SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
PLANT ECOLOGY
MOLECULAR ECOLOGY

Received
10 February 2014

Accepted
14 July 2014

Published
4 August 2014

Correspondence and requests for materials should be addressed to S.K. (sanjaykumar@ihbt.res.in; sanjayplp@rediffmail.com)

* These authors contributed equally to this work.

† Current address: National Institute of Plant Genome Research, Aruna Asaf Ali Marg, P.O. Box No. 10531, New Delhi-10067, India.

RNA-seq-mediated transcriptome analysis of actively growing and winter dormant shoots identifies non-deciduous habit of evergreen tree tea during winters

Asosii Paul[1]*†, Ashwani Jha[2]*, Shruti Bhardwaj[1], Sewa Singh[1], Ravi Shankar[2] & Sanjay Kumar[1]

[1]Biotechnology Division, CSIR-Institute of Himalayan Bioresource Technology, P.O. Box No. 6, Palampur-176061, Himachal Pradesh, India, [2]Studio of Computational Biology & Bioinformatics, CSIR-Institute of Himalayan Bioresource Technology, P.O. Box No. 6, Palampur-176061, Himachal Pradesh, India.

Tea [*Camellia sinensis* (L.) O. Kuntze] is a perennial tree which undergoes winter dormancy and unlike deciduous trees, the species does not shed its leaves during winters. The present work dissected the molecular processes operating in the leaves during the period of active growth and winter dormancy through transcriptome analysis to understand a long-standing question: why should tea be a non-deciduous species? Analyses of 24,700 unigenes obtained from 57,767 primarily assembled transcripts showed (i) operation of mechanisms of winter tolerance, (ii) down-regulation of genes involved in growth, development, protein synthesis and cell division, and (iii) inhibition of leaf abscission due to modulation of senescence related processes during winter dormancy in tea. These senescence related processes exhibited modulation to favour leaf abscission (i) in deciduous *Populus tremula* during winters, and (ii) also in tea but under osmotic stress during which leaves also abscise. These results validated the relevance of the identified senescence related processes for leaf abscission and suggested their operation when in need in tea.

Tea [*Camellia sinensis* (L.) O. Kuntze] is an evergreen tree species that yields a non-alcoholic beverage, tea. Tea tree is trimmed to a bush of about 0.9 to 1.25 m to ease plucking of apical bud and the associated two leaves (popularly known as two and a bud) that is used for commercial production of tea. Unlike deciduous trees such as *Populus tremula*, tea leaves do not exhibit the phenomenon of autumnal senescence, rather the growth of two and a bud is diminished, a phenomenon popularly known as winter dormancy (WD)[1,2]. Therefore, work on deciduous tree species was focused on autumn senescence[3], whereas WD was studied in non-deciduous tree species such as tea[4]. Autumn senescence is triggered by reduction in the photoperiod wherein phytochromes played a central role[5]. Detailed molecular analyses on autumn senescence in *P. tremula* showed major changes in gene expression including up-regulation of genes encoding for a variety of catabolic enzymes (proteases, lipases, nucleases)[3,6].

Leaf senescence leading to deciduous leaf habit is considered an 'opportunist' strategy and is characterized by having higher (i) leaf area per unit mass, (ii) leaf nutrient contents and (iii) photosynthetic capacity[7,8]. Such trees have high rates of carbon gain when environmental conditions are favorable and avoid maintenance and adaptation costs by shedding their leaves during unfavorable seasons[8]. The evergreen leaf habit, on the other hand with increased leaf life span, continues to photosynthesize during unfavorable season when deciduous species cannot[7], and compensates for ongoing maintenance costs and low carbon gain[8].

WD in tea sets in when the day light period becomes shorter than a critical light period of 11 h 15 min and minimum temperature falls below 13°C for at least six weeks[9]. Shorter day light period alters the balance of endogenous growth regulators in favor of dormancy and longer light period in the favor of growth in tea[1]. WD accompanies accumulation of abscisic acid (ABA) and reduction of gibberellins (GAs) levels[10,11]. Also, photosynthesis rates were reduced with concomitant imposition of oxidative stress during winters in tea[12–14]. Molecular analyses during WD in tea showed down-regulation of genes associated with protein synthesis and cell division leading to diminished growth and developmental activities during winter season[4,15]. Targeted gene analysis in tea showed an association of *histone H3 gene*[16], *QM like protein homologue*[17], and *alpha-tubulin*[18] with WD.

Table 1 | Summary of transcriptome data generated on Illumina Genome Analyzer IIx for two and a bud during the period of active growth and winter dormancy

	Period of active growth	Winter dormancy	Pooled
Total number of paired-end reads	81,182,096	49,594,708	130,776,804
Number of reads obtained after quality filtering	62,471,502	41,600,636	104,072,138
Number of primary assembled transcripts of pooled data	Not applicable	Not applicable	57,767
Average length of transcripts (bp) of pooled data	Not applicable	Not applicable	505.44
Average coverage of pooled data	Not applicable	Not applicable	111.28

Perennial, evergreen tree tea provides an opportunity to decipher the molecular processes that operate during winters in tea to make it a non-deciduous species. A transcriptome-based approach was followed to understand the processes in an integrated manner during winters (i.e. when the species experiences WD) and the period of active growth (PAG). Also, the identified processes were validated using relevant systems.

Results and Discussion

Read generation and *de novo* assembly. Six and eight picomoles of the libraries were used to generate Paired-End (PE) reads. Two different quantities of library were used to account for any technical variance in unigenes in the transcriptome data[19]. Six picomoles of library generated 25,815,706 and 28,154,978 PE reads from the tissues during PAG and WD, respectively; since two and a bud during winters were dormant, WD was used interchangeably for winters to express growth phase of tea. The read numbers for 8 picomole library were 55,366,390 and 21,439,730 in the same order. A total of 81,182,096 and 49,594,708 PE reads were obtained for PAG and WD library, respectively. After filtering for quality and contamination, a total of 62,471,502 and 41,600,636 reads were obtained for PAG and WD libraries, respectively. A total of 104,072,138 PE reads were obtained (PE read of 36×2 bp, fragment size 200 bp) from PAG and WD libraries (Table 1). Best primary assembly of short reads was obtained at a k-mer size of 21 nucleotides (Table 2). A total of 57,767 primarily assembled transcripts (Table 1) were generated from the pooled data, having an average length size of 505.44 bp and average coverage of 111.28; 13.84% of sequences were 1 kb or longer. The longest sequence length obtained was 5.828 kb.

Homology search and sequence clustering. Using hierarchical clustering approach involving TGICL-CAP3 and CD-HIT[20], a total of 57,027 unique assembled transcript sequences were obtained (http://scbb.ihbt.res.in/Tea-Teenali-IHBT/Tea-Teenali/; Supplementary Table S1). BLAST[21] hits were found for 33,784 sequences while 23,243 sequences showed no hit (Supplementary Table S1). Dissimilar sequence clustering[20] was performed to cluster the assembled unique transcript sequences in the form of unigene representation and to curtail inflated representation of total unigenes represented by the assembled sequences. This way, a total of 24,700 unigenes were identified from the assembled sequences (Supplementary Table S1). A total of 23,243 transcripts, which did not show any homologue from Non-Redundant (NR) database, were translated into six open reading frames (ORFs) and searched for functional domains in Conserved Domain Database (CDD)[22] using RPS-BLAST. Significantly conserved domains were found for 253 sequences (Supplementary Table S1). The highly representative domain was of fibronectin-attachment protein (5.13%).

