malonyl-CoA : ACP Malonyltransferase	At2g30200	C1	65.6
CDP-Choline Synthetase	At4g15130	C1	69.3
Ceramide Fatty Acyl Amide a-Hydroxylase	At2g34770	C1	78.0
a-Carboxyltransferase (Heteromeric Acetyl-CoA Carboxylase)	At2g38040	C1	35.8
CDP-Ethanolamine Synthetase	At2g38670	C4	73.2
Plastidial Phosphatidylglycerol-Phosphate Synthase	At2g39290	S1	56.8
ER Dihydroxyacetone-Phosphate Reductase glycerol 3 phosphate dehydrogenase	At2g41540	C2	75.3
Plastidial CDP-Diacylglycerol Synthetase/ putative phosphatidate cytidylyltransferase	At4g26770	C1	68.7
Mitochondrial CDP-Diacylglycerol Synthetase	At4g26770	C1	68.7
Ceramide Sphingobase D8 Desaturase/	At3g61580	C2	71.2

putative fatty acid desaturase/ cytochrome b5 fusion protein			
Ketosphinganine Reductase	At3g06060	S1	65.4
a-Ketoacid Decarboxylase E2 subunit	At3g06850	C1	56.3
Plastidial Linoleate Desaturase/ omega-3 fatty acid desaturase	At5g05580	C1	71.7
ER Glycerol-Phosphate Acyltransferase	At3g11430	C1	49.3
Digalactosyldiacylglycerol Synthase	At4g00550_a	C1	63.7
ER Oleate Desaturase (Omega 6 fatty acid desaturase, ER, FAD2)	At3g12120	C2	75.9
Monogalactosyldiacylglycerol Desaturase (Palmitate-specific, FAD5)	At3g15850	C1	59.2
Plastidial Dihydrolipoamide Dehydrogenase (Plastidial Pyruvate Dehydrogenase Complex)	At3g16950	C4	49.1
Acyl-ACP Thioesterase Fat A	At3g25110	C1	68.3
ER Diacylglycerol Cholinephosphotransferase	At3g25585	C2	78.1
Phospholipid : Diacylglycerol Acyltransferase	At5g13640	C2	38.7
Isovaleryl-CoA Dehydrogenase	At3g45300	C1	83.1
Serine Palmitoyltransferase (LCB2 subunit)	At5g23670	C1	66.8
ER Phosphatidylglycerol-Phosphate Synthase	At3g55030	S1	69.4
Mitochondrial Enoyl-CoA Hydratase	At4g31810	C2	67.9
ER Phosphatidylserine Decarboxylase	At4g25970_b	C2	73.8
Plastidial 1-Acylglycerol-Phosphate Acyltransferase	At4g30580	C1	47.2
Plastidial Oleate Desaturase (chloroplast omega-6 fatty acid desaturase, FAD6)	At4g30950	C2	73.4
UDP-sulfoquinovose Synthase	At4g33030	C1	77.4
3-Methylcrotonyl-CoA Carboxylase (non-biotinylated subunit)	At4g34030_a	C1	75.1
Serine Palmitoyltransferase (LCB1 subunit)	At4g36480_a	C1	40.6
Plastidial Lipoate Synthase	At5g08410	C1	40.7
Biotin Carboxyl Carrier Protein (Heteromeric Acetyl-CoA Carboxylase)	At5g16390	C1	53.6
Biotin Carboxylase (Heteromeric Acetyl-CoA Carboxylase)	At5g35360	C2/S1	84.8
Plastidial Dihydroxyacetone-Phosphate Reductase	At5g40610	C4	22.2
Ketoacyl-ACP synthase I (KAS I)	At5g46290	C3	50.8
b-Carboxyltransferase (Heteromeric Acetyl-CoA Carboxylase)	AtCg00500	C1	42.3

was the only upregulated sample in our study whereas TV1 flowers showed the least expression amongst the down regulated samples (Fig 2).

Glycerol-3 phosphate dehydrogenase (G3PDH)

Glycerol-3-phosphate dehydrogenase involved in the synthesis of glycerol-3-phosphate from glycolytic intermediates (Vigeolas et al., 2007), was found to be highly up-regulated in the seed samples of TV1, TV17 and TV20 varieties with a 73, 28.6 and 29.04-fold increase in expression respectively. Apart from seeds, the roots, leaves and flowers of TV20 also exhibited gene up-regulation. Similar trend was also observed in the roots of TV17. However, G3PDH was down regulated in the flowers of TV1, TV17, roots of TV1 and leaves of TV17 (Fig 2).

