Plant Omics Journal POJ

POJ 9(4):270-280 (2016) ISSN:1836-3644
DOI: 10.21475/poj.16.09.04.p7882

De novo transcriptome sequencing and analysis of *Hydrilla verticillata* (L.f.) Royle

Etika Goyal[1¶], Singh Amit Kumar[1¶], Ravi Shankar Singh[1#2], Ajay Kumar Mahato[1] and Kumar Kanika[1*]

[1]Biotechnology and Climate Change Laboratory, ICAR-National Research Centre on Plant Biotechnology, New Delhi, India
[#2]Current address: Department of Plant Breeding and Genetics, Bihar Agricultural University, Sabour, Bihar, India

***Corresponding author: kumarkanika@rediffmail.com**
[¶]These authors contributed equally to this work.

Abstract

Hydrilla verticillata (L.f.) Royle, an aquatic plant, best documented example of an inducible C_4 photosynthetic system, which concentrates CO_2 in the chloroplasts without enzymatic compartmentation in mesophyll and bundle sheath cells. *H. verticillata* is a facultative C_4 plant, which shifts from C_3 to C_4 photosynthesis under certain conditions, but lacks Kranz anatomy. Little is known about the molecular changes required for the transition to the C_4 carbon concentrating mechanism (CCM). To gain insight into the processes that are involved in the C_3 to the C_4 transition, high throughput transcriptome sequencing and analysis were carried out using 454-GS FLX Titanium technology. A total of 533,595 reads were obtained after quality filtering. From these reads, 1,813 "Expressed sequence tags (ESTs)" were generated yielding 1,538 unigenes. Almost 95% of these unigenes aligned to the "Non-redundant (nr)" database, whereas 4.88% did not show homology to any other known gene. "Reads per kilobase per million sequenced reads (RPKM)" analysis indicated that 283 unigenes were up-regulated and 383 unigenes were down-regulated in the C_4 state as compared to the C_3 state. A number of transcription factors found in our study were involved in photosynthesis. These consisted of, HAP3/NF-YB, AP/EREBP and C2H2 family. GO, COG and KEGG analysis indicated that many genes were involved in energy metabolism, nucleotide metabolism, photosynthesis, signal transduction, stress responses and are worthy of further investigation. This transcriptome analysis is an important first step towards understanding the molecular changes underpinning the transition from the C_3 to the C_4 photosynthesis in *H. verticillata*. It could also be a useful genomic resource for engineering C_4 photosynthetic CCM in C_3 crops.

Keywords: *Hydrilla verticillata*, Transcriptome, Next generation sequencing, 454 pyrosequencing.
Abbreviations: HvC, *Hydrilla* control sample; HvT, *Hydrilla* treated sample; COG, Clusters of Orthologous Groups; GO, Gene Ontology; SRA, Sequence Read Archive.

Introduction

C_4 photosynthesis is a composite biological trait that enables plants to accumulate biomass at a much faster rate and thrive even in adverse environments compared to C_3 plants (Brautigam et al., 2011). C_3 plants such as rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), soybean (*Glycine max* L.) and potato (*Solanum tuberosum* L.) assimilate atmospheric CO_2 directly through the C_3 photosynthetic pathway. C_4 plants such as maize (*Zea mays* L.), sugarcane (*Saccharum officinarum* L.) and sorghum [(*Sorghum bicolor* L.) *Moench*] have a C_4 photosynthetic pathway in addition to the C_3 pathway (Miyao et al., 2011).

The most salient features of C_4 plants are the presence of 'Kranz' anatomy and the cell-specific compartmentation of C_4 enzymes (Hatch, 1988). Compartmentation, chloroplast differentiation and structural adaptations are considered prerequisites for C_4 photosynthesis. In contrast, many aquatic autotrophs possess different modes of adaptation to cope with limited CO_2 and high O_2 conditions, including various CO_2 concentrating mechanism (CCM) and the capacity to utilize HCO_3^- for photosynthesis (Bowes and Salvucci, 1989).

However, plants like *Hydrilla verticillata* and *Borszczowia aralocaspica* prove that 'Kranz' anatomy is not essential for C_4 photosynthesis and that C_4 metabolism can function even within single cell (Voznesenskaya et al., 2001; Edwards et al., 2004). Magnin et al., (1997) and Reiskind et al., (1989) established that, there is an intracellular compartmentalization of C_4 and Calvin cycle (Reiskind et al., 1997). The carbon concentrating mechanisms were able to concentrate inorganic CO_2 up to 400 mmol m^{-3} in the chloroplast.