Functional annotation and characterization of the unigenes. Gene ontology (GO) classification was found for 18,316 unigenes that were further classified into biological process and molecular function categories (Supplementary Table S2). Genes involved in metabolic processes were highly represented in biological process category (Supplementary Fig. S1A). Functional classification of the annotated unigenes in molecular function category (Supplementary Fig. S1B) revealed that DNA binding was the highly represented group.

Similarly, Enzyme Commission (EC) classification was obtained for 8,856 unigenes, while Kyoto Encyclopedia of Genes and Genomes (KEGG) classification was obtained for 9,819 unigenes (Supplementary Table S3). As per the EC classification, a large amount of assembled unigenes belonged to non-specific serine/threonine protein kinase enzyme class alone (16.76%) (Supplementary Fig. S2A, Supplementary Table S3). Whereas, KEGG classification identified highest number of sequences belonging to plant-pathogen interaction pathways (5.44%) (Supplementary Fig. S2B).

Identification, functional annotation and characterization of the differentially expressed unigenes (DEUs). The correlation coefficient of gene expression, as measured through reads per kilo base per million (RPKM), between six picomoles and eight picomoles libraries was 0.997 (p-value = $2.20\ e^{-16}$) and 0.997 (p-value = $2.20\ e^{-16}$) for PAG and WD, respectively (Supplementary Table S1). Hence, the two libraries served as replicates, apart from offering better confidence and higher coverage. Tool edgeR[23] was used to identify significantly up- and down-regulated unigenes on the read count values of unigenes from the tissues during PAG and WD. A total of 5,204 out of 24,700 unigenes exhibited significant alteration in expression after applying Fisher's exact test on a negative binomial distribution using edgeR (Supplementary Table S1). Analyses of biological processes and molecular functions in the tissues during PAG and WD showed that several genes associated with molecular functions such as catalytic activity, and DNA binding were significantly modulated during WD (Supplementary Fig. S3).

Table 2 | Effect of k-mer size on assembling performance of tea transcriptome

K-mer	Total number of sequence	Average length of sequence (bp)	Maximum length of sequence (bp)	Average coverage	Number of sequence length 1000 bp and above	Percentage of transcripts (1000 bp and longer)
19mer	64,696	493.95	6,984	110.71	8,237	12.731
21mer	57,767	505.44	5,828	111.28	7,996	13.841
23mer	50,433	492.56	7,403	112.11	6,725	13.334
25mer	42,916	459.40	4,848	113.73	4,967	11.573
27mer	33,293	411.41	4,198	116.35	2,762	8.296
29mer	19,677	345.39	3,266	123.92	901	4.578

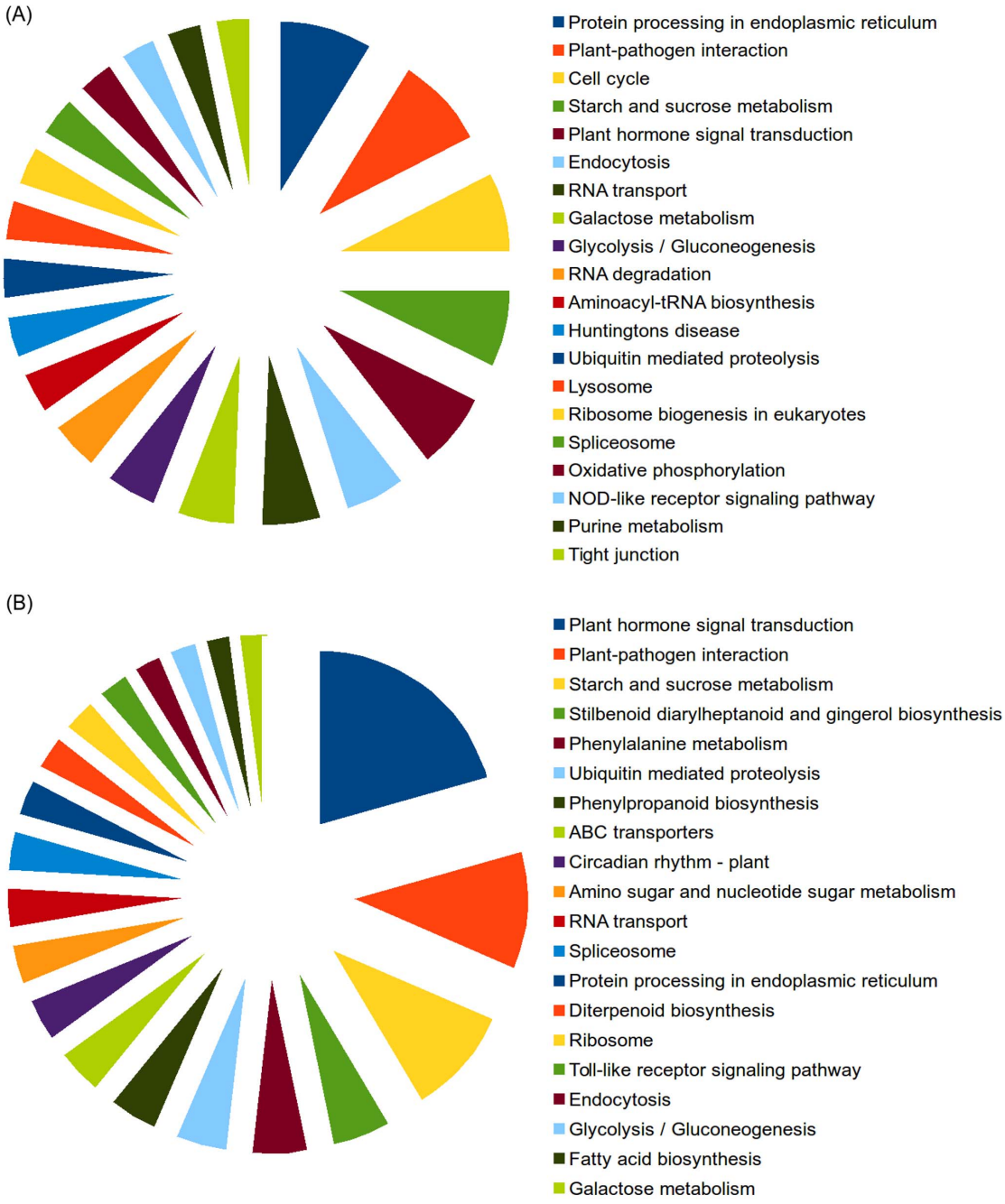

Figure 1 | Top 20 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways up-regulated during the period of active growth (A) and winter dormancy (B). Supplementary Table S3 has details on all the differentially expressed KEGG pathways for the two periods.

Biological processes such as protein transport and cell division were prominent in PAG as compared to those during WD (Supplementary Fig. S4). GO enrichment analysis showed that the genes associated with DNA binding and symporter activity were significantly enriched during PAG (Supplementary Fig. S5). GO Slim of DEUs showed that genes associated with transcription, DNA dependent and response to abiotic or biotic stimulus were prominently over-represented during WD (Supplementary Table S2); whereas cell organization and biogenesis, electron transport/energy pathways and DNA and RNA metabolism were down-represented during WD (Supplementary Table S2). Similar results were also observed in *Euphorbia esula* during seasonal dormancy transitions[24]. KEGG pathways analyses using DEUs showed that those associated with protein processing in endoplasmic reticulum, cell cycle, endocytosis and RNA transport were significantly down-regulated during WD whereas, up-regulated pathways included plant hormone signal transduction and plant-pathogen interaction (Fig. 1; Supplementary Table S3).