Discussion

Distribution of conserved motifs of lipid biosynthesis genes in tea

Lipids are one of the major macromolecules which serve as the structural, functional and storage components of the living systems. In plant lipid biosynthesis, a number of enzymes are involved and each being the product of a specific gene. Information about the genes involved in lipid biosynthesis pathway of tea is important to understand their role in the production of various types of lipids. In this study, 75 homologous genes involved in lipid biosynthesis were identified in tea (Table 1) through sequence homology search which were further screened for the presence of conserved motif through comprehensive computational analyses. Most genes showed conservation of specific motifs among which *biotin carboxylase* (BC), *acyl CoA: diacylglycerolacyltransferase* (DGAT), *monogalactosyldi-acylglycerol synthase* (MGDGS), *glycerol-3-phosphate dehydrogenase* (G3PDH) and *phosphatidylinositol synthase* (PIS) were considered for further analyses due to their important roles in the biosynthesis of different lipids. Motif analysis showed that the BC gene of *C. sinensis* showed 100% motif similarity with *C. oleifera*. Report of Wei et al. (2015) also stated that evolutionary rates of genes involved in metabolic pathways have no significant difference between *C. oleifera* and *C. sinensis*. The conserved nature of the BC gene is further evident from the common motifs (Fig 1) with *Arabidopsis thaliana* and *Ricinus communis*. The other four genes in this study also showed similar pattern of conserved motifs when compared to *Arabidopsis* and a perennial plant (*Theobroma cacao* or *Ricinus communis*). However, few motifs present in *Arabidopsis* and other perennial plants were absent in *C. sinensis* sequences and vice versa. This may be attributed to the unavailability of full length sequences of lipid biosynthesis genes of *C. sinensis* in public databases.

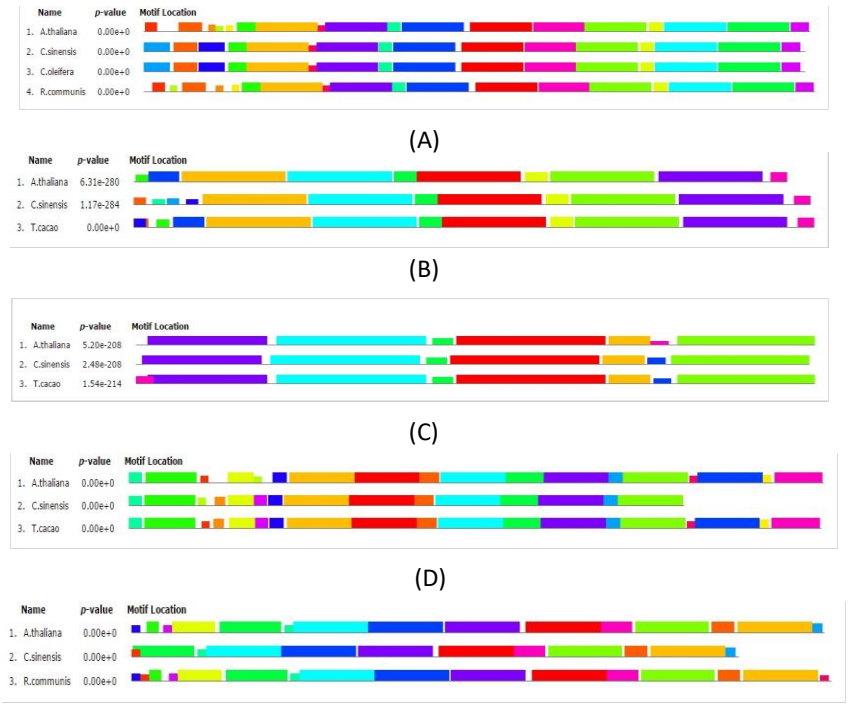

Fig 1 Motif comparison of the protein sequences of different enzymes involved in lipid biosynthesis in *Camellia sinensis* with those of *Arabidopsis thaliana* and a perennial plant. (A) Biotin Carboxylase; (B) Acyl CoA DiacylglycerolAcyltransferase; (C) Phosphatidylinositol Synthase; (D) Monogalactosyldiacylglycerol Synthase; (E) Glycerol 3 Phosphate Dehydrogenase.