H. verticillata is an aquatic macrophyte that can shift from C_3 to C_4 photosynthesis under low CO_2 conditions, without its leaf cells undergoing any structural modifications (Rao et al., 2006a; Rao et al., 2002; Estavillo et al., 2007). During this transition, genes encoding C_4-specific isozymes of phosphoenolpyruvate carboxylase (PEPC), pyruvate orthophosphate dikinase (PPDK), NADP-malic enzyme (NADP-ME) and carbonic anhydrase (CA) are up-regulated (Rao et al., 2006a; Rao et al., 2002; Estavillo et al., 2007; Brautigam et al., 2014; Reiskind and Maberly, 2014). In *H. verticillata* there are three isoforms of *PEPC*, of which only

one isoform is up regulated in C_4 leaves and has C_4-like isoform kinetics (Rao et al., 2008). Two isoforms of PPDK and three of NADP-ME were also detected. One isoform of NADP-ME from the chloroplast was upregulated in light. This isoform had kinetic characteristics intermediate between C_4 and C_3 isoform of terrestrial plants (Estavillo et al., 2007; Rao et al., 2006b).

Higher temperatures and elevated CO_2 level can reduce Rubisco activation (Kobza and Edwards, 1987; Sage, 1990; Brandner and Salvucci, 2000). In rice leaves rubisco de-activation triggered by increase in temperature is greater under higher CO_2, conditions (Vu et al., 1997). If the earth's surface temperature continues to increase with rising atmospheric CO_2 there will be a need for agronomic crops with improved heat tolerance. There is a need for developing genotypes, suited to cope with global climate changes. C_4 photosynthesis is advantageous when limitations on carbon acquisition are imposed, by high temperature, drought and saline conditions (Voznesenskaya et al., 2001). Further, C_4 plants also have better nitrogen-use efficiency as compared to C_3 plants, since the rate of photosynthesis per unit nitrogen in the leaf is increased as C_4 plants need less of Rubisco enzyme. This enzyme works more efficiently under high CO_2 concentrations in the bundle sheath cells unlike mesophyll cells in C_3 plants. It is by far the most abundant protein in leaves of C_3 plants (Oaks, 1994; Gowik and Westhoff, 2000). This has led to attempts to introduce components of C_4 photosynthetic pathways in C_3 crops, in order to increase productivity, especially under climate change scenario (Sheehy et al., 2000; Matsuoka et al., 2001). Introduction of all the characteristics of C_4 photosynthesis to C_3 crop plant like rice, will increase its photosynthetic potential and enable us to reap the net agronomic benefit in term of yield (Reynolds et al., 2010). An International Consortium is attempting to establish C_4 pathway in rice (Furbank et al., 2009). To understand the complexity of the biochemical and anatomical traits necessary for this mechanism to operate is daunting and the minimal set of necessary genes is also unknown.

H. verticillata is a good system to decipher the minimal components required to run C_4 photosynthesis without 'Kranz' anatomy. The inducible, single cell C_4 system in *H. verticillata* is a minimalistic C_4 model, which has led to different attempts to introduce a similar C_4 pathway into the mesophyll cells of C_3 cereals (Miyao, 2003; Hausler et al., 2002; Matsuoka et al., 2001; Miyao et al., 2011). This plant is a member of the hydrocharitaceae family and could provide insights into early events in the development and evolution of C_4 photosynthesis (Rao et al., 2002).

Next generation sequencing (NGS) is a powerful tool, that sequences the whole genome and quantifies gene expression (Metzker, 2010). NGS enables *de novo* transcriptome sequencing and analysis of non-model species without annotated genomes. This has led to the expansion of databases related to various biological processes of even non-model species. The 454 GS FLX genome sequencer provides quarter million sequence reads of nearly 200-400 bases in each run, from a cDNA template generated from mRNA (http://www.454.com/) (Metzker, 2010). These reads can be mapped onto a closely related reference to quantify the number of reads matching a gene locus, thereby providing a measure of transcript abundance (Flicek and Birney, 2009; Brautigam and Gowik, 2010).