Functional and pathway assignments of the DEUs using GO Slim and KEGG classification revealed numerous hormonal, physiological, and developmental changes during WD. These included alterations in (i) responses to plant growth regulator, (ii) cell cycle, (iii) stress-tolerance, (iv) transport, (v) signaling, (vi) protein synthesis and turnover, (vii) energy and (viii) metabolism (Fig. 1, Supplementary Fig. S3). Genes related to cell rescue/defense, metabolism, protein synthesis and transcription were shown to be most regulated during WD and dormancy break in sessile oak[25]. In leafy spurge, genes involved in catalytic activity were dominant in the growing buds, whereas those involved in DNA/RNA binding were the most prominent in dormant buds[26]. The genes related to stress

tolerance/detoxification dominated during dormancy in *Rubus idaeus*[27]. The present data suggested establishment of a metabolic equilibrium during WD to enable tea to tolerate the "harsh" environment of winters.

Differentially expressed transcription factors (TFs). Transcription factors are sequence specific DNA-binding proteins that interact with the *cis*-acting element in the promoter regions of respective target genes, and modulate gene expression[28]. A total of 455 transcription factor unigenes (224 from PAG and 231 from WD; Supplementary Table S2) representing 31 transcription factor families (Supplementary Table S4) exhibited significant difference in expression (Supplementary Table S2). The TFs exhibiting down-representation during WD included those encoding cysteine-3/histidine zinc finger domain (C3H), cysteine-rich polycomb-like protein (CPP), E2 promoter binding factor-dimerization partner (E2F-DP), forkhead-associated domain (FHA), and mitochondria transcription termination factor (mTERF) (Supplementary Table S2). Whereas genes encoding biotic and abiotic stresses and development related TFs were significantly over-represented during WD (Fig. 2). These TFs included APETALA2-ethylene-responsive element binding proteins (AP2-EREBP), cysteine-2/histidine-2 zinc finger proteins (C2H2), bri1-EMS-suppressor 1 (BES1), GAI-RGA-SCR (GRAS), lateral organ boundaries (LOB), and WRKY class. Interestingly, TFs encoding for SHI related sequence (SRS) class of TF, were up-regulated during WD (Supplementary Table S4). SRS class of TFs were implicated in suppression of GA responses in young organs to prevent premature growth or development[29] suggesting modulation of GA signaling and biosynthesis during WD. Indeed, GAs were shown to be modulated during WD and PAG[11]. Sensitivity to GA is also regulated by proteins belonging to GRAS family of plant transcriptional regulators[28]. Consistent with the down-regulation of GA biosynthesis, two genes encoding for the GA-INSENSITIVE (GAI) proteins, considered to maintain a repressed state of GA signaling, were rapidly up-regulated in apical buds of *Populus* (*P. tremula* × *P. alba*) upon transfer to shorter day light period (dormancy inducing condition)[30]. Induction of *REPRESSOR OF ga1-3* (*Rga*), which encodes a negative regulator of growth in the autumn and that of a *Ga 20-oxidase*, was reported during dormancy break in *P. tremula*[31]. A differential modulation of several TFs associated with growth, development, biotic and abiotic stress during WD (Fig. 2) suggested fine tuning of growth and developmental processes in response to environmental stress, which might be mediated through coordinated expression of TFs and their corresponding regulon (a group of genes controlled by a certain type of TF).

Genes unique to PAG and WD. A total of 818 and 249 unigenes were found to be exclusively expressed during PAG and WD, respectively (Supplementary Table S1). Some specific unigenes during PAG included *glycine decarboxylase* (C639578_124.0), *sucrose-6-fructosyltransferase* (C597661_210.0), *beta-galactosidase* (C712352_190.0, scaffold1188_150.4, scaffold14750_144.8, and scaffold17329_177.1), unigenes involved in chromatin modification (C642230_145.0, C638978_155.0) and maintenance of chromosomal structure (C674584_190.0) (Supplementary Table S1). Presence of these unigenes in tissues during PAG suggested the need to produce larger amounts of metabolites for the newly forming and dividing cells of the actively growing meristems during PAG. Several genes related to abiotic stress were present in tissues during WD (Supplementary Table S1). These were *Cbf-like protein 1* (scaffold22200_219.0), *serine/threonine protein phosphatase 2C* (C669110_74.0), *cytochrome P450* (C679574_59.0), *Mate efflux family protein* (C631054_30.0), *glycosyltransferase* (C628784_20.0), *proton-dependent oligopeptide transport family* (scaffold13797_24.0) *and annexin* (C629880_10.0). Wang *et al.*[32] reported induction of *Cbf-like protein 1* by low temperature in tea and suggested

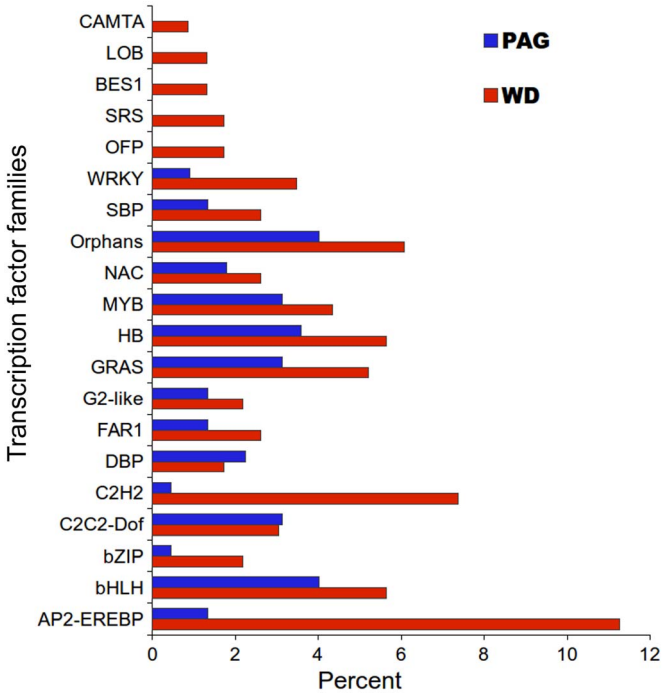

Figure 2 | Relative abundance and distribution of top 20 transcription factor (TF) families during the period of active growth (PAG) and winter dormancy (WD) for unigenes exhibiting significant differential expression. "Percent" on X-axis represents percent TF families out of total differentially expressed TF families in the tea transcriptome. Supplementary Table S2 has details on all the TF families up-regulated during PAG and WD. Full name of various TF families are expanded in Supplementary Table S6.

its role in cold responses. A peach *Cbf* increased cold hardiness as well as promoted short day-induced dormancy of apple trees[33]. *Serine/threonine protein phosphatase 2C*, is involved in stress sensing and signaling, while *cytochrome P450*, *MATE efflux family protein*, *glycosyltransferase* and *proton-dependent oligopeptide transport family*, and *annexin* are associated with detoxification and transport activities in the cell[26,30,34]. These stress responsive genes would help in maintaining cellular homeostasis during the environment of winters.

Additionally, *auxin signaling components* (C680134_27.0), *gibberellin 3-beta hydroxylase* (C669478_20.0) and *isopentenyl transferase* (C638160_20.0) were present in the tissues during WD. Up-regulation of *isopentenyl transferase* in the tissues during WD was one of the significant observations since it was shown to suppress leaf senescence[35]. Similar expression of the gene in tea might help inhibiting leaf senescence during WD (Supplementary Table S1).