Expression pattern of lipid genes in different organs of C. sinensis

The expression patterns of the above discussed five important genes were analysed in different parts of tea plant in three different varieties (TV1, TV17 and TV20). Seeds being the lipid storage part, showed up-regulation of all the studied genes compared to the other parts in all the varieties. Expression results in this study conform to the report of Gu et al. (2002) that the high expression of fatty acid and lipid biosynthetic genes are associated with the accumulation of storage lipids (triglycerides) in the seeds. Expression study in the seeds of the three varieties showed high up-regulation of MGDGS, G3PDH, PIS and DGAT genes in TV1 as compared to TV17 and TV20. BC showed lower expression in TV1 compared to the other two varieties. Reports suggest that differences in the content and profile of oil can be attributed to the differential expression pattern of lipid biosynthesis genes (Rusanov et al., 2011; Sharma and Chauhan, 2012). TV1 showed comparatively better expression of lipid biosynthesis genes suggesting a possibility of higher lipid content in TV1 than the other two varieties.

Expression pattern of biotin carboxylase (BC)

Acetyl CoA Carboxylase (ACCase) catalyses the carboxylation of acetyl CoA to malonyl CoA, the first committed step of fatty acid synthesis, which is essential for survival of plants. Biotin carboxylase is one of the most conserved subunits of the four subunits of ACCase among plants (Nikolau et al., 2003). This subunit was found to be highly up-regulated in the seeds of three TV varieties. The key involvement of ACCase subunits in lipid biosynthesis in seeds was earlier reported by Roesler et al. (1997) in *B. napus*. The transcripts of the ACCase subunit accumulates in high amount in the developing seeds where the synthesized fatty acids are used in the assembly of membrane lipids or storage lipids in the form of triacylglycerols (Nikolau et al., 2003). The high accumulation of BC subunit in the seeds of *C. sinensis* can be associated with the high lipid content of seeds, particularly the triacylglycerols.

Expression pattern of acyl CoA: diacylglycerol acyltransferase (DGAT)

TAGs represent the major storage lipids in developing seeds, petals, pollen grains and fruits. DGAT catalyses the conversion of DAG to TAG, which is considered to be the rate - limiting enzyme in TAG accumulation in plants (Turchetto-Zolet et al., 2011). TV1 and TV17 seeds showed increased expression of DGAT gene than other parts which was reported earlier in relation to the involvement of this gene in the up-regulation of TAGs in seeds of *Jatropha* (Xu et al., 2011), castor (Chandrasekaran et al., 2014) and *Camellia oleifera* (Xia et al., 2014). Overexpression of DGAT gene through the application of genetic engineering techniques has resulted in a substantial increase in the seed oil content of *Arabidopsis* and *B. napus* (Taylor et al., 2009; Jako et al., 2001). This implies that DGAT gene expression level has direct influence on the accumulation of seed oil and increase in seed-specific expression can be a probable strategy for regulating the quantity of TAGs and oil content of *C. sinensis* seeds. Leaves of TV20 also showed comparatively higher expression of DGAT with a 2.7