In the present study, the transcriptome profile of *H. verticillata* during elevated temperature was generated using NGS technology. A number of genes/transcription factors that may be involved in the transition from the C_3 to C_4 photosynthetic pathway were putatively identified by *de novo* transcriptome analysis. Functional annotation and classification analysis of reads indicated that, many genes were related to photosynthesis, signal transduction, stress responses and various putative metabolic pathways.

De novo transcriptome sequencing and assembly

We followed the transition from the C_3 to the C_4 photosynthetic state prior to NGS, by monitoring the expression levels of the *H. verticillata PEPC* isoform-2 gene (*PEPC-2*) using semi-quantitative RT-PCR (Table S1). The expression level of *PEPC-2* gene increased on day 8 sample kept at 34°C. This increase in band intensity was confirmed by qRT-PCR using *PEPC 2* gene specific primer (Fig. S1). Therefore, the day 8 sample kept on 34°C was used as the treatment (transition towards C_4 state; HvT) and day 8 sample kept at 20°C was used as control (present in C_3 state; HvC). To develop a comprehensive understanding of the cues involved in the transition from the C_3 to the C_4 photosynthetic phase in *H. verticillata*, two cDNA libraries were prepared using mRNA samples from shoots of HvC (control) and HvT, respectively. Equimolar amounts of RNA from three biological replicates were pooled together prior to the cDNA synthesis. These samples were subjected to pair-end read (PE) sequencing using the Roche 454 sequencing platform. The sequencing run generated a total of 599,069 raw reads. After removal of adaptor sequences, a total of 533,595 high quality reads having 215,200,000 bp were obtained. An overview of the sequencing and assembly process is presented in Table 1. All the raw transcriptome data have been deposited at the Sequence Read Archive (SRA) of the National Centre for Biotechnology Information (NCBI) with the accession numbers SRR1300831 and SRR1300832for HvC and HvT, respectively. By using the Newbler assembler (V.2.6), a total of 1,831 ESTs were generated. Using Cd-hit (V.4.6.1) software (Li and Godzik, 2006), the ESTs were further assembled into 1,538 unigenes, with a minimum unigene length of 70bp, a maximum length of 5,300 bp and an average length of 993 bp. The size distribution of the unigenes is shown in Fig. S2A. Among these unigenes, 1,362 (88.56%) were longer than 500 bp, and 494 (36.27%) of these were longer than 1,000 bp. The GC contents of the putative genes were also examined as it can provide information about their evolution, structure (intron size and number), regulation and DNA stability. The average GC content of *H. verticillata* unigenes was 50.9%, which is in the range of GC levels of coding sequences in monocots (50-60%) (Kuhl et al., 2004) (Fig. S2B).

Gene sequence annotation and classification

A sequence similarity search was conducted using Blast2GO software against the NCBI nr protein database via Basic Local Alignment Search Tool (BLAST) algorithm specifying E-values $1.0E^{-3}$ followed by functional annotation (Conesa and Gotz, 2008; Gotz et al., 2008; Myhre et al., 2006). Among the assembled unigenes, 95.12% (1,463 unigenes) were aligned to the nr database (Table S2). These assembled unigenes aligned mainly to *Medicago truncatula* (5.33%), *Magnetospirillum gryphiswaldense* (3.28%), *Oryza sativa* (3.00%), *Oxytricha trifallax* (6.83%) and *Vitis vinifera* (5.33%) (Fig. 1).

GO classification revealed that, out of 1,463 nr hits, a total of 786 unigenes were categorised into 27 functional groups (Fig. 2), including biological process, cellular component and molecular function. In total, 5,622 GO assignments were

obtained, with 1,985 (35.31%) for biological process, 1,720 (30.59%) for cellular component and 1,043 (18.55%) for molecular function at level 2. In the molecular function category, binding (39.12%) and catalytic activity (37.39%) were predominant. In the cellular component category, cell (36.22%) and organelle (31.51%) were the most prominent classes. In the biological process category, most of the unigenes were involved in metabolic processes (29%) and cellular processes (22.97%) (Table S3).