In order to ascertain the relevance of RPKM-based expression values, quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) was carried out for randomly selected 19 genes. The expression patterns observed through the two different approaches were in agreement with each other displaying a significant correlation coefficient of 0.899 (p-value = 1.70 e^{-07}) and 0.862 (p-value = 2.13 e^{-06}) for first and second year, respectively (Fig. 3, Supplementary Table S5). Such values suggested a significant agreement between the expression patterns observed through the two different platforms (RPKM versus qRT-PCR)[36,37].

Analysis of biological processes during PAG and WD identifies modulation of senescence related unigenes. A worth noting point in the analysis of unigenes during PAG and WD was modulation of genes related to leaf senescence that would ultimately lead to leaf senescence. These genes were *cytokinin receptor 1* (*Cre1*), *auxin*

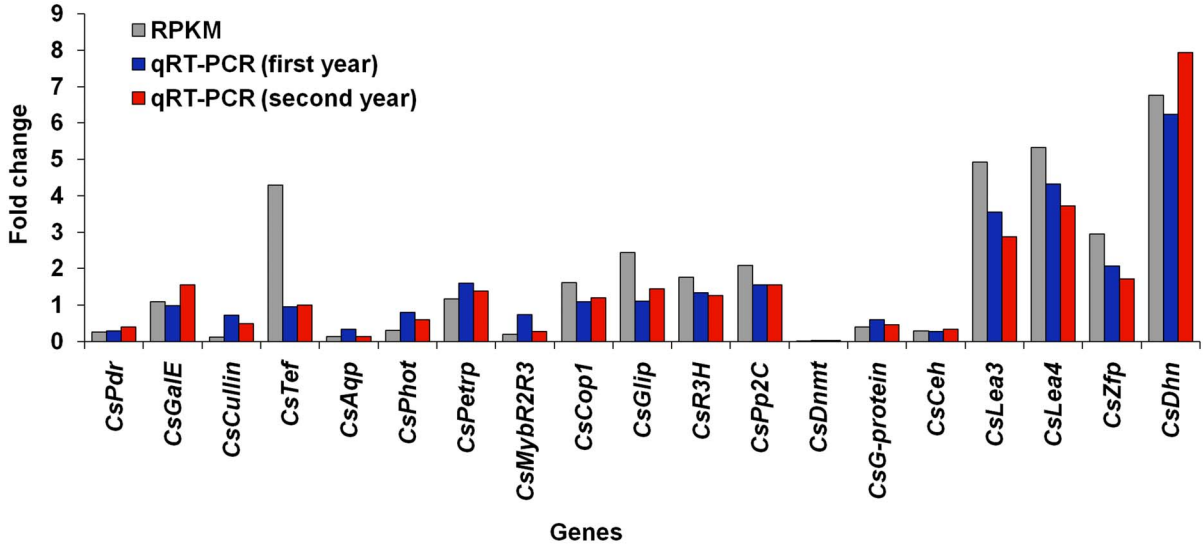

Figure 3 | Relationship of gene expression between reads per kilo base per million (RPKM) and quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) data obtained during winter dormancy as compared to the period of active growth for the randomly selected nineteen genes; qRT-PCR was performed using tissues for two consecutive years of the field grown tea plants. Correlation coefficient of RPKM and qRT-PCR for first and second year was 0.899 (p-value = 1.70 e^{-07}) and 0.862 (p-value = 2.13 e^{-06}), respectively. RPKM data, primers and qRT-PCR condition are detailed in Supplementary Table S5. Full name of various genes are expanded in Supplementary Table S6.

response factor 5 (*Arf5*), *auxin hydrogen transporter* (*Pin1*), *auxin hydrogen symporter* (*Pin2*), *ethylene response factor 2* (*Erf2*), *gibberellin 2-oxidase 1* (*Ga2-ox1*), *jasmonate o-methyltransferase* (*Jomt*), *polygalacturonase inhibiting protein 1* (*Pgip1*), *polygalacturonase inhibitor 1* (*Pgi1*), *polygalacturonase inhibitor 2* (*Pgi2*), *cellulase 2* (*Cel2*), and *polygalacturonase* (*Pg*) (Supplementary Table S1).

Cell wall degrading enzymes cellulase (CEL) and polygalacturonase (PG) are closely associated with disassembly and modification of the cell wall and participate in the senescence process[38]. Further, PG is regulated by polygalacturonase inhibitors (PGI)[39]. Leaf senescence also involves a network of hormone signalling pathways which may have indirect role as follows. ARFs are transcription factors that mediate responses to the plant hormone auxin[40,41]. Auxin and ethylene levels are shown to exhibit response analogous to leaf senescence[42]. PINs (*Pin1, Pin2*) are auxin transport factors that have several roles in plants including in modulating growth responses to environmental cues[43]. *Ga2-ox1* encodes for gibberellin oxidase that inactivates gibberellin and has an important role in the regulation of leaf senscence[44]. Cytokinin signals are perceived by histidine kinase CRE1 (a cytokinin receptor) and further relayed by a multistep variant of the two-component signaling system[45]. Increase in cytokinins and leaf senescence has a direct correlation[46]. Activation of *Jomt* expression leads to production of methyl jasmonate, which acts as (i) an intracellular regulator, (ii) a diffusible intercellular signal transducer, and (iii) an airborne signal that mediates intra- and inter-plant communications[47]. ERF proteins are involved in biosynthesis of ethylene and its production, which in turn affects leaf senscence[48]. Precocious leaf senescence was observed in transgenic Arabidopsis plants with enhanced expression of *AtErf4*, or *AtErf8*. *AtErf4* and *AtErf8* targeted the *EPITHIOSPECIFIER PROTEIN/EPITHIOSPECIFYING SENESCENCE REGULATOR* gene (a negative regulator of leaf senescence) and regulated the expression of many genes involved in the progression of leaf senescence[49].

RPKM data showed down-regulation of *CsCre1*, *CsArf5*, *CsPin1*, *CsPin2*, *CsErf2*, *CsJomt*, *CsCel2* and *CsPg* during WD. Whereas, *CsGa2-ox1*, *CsPgip1*, *CsPgi1* and *CsPgi2* exhibited up-regulation during WD (Fig. 4). RPKM and qRT-PCR based expression data were in accordance with each other with a correlation coefficient of 0.754 (p-value = 0.0046) (Fig. 4, Supplementary Table S5). As discussed elsewhere, this value of correlation coefficient is considered significant ensuring confidence in the two methods of gene expression[36,37]. Down-regulation of leaf senescence related genes during winters

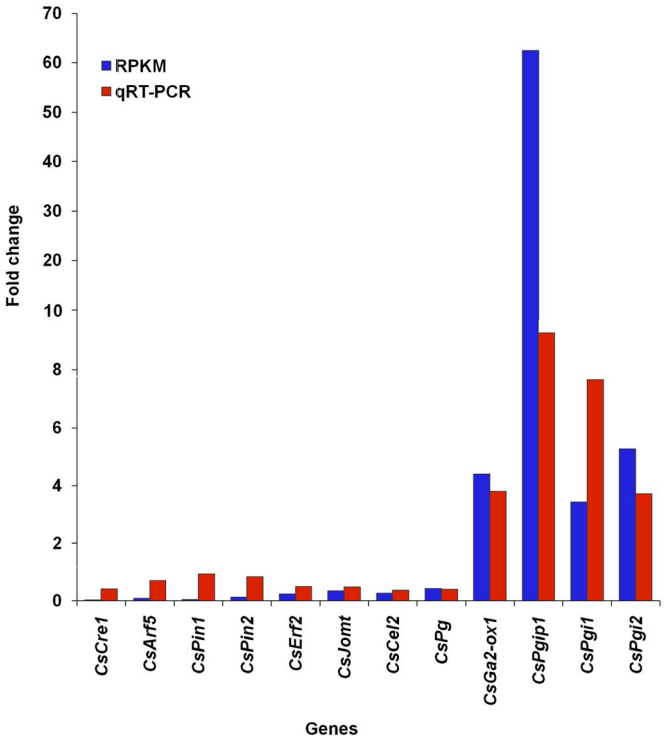

Figure 4 | Relative expression of various genes associated with leaf abscission in two and bud during winter dormancy as compared to the period of active growth based upon the data obtained by reads per kilo base per million (RPKM) values and validated by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR). Correlation coefficient of RPKM and qRT-PCR was 0.754 (p-value = 0.0046). RPKM data, primers and qRT-PCR condition is detailed in Supplementary Table S5. Full name of the genes are expanded in Supplementary Table S6.