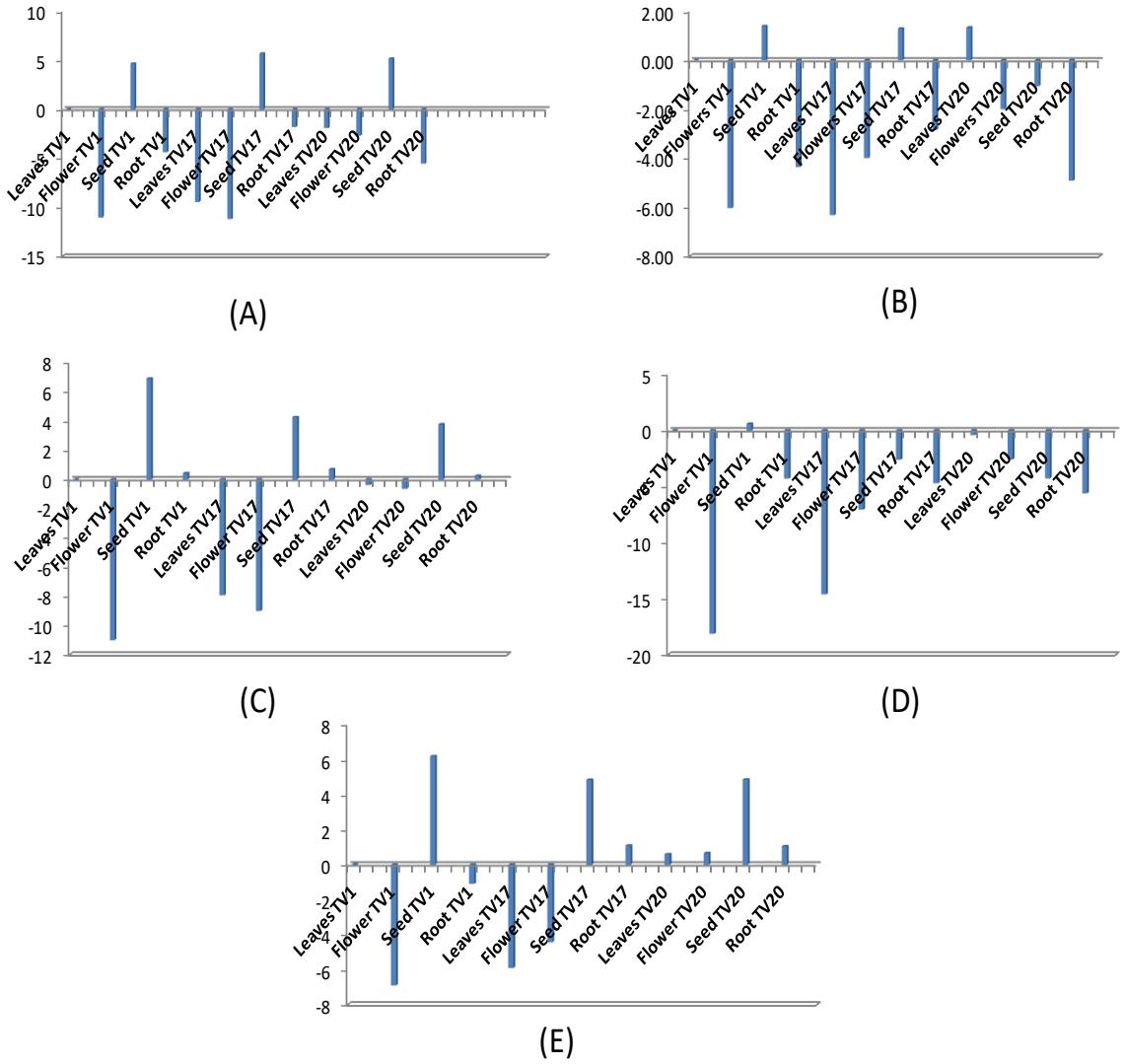

Fig 2. Diagrammatic representation of the fold change (Log 2 Fold) expression of different lipid biosynthesis genes across different organs of *Camellia sinensis*. (A) Biotin Carboxylase; (B) Acyl CoA Diacylglycerol Acyltransferase; (C) Phosphatidylinositol Synthase; (D) Monogalactosyldiacylglycerol Synthase; (E) Glycerol 3 Phosphate Dehydrogenase.

- fold up-regulation than TV1 leaves. The high expression of DGAT in leaves may be linked with growth and development or can be associated with comparatively higher level of lipid content. Increase in the expression of DGAT genes in genetically modified tobacco resulted in high oil content in leaves comparable to seeds (Andrianov et al., 2010).

Expression pattern of phosphatidylinositol synthase (PIS)

PIS is a key enzyme of the phospholipid pathway involved in the synthesis of phosphatidylinositol (PtdIns), a structural component of cell membranes (Liu et al., 2013). The oil bodies present in the seeds contain a monolayer of phospholipids (phosphatidylserine and phosphatidylinositol) and small amounts of free fatty acids surrounding a triacylglycerol matrix. The increased expression of PIS in the seeds of the three studied varieties indicated presence of oil bodies in *C. sinensis* seeds as was reported in mature seeds of other plants like mustard (*Brassica juncea* L.), cotton (*Gossypium hirsutum* L.), flax (*Linus usitatissimum*), rape (*Brassica napus* L.), etc. (Tzen et al., 1992; Tzen et al., 1993). The up-regulation of PIS in the roots of TV1, TV17 and TV20 varieties is comparable with earlier reports which showed the involvement of certain phospholipids in root growth (Xue et al., 2009).