Finally, to evaluate the completeness of the *H. verticillata* transcriptome and the effectiveness of the annotation process, all the unigenes were submitted to the Clusters of Orthologous Groups (COG) database for functional prediction and classification. COG classification of 1,463 nr hits indicated that 1,394 (95.28%) unigenes were clustered into 24 functional categories. Some unigenes had multiple COG functions, therefore, a total of 1,461 functional annotations were produced. Among the 24 functional categories, "energy production and conversion" (272, 18.62%) represented the largest group, followed by "translation" (188, 12.87%), "general functional prediction only" (186, 12.73%), "post translational modification, protein turnover, chaperone function" (134, 9.17%), "Inorganic ion transport and metabolism" (104, 7.12%), "carbohydrate metabolism and transport" (93, 6.37%) and "amino acid metabolism and transport" (87, 5.95%). The smallest groups were "RNA processing and modification" (3, 0.20%) and "defence mechanism" (9, 0.62%) (Fig. 3).

KO (KEGG Orthology) numbers were assigned to 609 unigenes that comprised only two contigs and 607 isotigs, which were involved in 216 different pathways. A summary of the sequences involved in these pathways is described in Table S4. The six hundred and nine KEGG annotated unigenes were categorised into six different functional groups mainly "metabolism", "genetic information processing", "cellular process" and "environmental information processing" (Table 2).The pathways represented by the KEGG annotated unigenes were mostly involved in Ribosome (ko03010), Photosynthesis (ko00195), Oxidative phosphorylation (ko00190), Carbon metabolism (ko01200), Carbon fixation in photosynthetic organisms (ko00710), Biosynthesis of amino acid (ko01230) and Glycolysis/ Gluconeogenesis (ko00010) (Fig. 4, Table 2).

Analysis of differentially expressed transcripts and identification of transcription factors

To study the expression level of unigenes in HvC and HvT, the number of clean reads was compared between the libraries for each of the 1,538 assembled unigenes through RPKM analysis (Mortazavi et al., 2008). A total of 666 unigenes were differentially expressed between HvC and the HvT (Fig. 5). Out of 666 unigenes, 283 unigenes were up-regulated in HvT and down regulated in HvC while 383 unigenes were down-regulated in HvT and upregulated in HvC (Table S5). Interestingly, of the 1,463 unigenes under investigation, 204 and 186 unigenes were uniquely expressed in HvT and HvC, respectively. All the assembled unigenes were analysed for the presence of transcription factors using the PlantTFcat online tool (http://plantgrn.noble.org/ PlantTFcat). A total of 44 unigenes, representing 2.86% of the transcriptome, belonged to 18 putative transcription factor (TF) families (Table S6) and those are represented by "Hap3/NF-YB", "C2H2" and AP2-EREBP TF family. These transcription factors are reported to play role in abiotic stress tolerance and photosynthesis improvement (Saibo et al., 2009).

In silico identification of SSRs markers

All the unigenes were used to mine potential SSRs to develop SSR markers. In total, 1041 SSRs were identified by SSR Locator in 650 (42.26%) unigenes of *H. verticillata*. Out of these unigenes, 170 unigenes contained more than one SSR (Table 3). Hexa-nucleotide repeats (77.33%) were the most abundant SSRs in the transcriptome. The second major class was tri-nucleotide (14.32%) and the remaining repeat motifs were mononucleotide (0.38%), dinucleotide (1.83%), tetranucleotide (5.28%) and pentanucleotide (0.86%) (Fig. 6a, 6b).

SSRs with two tandem repeats (76.37%) were the most common, followed by four, three, five and six tandem repeats. The dominant repeat motif in all the SSRs was AAGCGA/TCGCTT (33, 3.17%), followed by AAAATC/GATTTT (15, 1.44%), CTC/GAG (15, 1.44%), CCG/CGG (13, 1.25%) and GGA/TCC (12, 1.15%) (Table S7). The analysis of GC content of microsatellites indicated that most of the SSRs were rich in GC content, except for mono- and penta- nucleotide repeats (Fig. 6c). Focusing on tri- and hexa- nucleotide repeats, SSRs with 50%-100% GC content were twice as prevalent as those with only 0-50% GC content. These SSRs can be used as functional markers in various genetic studies in the future. To make these SSR markers useful, we employed Primer 3 to design primer pairs for each SSR. In total, 872 primer pairs for 1,041 (83.77%) SSRs were designed from the microsatellites with sufficient flanking sequences (Table S8).