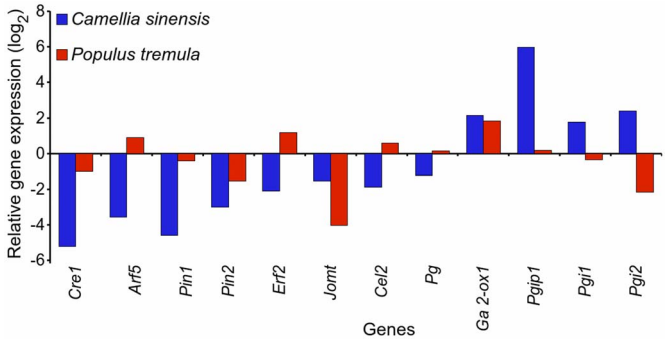

Figure 5 | Comparative analysis of expression of various genes associated with leaf abscission in *Camellia sinensis* and *Populus tremula*. Gene expression of *C. sinensis* was based upon reads per kilo base per million (RPKM; during winter dormancy as compared to the period of active growth) values (Supplementary Table S5), whereas gene expression for *P. tremula* was based on microarray data [during autumn senescence as compared to the period of active growth (before onset of senescence)] published by Anderson *et al.*[3] (Supplementary Table S5). Full name of genes are expanded in Supplementary Table S6.

ensures tea not to set in leaf senescence and hence leaf abscission is not observed.

Comparative analysis of leaf senescence related unigenes between tea and *P. tremula*. Unlike tea, *P. tremula* is a perennial deciduous tree and hence the above twelve leaf senescence associated unigenes were also studied in this tree species before and during autumn senescence (winters) using the microarray data of Anderson *et al*[3]. In contrast to the gene expression in tea, *Arf5*, *Erf2*, *Cel2*, and *Pg* showed up-regulation whereas, *Pgi1* and *Pgi2* exhibited down-regulation during autumn senescence in *P. tremula* (Fig. 5). Expression of *Cre1*, *Pin1*, *Pin2*, and *Jomt* was also down-regulated during autumn senescence in *P. tremula* similar to the expression recorded during WD in tea (Fig. 5). Expression of *Ga2-ox1* and *Pgip1* was up-regulated during autumn senescence in *P. tremula* which is in line with the expressions observed for these genes in tea during WD (Fig. 5). Particularly, up-regulation of *Cel2* and *Pg*, and down-regulation of *Pgi1* and *Pgi2* in *P. tremula* as compared to that in tea suggested that the former but not the tea has a tendency to abscise its leaves during winters.

Analysis of senescence related unigenes in polyethylene glycol (PEG) induced leaf abscission in tea. Polyethylene glycol (PEG-8000; 10%) was used to induce osmoticum-induced abscission of mature leaves e.g. at position 4 and 5 (leaf position was with reference to apical bud at "0" position; the leaf adjacent to apical bud was designated to be at position 1). Senescence was noticeable at 72 h of the treatment and the leaves abscised thereafter (Supplementary Fig. S6). PEG significantly affected relative electrolyte leakage (REL) (Fig. 6A) and relative water content (RWC) of the leaf tissue in a time dependent manner (Fig. 6B). PEG treatment (72 h) led to increase in REL by 226.09%, and a decrease in RWC by 36.96% as compared to the respective control value of the same time period. Gene expression data showed down-regulation of *CsCre1*, *CsArf5*, *CsPin1*, *CsPin2*, *CsErf2*, *CsCel2* and *CsGa2-ox1*, while up-regulation was observed for *CsJomt*, *CsPg*, *CsPgip1*, *CsPgi1* and *CsPgi2* at 24 h of the PEG treatment (Fig. 6C). *CsCre1*, *CsPin1*, *CsPin2*, and *CsErf2* continued to be down-regulated even at 48 h of PEG treatment; whereas *CsArf5*, *CsCel2* and *CsGa2-ox1* started exhibiting up-regulation along with *CsJomt*, *CsPgip1*, *CsPgi1* and *CsPgi2*. Increasing the PEG treatment time to 72 h led to up-regulation of *CsCre1, CsArf5*, *CsPin1*, *CsPin2*, *CsErf2*, Cs*Ga2-ox1*, *CsJomt, CsCel2* and *CsPg*, and down-regulation of *CsPgip1 CsPgi1* and *CsPgi2* as compared to the respective control (Fig. 6C). Gene expression data was in accordance to the observation of leaf retention upto 48 h followed by setting-in of senescence at 72 h of the PEG treatment. This experiment further strengthened our conclusion on association of the identified senescence related genes with leaf abscission in tea. Also, the data suggested that tea has mechanism of leaf abscission, but it does not operate during winter season.

To conclude, transcriptome analysis during the PAG and WD suggested operation of mechanisms that (i) permit tea to tolerate winter through expression of genes associated with stress tolerance, (ii) minimize growth during winters by down-regulation of genes involved in growth, development, protein synthesis, DNA processing, and cell division, and (iii) does not allow leaf abscission due to modulation of leaf abscission related genes during WD. Since the leaves are retained during winter season, tea develops the mechanisms of stress tolerance to tolerate the "harsh" conditions of winters and also slows down the molecular machinery of growth and development that is reflected as WD. On the contrary to situation in tea, expression of leaf senescence related gene homologues favored leaf abscission during winter season in deciduous tree *P. tremula*. PEG-induced leaf senescence not only validated the relevance of the identified mechanism that lead to leaf abscission, but also suggested their operation in tea when needed.

Methods

Plant material. TEENALI, an Assamica type of tea clone[4], growing in the experimental tea farm of the Institute was used for various experiments. Experiments were performed on two and a bud (apical bud and associated two leaves), which are the biologically active aerial portion of the plant[4] and also used for commercial production of tea[1]. Two and a buds were harvested during PAG (July; maximum temperature, 25 ± 2°C; minimum temperature, 20 ± 2°C) and winter season [December; maximum temperature, 15 ± 2°C; minimum temperature, 4 ± 2°C; during the period the tea was in the phase of WD wherein the growth of two and a bud is diminished]. Tissues were harvested between 9 to 11 am, immediately frozen in liquid nitrogen, and stored at −80°C until further use.

Library preparation, Illumina sequencing, *de novo* assembly and sequence clustering. Total RNA was isolated from tissues during PAG and WD as described previously[50]. Preparation of cDNA and transcriptome sequencing was performed essentially as described by Gahlan *et al.*[20]. Briefly, poly (A) mRNA was purified using Oligotex mRNA Midi Prep Kit (Qiagen, Germany), re-purified using mRNA-Seq 8 Sample Prep Kit (Illumina, USA), and reverse transcribed using SuperScript III (Invitrogen, USA) for the synthesis of first strand cDNA. Second-strand cDNA synthesis, cDNAs end repairing, and the Illumina adaptors ligation was performed using mRNA-Seq 8 Sample Prep Kit (Illumina, USA) as per the suggestions of the manufacturer. The products were sequenced [PE, 36 × 2 bp] on an Illumina Genome Analyzer IIx (Illumina, USA) using six and eight picomoles of libraries following the manufacturer's instructions.