Expression pattern of monogalactosyldiacylglycerol synthase (MGDGS)

MGDG is an important galactolipid constituting around 50% of the total chloroplastic membrane (Nakamura et al., 2010). Reports showed that MGDGS up-regulation during inorganic phosphate starvation resulted in oil accumulation in the vegetative tissues (Shimojima et al., 2015). Our results showed 1.5- fold increase in the level of MGDGS in the TV1 seeds and down regulation in other parts relative to the TV1 leaves (Fig 2). This may be due to abundance of inorganic phosphates which led to lower expression of MGDGS gene in majority of the samples.

Expression pattern of glycerol-3-phosphate dehydrogenase (G3PDH)

G3PDH catalyses the synthesis of glycerol-3- phosphate which is required for the assembly of triacylglycerols (TAG), the main constituent of edible oils. Reports indicated that 2

fold increase in the activity of glycerol-3-phosphate dehydrogenase in developing seeds resulted in 3 - 4 fold increase in glycerol-3- phosphate level which increased the lipid content to about 40% (Vigeolas et al., 2007). Similarly, overexpression of G3PDH results in a 60% increase in neutral lipid content in the marine algae *Phaeodactyl umtricornutum* (Yao et al., 2014). Up-regulation of G3PDH in the seeds of the three TV varieties can be assumed to follow similar pattern leading to an increase in the total lipid content. High expression of G3PDH in the leaves, flowers and roots of TV20 variety as well as in the TV17 roots also indicated a probable role of this gene in plant growth and development.

The comparatively higher expression of some studied genes (DGAT, PIS and G3PDH) in the vegetative tissues indicated that the lipid biosynthesis genes may have important role in growth and development of tea plant besides lipid accumulation. Earlier studies proposed that apart from using seeds for oil extraction transgenic approaches can be undertaken for increasing lipid content in vegetative tissues as a novel platform for meeting global production needs for low-cost, energy-dense lipids (Chapman et al., 2013). This kind of strategy may be used in *C. sinensis* to convert unharvested mature leaves into valuable co-products for tea diversification.

Materials and Methods

In-silico identification of lipid biosynthesis genes in *Camellia sinensis*

The transcripts involved in lipid biosynthesis pathway of *C. sinensis* were mined from different databases using *Arabidopsis* sequences as reference. The *Arabidopsis* reference protein sequences were collected from The *Arabidopsis* Lipid Gene Database (http://lipids.plantbiology.msu.edu/) and subsequently the protein sequences were subjected to NCBI BLAST (tblastn) with *C. sinensis* EST (Expressed Sequence Tags), TSA (Transcriptomics Shotgun Assembly) and NR (Non Redundant) databases. The sequences showing significant similarity with an E-value ≤ $1e^{-15}$ were selected and assembled using CAP3 program to remove redundancy and get consensus sequences. Each of these sequences were screened for the presence of open reading frame using NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) and the sequences with the longest frame having both start and stop codon were sorted out and the corresponding translated protein sequence was obtained. The sequences were further subjected to BLAST with NCBI NR (Non-Redundant) database to confirm their annotation. Based on the BLAST annotation and alignment results, more sequences were screened and eliminated. For each of the sequence of *C. sinensis*, the best two BLAST hits were selected: one of *Arabidopsis* and one of any perennial plant (*Theobroma cacao* or *Ricinus communis*).The sequences were then aligned and compared by creating a multiple sequence alignment (MSA). The Identity Matrix from the MSA was analyzed to find the percentage similarity between the sequences.

Motif identification

MEME program was used for the identification of the motif cluster present in the lipid biosynthesis protein sequences. The sequences from *C. sinensis*, *Arabidopsis* and other perennial plants were together analyzed for the easy identification of common motifs between them.

Sample collection

Tocklai vegetative, TV1 (Assam- China hybrid), TV17 (Assam hybrid) and TV20 (Cambod) varieties were selected from TTRI germplasm collection plot for collection of roots, leaves, flowers and seeds. The collected samples were immediately kept at -80^0C to prevent any enzymatic activity.