Discussion

H. verticillata, an aquatic plant, is the best documented example of an inducible C_4 photosynthetic system that concentrates CO_2 in the chloroplasts without enzymatic compartmentation in mesophyll and bundle sheath cells. This plant lacks 'Kranz' anatomy and is a facultative C_4 plant that shifts from C_3 to C_4 photosynthesis under certain conditions. The primitive types of C_4 photosynthetic system without 'Kranz' anatomy in *H. verticillata* and *Borszczowia aralocaspia* (Chenopodiaceae) suggests that C_4 photosynthetic system could run in single mesophyll cells (Freitag and Stichler, 2000; Magnin et al., 1997). Information at the molecular level related to the transition from the C_3 photosynthetic pathway to the C_4 photosynthetic pathway is not available. To gain insight into the processes that are involved in the transition from the C_3 mode to the C_4 mode of photosynthesis in *H. verticillata*, high throughput transcriptome sequencing and analysis was carried out in this study using 454-GS FLX Titanium technology.

Inducible photosynthetic system of *H. verticillata* provides an excellent opportunity to study the minimum essential biochemical elements required to operate a C_4 photosynthetic system. Exceptionally all the enzymes and metabolite transporters involved in the C_4 pathway also occur in C_3 plants, although at much lower activities and having different tissue specificities (Hausler et al., 2002). Therefore, the introduction of individual C_4 cycle genes might perturb the primary metabolism, and also in certain cases secondary metabolism (Hausler et al., 2002). Advent of genetic engineering in crop plants has raised hope for introduction of C_4 characteristics in C_3 plants by genetic manipulation of rice, tobacco and potato (Matsuoka et al., 2001). The facultative nature of *H. verticillata* also enables us to study the genes involved both in the C_3 and the C_4 states and to identify the differences in the expression of genes and

Table 1. Summary of 454 transcriptome sequencing and assembly for *H. verticillata.*

Item	Library	Number	Total Bases (Mb)
Raw reads	HvC	284,588	112.8
	HvT	314,481	123.5
Clean reads	HvC	250,969	101.6
	HvT	282,626	113.6
Contigs			
No. of Contigs (n)		71	-
Average Length (bp)		1,190.76	-
Maximum Length (bp)		4,177	-
Minimum Length (bp)		516	-
Singletons			
No. of Singletons (n)		1,467	-
Average Length (bp)		983.23	-
Maximum Length (bp)		5,300	-
Minimum Length (bp)		70	-

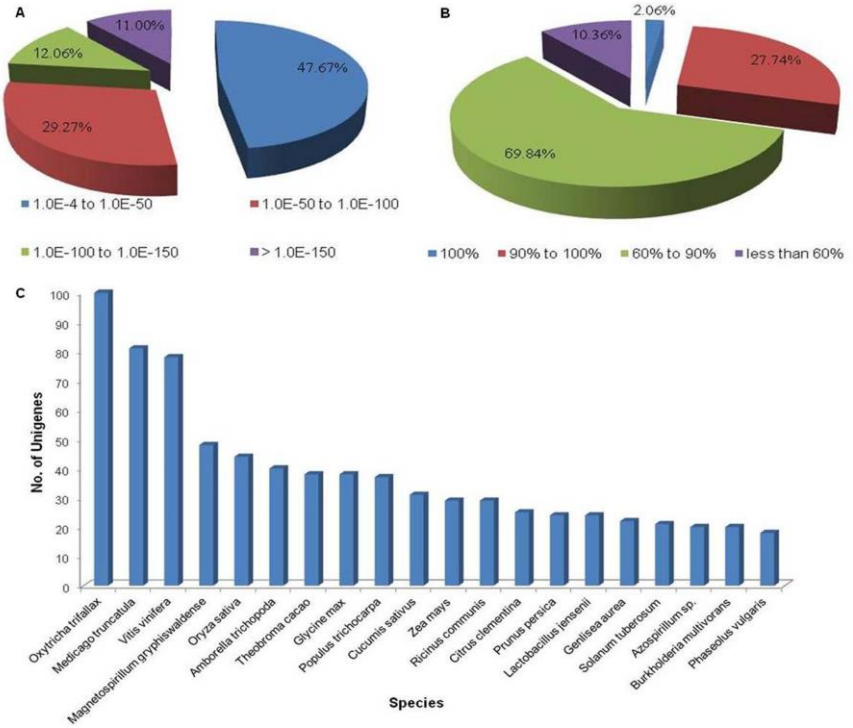

Fig 1. Characteristics of the homology search of the *H. verticillata* transition phase unigenes. (a) E-value distribution of the BLASTX hits against the nr protein database for each unigene, using an E-value cut-off of $10E^{-3}$; (b) Similarity distribution of the top BLAST hits for each unigene; (c) Number of unigenes matching the top 20 species using BLASTX.

variations in the regulatory and catalytic domains of the protein products.