PE reads from the two libraries were generated using CASAVA version 1.3 package in FASTq format. FilteR[20] was used to filter out poor quality reads, read trimming as well as for adapter removal as described previously[20]. Only those reads were retained which showed quality score of 30 or higher[20]. A minimum of 70% of the read nucleotides should pass the quality score of 30[20]. The obtained reads for different experimental conditions were merged before the assembling step. Evaluation of the assembly quality was also done by calculating N50, coverage, % transcripts having length > 1 kb, maximum length obtained and average length of the assembled transcript sequences.

De novo assembly was done using SOAP*denovo*[51]. The high quality reads were split into smaller fragments, the 'k-mers', to assemble the reads into contigs using de Bruijn graphs. K-mer size of 21 achieved the best balance between the number of contigs produced, coverage and average sequence length attained. The PE option of assembling with distance of 200 bp was applied. The same parameters were also used to build scaffold sequences by merging two contigs into single scaffold sequence that shared read pairs. The primarily assembled sequences were subjected to hierarchical clustering steps to curtail redundancy by merging of significantly overlapping contigs/scaffolds using TGICL-CAP3 and CD-HIT-EST at 90% similarity cut-offs. The final assembled sequences were searched against the NR database using BLASTX at E-value cutoff of E^{-05} to identify the unigenes. Dissimilar sequence clustering[20] was performed over the sequences returning the BLAST hit to cluster the assembled transcript sequences into single unigene representations and to curtail inflated representation of total unigenes for the assembled sequences. The quantification of genes abundance was measured by mapping the reads across the assembled unigene sequences following a well established protocol described previously[20]. The abundance of transcripts was measured using RPKM.

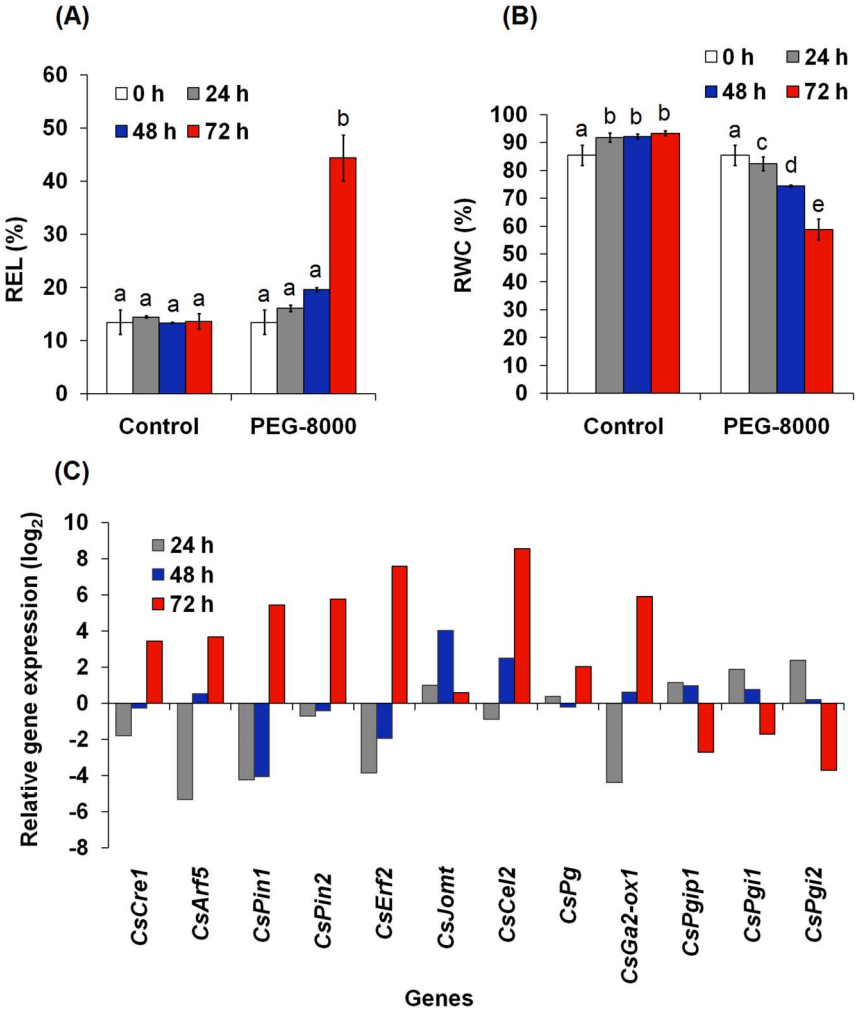

Figure 6 | Effect of 10% polyethylene glycol-8000 (PEG-8000) on physiological parameters and gene expression in leaf tissue of tea. Relative electrolyte leakage (REL) and relative water content (RWC) are shown in panels (A) and (B), respectively. Error bars are standard error of the mean of three biological replicates. Different letters above the bar show significant difference at $p < 0.05$. Relative expression of genes associated with leaf abscission in response to 10% PEG-8000 as compared to the corresponding control of the same time is shown in panel (C). Gene expression was analyzed by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR). Primer details and qRT-PCR conditions are detailed in Supplementary Table S5. Abbreviations of the genes are expanded in Supplementary Table S6.

Functional annotation and characterization of unigenes and DEUs. Assembled sequences were searched against UniProt databases (http://www.uniprot.org/downloads) and associated entries for gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Enzyme Commission (EC) (http://www.chem.qmul.ac.uk/iubmb/enzyme/) with a cut-off E-value of 10^{-1} to annotate these sequences. E-value of 10^{-1} allows identification of the most agreeable functional category. It captures even small functional domains/regions despite of poor overall sequence similarity. It reduces the chances of missing out of the functional annotation of the assembled sequences which otherwise might have been eliminated at stringent cut-off. Majority of GO, EC and KEGG-based annotation and statistics were performed using annotation tools Annot8r[52] and blast2GO[53]. Use of two different tools for annotation was used to remove the chances of false annotation of genes. Only those annotations were retained which were common in both the tools.

Plant Transcription Factor Database (http://plntfdb.bio.uni-potsdam.de) was used for identification and classification of transcriptional factors. GO term enrichment analysis was performed using agriGO[54] for hyper-geometric test. This enrichment analysis was performed to evaluate the enrichment of various GO categories for the unigenes having significant expression level for tissues during PAG and WD. Significant DEUs were identified using edgeR[23] tool in "R" Bioconductor package, with replicates obtained from six and eight picomoles libraries[19]. Significantly differentially expressed unigenes were identified using edgeR[23] package which compares the read count values of unigene for the two conditions and statistically evaluates the significant change in gene expression (Supplementary Fig. S7). The tool edgeR applies trimmed mean of M-values and considers dispersion of expression values around that mean to get the corrected representation of the sample. Significantly up- and down-regulated genes were those whose p-value was less than 0.05 (Supplementary Table S1). Genes with positive and negative log fold change (logFC) value are considered to be significantly up- and down-regulated genes, respectively.

Nineteen randomly selected unigenes were validated by qRT-PCR[55] in two and a bud harvested from the field grown tea bushes during PAG and WD for two consecutive years (Supplementary Table S5 has details on qRT-PCR including primers).