RNA isolation and cDNA preparation

RNA was isolated from root, leaf, flower and seed samples separately by little modification of the protocol described by Das *et al.* (2013). The isolated RNA were resolved in a 1% agarose gel and quantified using a Biophotometer (Eppendorf, Hamburg, Germany). After verifying the integrity of the RNA, 500ng of RNA from each sample was used for first strand cDNA preparation using Verso first strand cDNA synthesis kit (Thermo Scientific, USA).

Quantitative Real-Time PCR analysis

Five important genes known to have a major role in regulating the lipid biosynthesis pathway in plants were selected based on *in-silico* identification of *C. sinensis* genes to study their expression patterns in the roots, leaves, flowers and seeds of three selected tea varieties. The selected genes included Biotin carboxylase (BC) subunit of the Acetyl CoA carboxylase (ACC) enzyme which is the initial enzyme of fatty acid biosynthesis; Acyl CoA: diacylglycerolacyl transferase (DGAT), the final enzyme of triacylglycerol biosynthesis; Monogalactosyldiacylglycerol synthase (MGDGS), the final enzyme for mono and diacylglycero lipids; Glycerol 3 phosphate dehydrogenase (G3PDH), the branch point enzyme for synthesis of phospholipids, galactolipids and triacylglycerols; and Phosphatidylinositol synthase (PIS), the final enzyme for synthesis of phosphatidyl 1D- myoinositol phospholipid.

Gene specific primers were designed using the Primer3 (v.0.4.0) software to generate products in the 150-250bp range for quantitative Real Time-Polymerase Chain Reaction (qRT-PCR). qRT-PCR was performed in a Roche Light Cycler 480 using the cDNA as template and amplified by gene-specific primers. The PCR was performed in a 10 µl reaction volume containing 0.5 µl cDNA template, 10ng of each primer, 1X SYBR Green master mix (Thermo-Scientific, USA) and nuclease free water to adjust the reaction volume.

The expression levels of all the genes were calculated using the ddCt method. The raw Ct values were normalized against 18S rRNA house-keeping gene. The fold change of each sample was calculated relative to TV1 leaves and the log 2 fold-change data thus obtained was used for the representation of gene expression.
(The sequences of the five gene-specific primers used for qRT-PCR are provided in the supplementary table S1)

Conclusion

Studies related to the lipid biosynthesis pathway was reported from many plants but limited work was done for *C. sinensis*. The present study led to the identification of many homologous genes involved in lipid biosynthesis pathway of tea (*C. sinensis*) using comparative genomics. The gene expression analysis revealed that like other oil producing

plants, the seeds of *C. sinensis* showed highest expression of the lipid biosynthesis genes compared to other tissues. This study put forth an idea for diversification of non-harvested tea seeds for extraction of edible seed oil. Further research work also can be carried for studying the content and composition of essential oils in tea seeds for their probable industrial applications. The possible growth and developmental role of lipids have also been reflected in the present study as leaves and roots of *C. sinensis* showed high upregulation of few lipid genes. More comprehensive study may lead to elucidation of molecular mechanism governing the lipid accumulation in seeds of tea plant and thus may help in development of new plant varieties having high seed oil content using genetic engineering approaches.

Acknowledgement

The authors are grateful to Director, TRA, for constant encouragement. M Gogoi and A Zaman are thankful to DeitY for their fellowship from "Tea Bioinformatics" project. The authors acknowledge Department of Biotechnology for financial support through the project "Establishment of Institutional Biotech Hubs by DBT under special programme for the North Eastern States of India". We would also like to thank our ex-technical officer Mr. Ranjit Bordoloi for his help in sample collection.

References

Abdelaziz AIM, Elamin IHM, Gasmelseed GA, Abdalla BK (2014) Extraction, refining and characterization of Sudanese castor seed oil. J Chem Eng. 2:1-4.

Andersson M (2004) Lipid trafficking: into, within and out of the chloroplast. Thesis, Department of Botany, Göteborg University, Sweden

Andrianov V, Borisjuk N, Poqrebnyak N, Brinker A, Dixon J, Spitsin S, Flynn J, Matyszczuk P, Andryszak K, Golovkin M, Koprowski H (2010) Tobacco as a production platform for biofuel: overexpression of *Arabidopsis* DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. Plant Biotechnol J. 8:277-287.