One of the three primary carboxylases initiating the C_4 CCM in *H. verticillata PEPC* isoform 2 is differentially expressed in C_4 tissue (Rao et al., 2002; Magnin et al., 1997). In the present study, the expression level of the *H. verticillata PEPC-2* expressed in HvT was monitored to follow the transition from the C_3 to the C_4 photosynthetic state prior to NGS. The *PEPC-2* gene transcript level had exhibited an increase in expression on day 8 indicating an induction of the C_4 CCM under modified growth conditions. We specifically harvested samples kept at 34°C on day 8 (HvT) (Fig. S1).

To develop a comprehensive understanding of the cues involved in the transition from the C_3 to C_4 photosynthetic phase in *H. verticillata*, two cDNA libraries HvC and HvT were prepared. The pair-end read (PE) sequencing generated a total of 599,069 raw reads and a total of 533,595 high quality reads (215,200,000 bp). From these reads a total of 1,831 ESTs and 1,538 unigenes, were obtained. Among the assembled unigenes, 95.12% (1,463 unigenes) aligned to the nr database (Table S2). Among these aligned unigenes in nr protein database, 52.33% of these genes indicated high homology with the gene sequences in the database, having E-values less than 1.0 E^{-50}. The remaining 47.67% had an E-value between 1.0 E^{-4} to 1.0 E^{-50} (Fig. 1a). The similarity distribution showed that 42.76% of the aligned unigenes had a similarity value higher than 85%, whereas 10.36% of these genes had a similarity value lower than 60% (Fig. 1b). For species distribution, approximately 23.99% of the total unigenes matched with sequences from the five top-hit species, i.e., *Medicago truncatula*, *Magnetospirillum gryphiswaldense*, *Oryza sativa*, *Oxytricha trifallax* and *Vitis vinifera*. The twenty top-hit species based on nr annotation are shown in Fig. 1c. Among the unique sequences derived from these unigenes, coding sequences exhibited homology to sequences for proteins such as "cytochrome P450", "transcript antisense to rRNA protein", "cell wall associated

Table 2. KEGG classification of *H. verticillata* unigenes.

KEGG categories	No. of unigenes	KEGG categories	No. of unigenes
Metabolism	386	Environmental Information Processing	53
Carbohydrate Metabolism	111	Signal Transduction	109
Energy Metabolism	232	Organismal System	102
Lipid Metabolism	20	Immune System	28
Nucleotide Metabolism	12	Endocrine System	49
Amino Acid Metabolism	52	Circulatory System	22
Metabolism of other Amino Acids	9	Digestive System	13
Metabolism of Cofactors and Vitamins	15	Excretory System	7
Metabolism of Terpenoids& Polyketides	8	Nervous System	35
Biosynthesis of other Secondary Metabolites	13	Sensory System	22
Xenobiotics Degradation & Metabolism	27	Development	6
Genetic Information Processing	109	Environmental Adaption	17
Transcription	11	Human Disease	209
Translation	109	Cancer	39
Folding, Sorting & Degradation	27	Immune Disease	15
Replication & Repair	3	Neurodegenerative Disease	114
Cellular Process	50	Substance Dependence	21
Transport & Catabolism	40	Circulatory Disease	21
Cell Growth & Death	25	Endocrine &Metabolic Disease	30
Cell Communication	36	Infectious Disease	136
		Total	609