Twelve senescence related unigenes were also analyzed in *P. tremula* (a deciduous tree species) using the microarray data published by Anderson *et al.*[3]. Data for before and during autumn senescence were downloaded from Array Express[3]. Deduced amino acid sequences of the expressed sequence tag (EST) of *P. tremula* were retrieved from GenBank. Assembled unigenes of tea (twelve target genes) were searched using BLASTX against the downloaded *P. tremula* ESTs and the best matching top hit was assigned to the query as described previously[56]. Details of expression data of these twelve genes is mentioned in Supplementary Table S5.

Studies on polyethylene glycol induced leaf abscission. In a separate experiment shoot cuttings of clone TEENALI were collected during PAG and transferred to 150 ml deionized water in plant growth chamber set at 25 ± 3°C (growth temperature, GT). After 24 h, cuttings were transferred to deionized water (control) and 10% polyethylene glycol-8000 (PEG-8000; Sigma, USA), separately and housed in plant growth chamber set at GT (Supplementary Fig. S6).

Leaf at position 4 and 5 (leaf position was with reference to apical bud at "0" position; the leaf adjacent to apical bud was designated to be at position 1) were harvested at an interval of 24 h starting from 0 h till 72 h. Fourth and 5th leaves senesce during the period and abscise thereafter. Fourth leaf was used for gene expression analysis and estimating relative electrolyte leakage (REL), whereas 5th leaf was used for estimating relative water content (RWC). Our previous work[55] did show

REL to be a better estimate of osmotic stress in tea and hence REL was estimated for the same leaf that was selected for estimating gene expression. RWC was determined as described previously by Munné-Bosch and Peñuelas[57], and REL was estimated essentially as described by Blum[58]. Five leaf discs from fourth leaf of equal diameter were cut with cork borer for REL analysis and complete fifth leaf was taken for RWC analysis. Tissues harvested for gene expression were immediately frozen in liquid nitrogen and stored at −80°C for further analysis. Expression of twelve leaf senescence related genes was analyzed by qRT-PCR as described previously[55]. Each experiment had three separate biological replicates. Duncan's multiple range test ($p < 0.05$) was used to compare means post hoc using the software Statistica version 7.0 (Starsoft Inc. USA).