Bates PD, Stymne S, Ohlrogge J (2013) Biochemical pathways in seed oil synthesis. Curr Opin Plant Biol. 16(3):358-364.

Chandrasekaran U, Liu A (2014) Differential expression patterns of lipid genes in developing seeds of castor bean (*Ricinus communis* L.). Bangladesh J Botany. 43:37-43.

Chapman K, Dyer JM, Mullen RT (2013) Commentary: why don't plant leaves get fat? Plant Sci. 207:128-134.

Das A, Saha D, Mondal TK (2013) An optimized method for extraction of RNA from tea roots for functional genomics analysis. Indian J Biotechnol. 12:129-132.

George KO, Kinyanjui T, Wanyoko J, Moseti OK, Wachira F (2013) Extraction and analysis of tea (*Camellia*) seed oil from different clones in Kenya. Afr J Biotechnol. 12:841-846.

Gu K, Yi C, Tian D, et al (2012) Expression of fatty acid and lipid biosynthetic genes in developing endosperm of *Jatropha curcas*. Biotechnol Biofuels. 5:47-61.

Hammond EG, Johnson LA, Su C, Wang T, White PJ (2005) Soybean oil. In: Shahidi F (ed) Bailey's Industrial Oil and Fat Products, 6th edn. John Wiley & Sons, New York.

Jako C, Kumar, A, Wei Y, Zou J, Barton DL, Giblin EM, Covello PS, Taylor DC (2001) Seed-Specific Over-Expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiol. 126:861-874.

Liu X, Zhai S, Zhao Y, Sun B, Liu C, Yang A, Zhang J (2013) Overexpression of the phosphatidylinositol synthase gene (ZmPIS) conferring drought stress tolerance by altering membrane lipid composition and increasing ABA synthesis in maize. Plant Cell Environ. 36:1037-1055.

Ma J, Ye H, Rui Y, Chen G, Zhang N (2011) Fatty acid composition of *Camellia oleifera* oil. J Verbrauch Lebensm. 6:9-12.

Meijer HJG, Munnik T (2003) Phospholipid- based signaling in plants. Annu Rev Plant Biol. 54:265-306.

Nakamura Y (2013) Galactolipid biosynthesis in flowers. Bot Stud. 54: 29-35.

Nakamura Y, Shimojuma M, Ohta H, Kobayashi K (2010) Biosynthesis and function of monogalactosyldiacylglycerol (MGDG), the signature lipid of chloroplasts. In: Rebeiz CA, Benning C, Bohnert H, Daniell H, Hoober JK, Lichtenthaler HK, Portis AR, Tripathy BC (Eds.) The Chloroplast: Basics and Application, Springer, Netherlands.

Nikolau BJ, Ohlrogge JB, Wurtele ES (2003) Plant biotin-containing carboxylases. Arch Biochem Biophys. 414:211-222.

Riekhof WR, Benning C (2009) Glycerolipid biosynthesis. In: Stern DB (Ed) The chlamydomonas sourcebook: organeller and metabolic processes, 2nd edn. Elsevier, Dordrecht, Netherlands.

Roesler K, Shintani D, Savage L, Boddupalli S, Ohlrogge J (1997) Targeting of the *Arabidopsis* homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiol. 113:75-81.

Rusanov KE, Kovacheva NM, Atanassov II (2011) Comparative GC/MS analysis of rose flower and distilled oil volatiles of the oil bearing *Rosa Rosa Damascena*. Biotechnol Biotec Eq. 25:2210-2216.

Ruter JM (2002) Nursery production of tea oil *Camellia* under different light conditions. In: Janick J and Whipkey A (Eds.) Trends in new crops and new uses, ASHS Press, Alexandria, VA

Salas JJ, Force EM, Garces R (2006a) Accumulation of phospholipids and glycolipids in seed kernels of different sunflower mutants (*Helianthus annus*). J Am Oil Chem Soc 83:539-545.

Salas JJ, Martinez-Force E, Garces R (2006b) Phospholipid molecular profiles in the seed kernel from different sunflower (*Helianthus annus*) mutants. Lipids. 41:805-811.

Schmid KM, Ohlrogge JB (2002) Lipid metabolism in plants. In: Vance DE and Vance JE (Eds.) Biochemistry of Lipids, Lipoproteins and Membranes, 4th edn. Elsevier, Amsterdam.