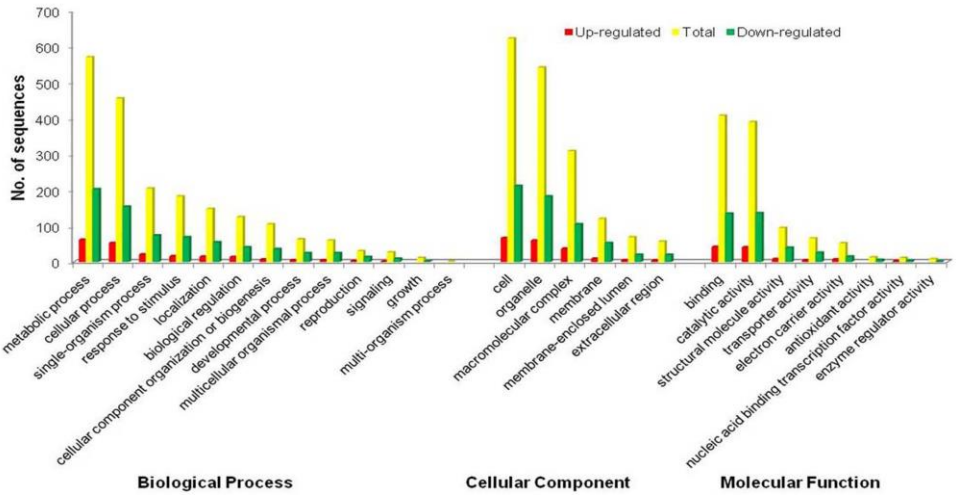

Fig 2. Gene Ontology classification of *H. verticillata* transcriptome and identification of differentially expressed genes in the C_3 and C_4 photosynthetic pathway. The Y-axis represents number of unigenes and the X-axis shows the GO categories.

hydrolase", "senescence protein", "amino transferase II", "thiazol biosynthesis enzyme", "Zinc finger", "serine hydroxyl methyl transferase", "NADH dehydrogenase", "ascorbate peroxidase esterase lipase", "tpr domain" and "chaperone protein chloroplastic like", etc. In the present study, the above mentioned putative functional transcripts that were identified can provide leads for future investigations.

Functional annotation

GO analysis provides a controlled vocabulary for the dynamic and hierarchical relationships between biological processes, cellular components and molecular functions, allowing a coherent annotation of gene products (Harris et al., 2004). In GO "biological process", includes "metabolic process", "cellular process", "response to stimulus", "single-organism process", "localization" and "biological regulation". "Cellular process" comprises of "cell", "organelle", "macromolecular complex", "membrane", "extracellular region" and "membrane-enclosed lumen". "Molecular function" represented by "binding", "catalytic activity", "structural molecule activity", "transporter activity", "electron carrier activity" and "antioxidant activity" (Fig. 2).

COG is a database where orthologous gene products are classified (Tatusov et al., 2000). In the present transcriptome analysis, all the unigenes obtained represented 24 functional categories in the COG database (Fig. 3). The "energy production and conversion" represented the largest group,

Table 3. Statistics of SSRs in the *H. verticillata* transcriptome.

Items	Numbers
Total no. of SSRs	1,041
No. of SSR containing sequences	650 (42.26%)
No. of sequences containing more than one SSR	170
No. of mono-nucleotide SSR	4 (0.38%)
No. of di-nucleotide SSR	19 (1.83%)
No. of tri-nucleotide SSR	149 (14.32%)
No. of tetra-nucleotide SSR	55 (5.28%)
No. of penta-nucleotide SSR	9 (0.86%)
No. of hexa-nucleotide SSR	805 (77.33%)