1. Barua, D. N. Seasonal dormancy in tea (*Camellia sinensis* L.). *Nature* **224**, 514 (1969).
2. Kumar, S., Paul, A., Bhattacharya, A., Sharma, R. K. & Ahuja, P. S. Tea: Present status and strategies to improve abiotic stress tolerance. In: Tuteja, N., Gill, S. S., Tiburcio, A. F. & Tuteja, R., (ed) *Improving Crop Resistance to Abiotic Stress*. Wiley-VCH Verlag GmbH & Co. pp 1401–1424; DOI: 10.1002/9783527632930 (2012).
3. Andersson, A. *et al*. A transcriptional timetable of autumn senescence. *Genome Biol.* **5**, R24 (2004).
4. Paul, A. & Kumar, S. Responses to winter dormancy, temperature, and plant hormones share gene networks. *Funct. Integr. Genomics.* **11**, 659–664 (2011).
5. Olsen, J. E. *et al*. Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. *Plant J.* **12**, 1339–1350 (1997).
6. Bhalerao, R. *et al*. Gene expression in autumn leaves. *Plant Physiol.* **131**, 430–442 (2003).
7. Wright, I. J. *et al*. The worldwide leaf economics spectrum. *Nature* **428**, 821–827 (2004).
8. Kloeke, A. E. E. O., Douma, J. C., Ordoñez, J. C., Reich, P. B. & van Bodegom, P. M. Global quantification of contrasting leaf life span strategies for deciduous and evergreen species in response to environmental conditions. *Global Ecol. Biogeogr.* **21**, 224–235 (2012).
9. Laycock, D. H. Latitude, day length and crop distribution. *Proceedings of the Tocklai Experiment Station, Jorhat, Assam, India for 1969* 1–9 (1969).
10. Nagar, P. K. Changes in endogenous abscisic acid and phenols during winter dormancy in tea [*Camellia sinensis* (L.) O. Kuntze]. *Acta Physiol. Plant.* **18**, 33–38 (1996).
11. Nagar, P. K. & Kumar, A. Changes in endogenous gibberellin activity during winter dormancy in tea [*Camellia sinensis* (L.) O. Kuntze]. *Acta Physiol. Plant.* **22**, 439–443 (2000).
12. Vyas, D. & Kumar, S. Purification and partial characterization of a low temperature responsive Mn-SOD from tea [*Camellia sinensis* (L.) O. Kuntze]. *Biochem. Biophys. Res. Commun.* **329**, 831–838 (2005).
13. Vyas, D. & Kumar, S. Tea [*Camellia sinensis* (L.) O. Kuntze] clone with lower period of winter dormancy exhibits lesser cellular damage in response to low temperature. *Plant Physiol. Biochem.* **43**, 383–388 (2005).
14. Vyas, D., Kumar, S. & Ahuja, P. S. Tea (*Camellia sinensis*) clones with shorter periods of winter dormancy exhibit lower accumulation of reactive oxygen species. *Tree Physiol.* **27**, 1253–1259 (2007).
15. Paul, A. & Kumar, S. *Dehydrin2* is a stress-inducible, whereas *Dehydrin1* is constitutively expressed but up-regulated gene under varied cues in tea [*Camellia sinensis* (L.) O. Kuntze]. *Mol. Biol. Rep.* **40**, 3859–3863 (2013).
16. Singh, K., Kumar, S. & Ahuja, P. S. Differential expression of *histone H3* gene in tea [*Camellia sinensis* (L.) O. Kuntze] suggests its role in growing tissue. *Mol. Biol. Rep.* **36**, 537–542 (2009).
17. Singh, K., Paul, A., Kumar, S. & Ahuja, P. S. Cloning and differential expression of *QM like protein homologue* from tea [*Camellia sinensis* (L.) O. Kuntze]. *Mol. Biol. Rep.* **36**, 921–927 (2009).
18. Paul, A., Lal, L., Ahuja, P. S. & Kumar, S. *Alpha-tubulin* (*CsTUA*) up-regulated during winter dormancy is a low temperature inducible gene in tea [*Camellia sinensis* (L.) O. Kuntze]. *Mol. Biol. Rep.* **39**, 3485–3490 (2012).
19. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Biol.* **18**, 1509–1517 (2008).
20. Gahlan, P. *et al*. *De novo* sequencing and characterization of *Picrorhiza kurrooa* transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics* **13**, 126 (2012).
21. Altschul, S. F. 1., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–10 (1990).
22. Marchler-Bauer, A. *et al*. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–229 (2011).
23. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
24. Horvath, D. P., Chao, W. S., Suttle, J. C., Thimmapuram, J. & Anderson, J. V. Transcriptome analysis identifies novel responses and potential regulatory genes involved in seasonal dormancy transitions of leafy spurge (*Euphorbia esula* L.). *BMC Genomics* **9**, 536 (2008).
25. Derory, J. *et al*. A Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). *New Phytol.* **170**, 723–738 (2006).
26. Jia, Y. *et al*. Subtractive cDNA libraries identify differentially expressed genes in dormant and growing buds of leafy spurge (*Euphorbia esula*). *Plant Mol. Biol.* **61**, 329–344 (2006).
27. Mazzitelli, L. *et al*. Co-ordinated gene expression during phases of dormancy release in raspberry (*Rubus idaeus* L.) buds. *J. Exp. Bot.* **58**, 1035–1045 (2007).
28. Perez-Rodriguez, P. *et al*. PInTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* **38**, D822–D827 (2010).
29. Fridborg, I., Kuusk, S., Robertson, M. & Sundberg, E. The Arabidopsis protein SHI represses gibberellin responses in *Arabidopsis* and barley. *Plant Physiol.* **127**, 937–948 (2001).
30. Ruttink, T. *et al*. A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* **19**, 2370–2390 (2007).
31. Druart, N. *et al*. Environmental and hormonal regulation of the activity-dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. *Plant J.* **50**, 557–573 (2007).
32. Wang, Y., Jiang, C. J., Li, Y. Y., Wei, C. L. & Deng, W. W. CsICE1 and CsCBF1: two transcription factors involved in cold responses in *Camellia sinensis*. *Plant Cell Rep.* **31**, 27–34 (2012).
33. Wisniewski, M., Norelli, J., Bassett, C., Artlip, T. & Macarisin, D. Ectopic expression of a novel peach (*Prunus persica*) CBF transcription factor in apple (*Malus* × *domestica*) results in short-day induced dormancy and increased cold hardiness. *Planta* **233**, 971–983 (2011).
34. Conde, A., Chaves, M. M. & Geros, H. Membrane transport, sensing and signaling in plant adaptation to environmental stress. *Plant Cell Physiol.* **52**, 1583–1602 (2011).
35. Gan, S. S. & Amasino, R. M. Inhibition of leaf senescence by autoregulated production of cytokinin. *Science* **270**, 1986–1988 (1995).
36. Wang, J. *et al*. RNA-seq based identification and mutant validation of gene targets related to ethanol resistance in cyanobacterial *Synechocystis* sp. PCC 6803. *Biotechnol. Biofuels.* **5**, 89 (2012).
37. Xu, Li. *et al*. Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. *J. Exp. Bot.* **62**, 5607–5621 (2011).
38. Horton, R. F. & Osborne, D. J. Senescence, abscission and cellulase activity in *Phaseolus vulgaris*. *Nature* **214**, 1086–1088 (1967).
39. Mattei, B. *et al*. A Secondary structure and post-translational modifications of the leucine-rich repeat protein PGIP (polygalacturonase-inhibiting protein) from *Phaseolus vulgaris*. *Biochemistry* **40**, 569–576 (2001).
40. Ellis, C. M. *et al*. *AUXIN RESPONSE FACTOR1* and *AUXIN RESPONSE FACTOR2* regulate senescence and floral organ abscission in *Arabidopsis thaliana*. *Development* **132**, 4563–4574 (2005).
41. Lim, P. O. *et al*. Auxin response factor 2 (ARF2) plays a major role in regulating auxin-mediated leaf longevity. *J. Exp. Bot.* **61**, 1419–1430 (2010).
42. Abeles, F. B. & Rubinstein, B. Regulation of ethylene evolution and leaf abscission by auxin. *Plant Physiol.* **39**, 963–969 (1964).
43. Cin, V. D., Velasco, R. & Ramina, A. Dominance induction of fruitlet shedding in *Malus* × *domestica* (L. Borkh): molecular changes associated with polar auxin transport. *BMC Plant Biol.* **9**, 139 (2009).
44. Hedden, P. & Phillips, A. L. Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci.* **5**, 523–530 (2000).
45. Bishopp, A., Mahonen, A. P. & Helariutta, Y. Signs of change: hormone receptors that regulate plant development. *Development* **133**, 1857–1869 (2006).
46. van Staden, J. Changes in endogenous cytokinin levels during abscission and senescence of *Streptocarpus* leaves. *J. Exp. Bot.* **24**, 667–671 (1973).
47. Seo, H. S. *et al*. Jasmonic acid carboxyl methyltransferase: A key enzyme for jasmonate-regulated plant responses. *Proc. of Nat. Acad. Sci. USA* **98**, 4788–4793 (2001).
48. Zhang, Z. J., Zhang, H. W., Quan, R. D., Wang, X. C. & Huang, R. F. Transcriptional regulation of the ethylene response factor LeErf2 in the expression of ethylene biosynthesis genes controls ethylene production in tomato and tobacco. *Plant Physiol.* **150**, 365–377 (2009).
49. Koyama, T. *et al*. A regulatory cascade involving class II ETHYLENE RESPONSE FACTOR transcriptional repressors operates in the progression of leaf senescence. *Plant Physiol.* **162**, 991–1005 (2013).
50. Muoki, R. C., Paul, A., Kumari, A., Singh, K. & Kumar, S. An improved protocol for the isolation of RNA from roots of tea (*Camellia sinensis* (L.) O. Kuntze). *Mol. Biotechnol.* **52**, 82–88 (2012).
51. Li, R. *et al*. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).
52. Schmid, R. & Blaxter, M. L. annot8r: GO, EC and KEGG annotation of EST datasets. *BMC Bioinformatics* **9**, 180 (2008).
53. Conesa, A. *et al*. "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research" *Bioinformatics* **21**, 3674–3676 (2005).
54. Du, Z., Zhou, X., Ling, Y., Zhang, Z. & Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70 (2010).
55. Muoki, R. C., Paul, A. & Kumar, S. A shared response of *thaumatin like protein*, *chitinase*, and *late embryogenesis abundant protein3* to environmental stresses in tea [*Camellia sinensis* (L.) O. Kuntze]. *Funct. Integr. Genomics* **12**, 565–571 (2012).

56. Reichmann, J. *et al.* Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. *PLoS Comput. Biol.* **8**, e1002486 (2012).
57. Munné-Bosch, S. & Peñuelas, J. Drought-induced oxidative stress in strawberry tree (*Arbutus unedo* L.) growing in Mediterranean field conditions. *Plant Sci.* **166**, 1105–1110 (2004).
58. Blum, A. Cell membrane stability as a measure of drought and heat tolerance in wheat. *Crop Sci.* **21**, 43–47 (1981).

Acknowledgments

Authors acknowledge Council of Scientific and Industrial Research (CSIR) for providing financial assistance through network projects BSC-0107, BSC-0109 and BSC-0121. Authors thank the Director CSIR-IHBT for help in providing necessary facilities. A.P. and S.S. are grateful to CSIR for awarding Junior and Senior Research Fellowships. S.B. and A.J. gratefully acknowledge CSIR for providing assistantship. We are grateful to Vandna Chawla for her help during digital expression analysis. The manuscript represents CSIR-IHBT communication number 3556.

Author contributions

A.P. carried out experiment at PAG and WD, prepared cDNA library for Illumina sequencing, performed expression analysis and drafted the manuscript. A.J. performed read generation, process of assembling, clustering, homology searching, annotation, CDD search, entire computational analysis and drafted the manuscript. S.B. and S.S. performed polyethylene glycol-mediated experiment, qRT-PCR analysis, literature survey and organized the manuscript. R.S. conceived, planned, developed and tested the protocols for the entire computational part of this study, performed reads based expression analysis and associated studies, developed the algorithm and tool for dissimilar sequence clustering, supervised the entire computational part of the study and drafted the manuscript. S.K. conceived the study, designed the experiments, guided on transcriptome generation and wet-lab experiments, analysed, interpreted and integrated computational and wet-lab results and coordinated the study. S.K. also drafted and finalized the manuscript. All authors have read and approved the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Paul, A. *et al.* RNA-seq-mediated transcriptome analysis of actively growing and winter dormant shoots identifies non-deciduous habit of evergreen tree tea during winters. *Sci. Rep.* **4**, 5932; DOI:10.1038/srep05932 (2014).