Sharma A, Chauhan RS (2012) *In silico* identification and comparative genomics of candidate genes involved in biosynthesis and accumulation of seed oil in plants. Comp Funct Genom. 2012: article ID 914843. doi: http://dx.doi.org/10.1155/2012/914843

Shimojima M, Madoka Y, Fujiwara R, Murakawa M, Yoshitake Y, Ikeda K, Koizumi R, Endo K, Ozaki K, Ohta H (2015) An engineered lipid remodeling system using a galactolipid synthase promoter during phosphate starvation enhances oil accumulation in plants. Front Plant Sci. 6:664. doi: http://dx.doi.org/10.3389/fpls.2015.00664

Singh ID, Bhattacharjee H (1992) Tea seed oil. Two and a Bud. 39:7-11.

Taieb NH, Grati N, Ayadi M, Attia I, Bensalam H, Gargouri A (2012) Optimisation of olive oil extraction and minor compounds content of Tunisian olive oil using enzymatic formulations during malaxation. Biochem Eng J. 62:79-85.

Taylor DC, Zhang Y, Kumar, A, Francis T, Giblin EM, Barton DL, Ferrie JR, Laroche A, Shah S, Zhu W, Snyder CL, Hall L, Rakow G, Harwood JL, Weselake RJ (2009) Molecular modification of triacylglycerol accumulation by over- expression of DGAT1 to produce canola with increased seed oil content under field conditions. Botany. 87:533-543.

Tjellstrom H, Strawsine M, Ohlrogge JB (2015) Tracking synthesis and turnover of triacylglycerol in leaves. J Exp Bot. doi: 10.1093/jxb/eru500

Turchetto- Zolet AC, Maraschin FS, de Morais GL, Cagliari A, Andrade CM, Margis-Pinheiro M, Margis R (2011) Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. BMC Evol Biol. 11:263- 276.

Tzen JTC, Cao YZ, Laurent P, Ratnayake C, Huang AHC (1993) Lipids, proteins and structure of seed oil bodies from diverse species. Plant Physiol. 101:267-276.

Tzen JTC, Lie GC, Huang AH (1992) Characterization of the charged components and their topology on the surface of plant seed oil bodies. J Biol Chem. 267:15626-15634.

Vigeolas H, Waldeck P, Zank T, Geigenberger P (2007) Increasing seed oil content in oil – seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol- 3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnol J. 5:431-441.

Wang Y, Sun D, Chen H, Qian L, Xu P (2011) Fatty acid composition and antioxidant activity of tea (*Camellia sinensis* L.) seed oil extracted by optimized supercritical carbon dioxide. Int J Mol Sci. 12:7708-7719.

Wei Y, Jing W, Youxiang Z, Mingming Z, Yan G, Hua D, Lijun P, Dingjin H (2015) Genome-wide identification of genes probably relevant to the uniqueness of tea plant (*Camellia sinensis*) and its cultivars. Int J Genom. 2015: article ID 527054. doi: http://dx.doi.org/10.1155/2015/527054

Welti R, Wang X (2004) Lipid species profiling: a high-throughput approach to identify lipid compositional changes and determine the function of genes involved in lipid metabolism and signaling. Curr Opin Plant Biol. 7(3):337-344.

Xia EH, Jiang JJ, Huang H, Zhang LP, Zhang HB, Gao LZ (2014) Transcriptome analysis of the oil rich tea plant, *Camellia oleifera*, reveals candidate genes related to lipid metabolism. PLoS One. 9(8):e104150.

Xu R, Wang R, Liu A (2011) Expression profiles of genes involved in fatty acid and triacylglycerol synthesis in developing seeds of Jatropha (*Jatropha curcas* L.). Biomass Bioenerg. 35:1683-1692.

Xue HW, Chen X, Mei Y (2009) Function and regulation of phospholipid signaling in plants. Biochem J. 421:145-156.

Yao Y, Lu Y, Peng KT, Huang T, Niu YF, Xie WH, Yang WD, Liu JS, Li HY (2014) Glycerol and neutral lipid production in the oleaginous marine diatom *Phaeodactylum tricornutum* promoted by overexpression of glycerol-3-phosphate dehydrogenase. Biotechnol Biofuels. 7:110-118.