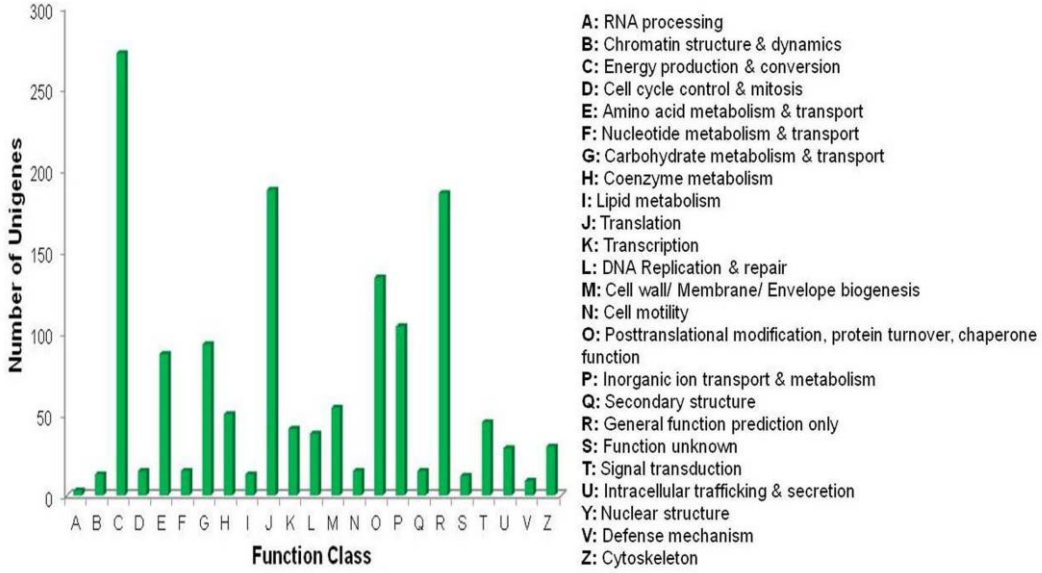

Fig 3. COG function classification of *H. verticillata* transcriptome. Graphical representation of clusters of orthologous (COG) classification. The number of unigenes is reported on the Y-axis.

followed by "translation", "general functional prediction only", "post translational modification, protein turnover, chaperone function", "Inorganic ion transport and metabolism", "carbohydrate metabolism and transport" and "amino acid metabolism and transport". The smallest groups were "RNA processing and modification" and "defence mechanism. This result indicated that most of the unigenes were involved in "energy production and conversion", "carbohydrate metabolism and conversion" which probably evolved because of high consumption rate of energy. But detailed studies are required to prove this hypothesis.

In addition to GO analysis, KEGG pathway mapping based on enzyme commission (EC) numbers for assignments was also carried out for the assembled sequences. This alternative approach to categorise gene functions emphasises biochemical pathways. KEGG annotated unigenes were classified into six different functional groups. Following these analyses, the largest group, with 42.46% of the unigenes, was classified into "metabolism", with most of these genes involved in "carbohydrate metabolism", "nucleotide metabolism" and "energy metabolism". The results suggest that carbohydrate, energy and nucleotide metabolism were active during the transition between these photosynthesis modes. Sequences classified into "genetic information processing (GIP)" accounted for 12.00% of the KEGG annotated sequences. Well-represented metabolic pathways included "folding, sorting and degradation", "replication and repair", "translation" and "transcription". "Cellular processes" represented 5.50% of the KEGG annotated sequences. Processes such as "transport and catabolism", "cell growth and death" and "cell communication" were well represented. Additionally, 5.83% of the sequences were classified into "environmental information processing (EIP)", for example, "signal transduction". "Organismal system" and "human disease" accounted for 11.22% and 22.99% of the unigenes, respectively (Fig. 4, Table 2). COG and KEGG pathway analysis are very useful techniques for predicting potential genes and the functions of these genes at a whole transcriptome level. The predicted molecular pathways, together with COG analysis, are useful for further investigations of the functions of genes.

Transcription factors (TFs) contain DNA-binding domains (DBDs) and regulate gene expression by binding to specific DNA sequences. TFs play important role in regulating growth, development and response to environmental factors by controlling gene variation. For example, Zhang et al., (2008) reported that ABP-9, a bZIP type TF has a role in improving the adaptability and stability of the photosynthetic apparatus to abiotic stress. They suggested that this TF potentially is related to expression of novel mechanism activated under water deficit and heat stress. In the present transcriptome profiling, a total of 18 putative transcription factors (TF) families were identified (Table S6). Among these, 11 unigenes represented the "Hap3/NF-YB" transcription factor, which were implicated in diverse developmental process. Further, the 2 unigenes represents AP2-EREBP and one unigene falls under C2H2 transcription factor family, which plays role in photosynthesis under stress condition (Saibo et al., 2009). HAP3 genes were shown to

Table 4. GO enrichment analysis of differentially expressed genes.

GO-ID	Term	Category	No. of differentially expressed genes	No. of unigenes in subgroup	P-value	FDR
Up-regulated in C_4 photosynthetic pathway						
Total			128	1293		
GO:0006950	Response to stress	P	5	131	9.56E-04	6.60E-02
GO:0006091	Generation of precursor Metabolites and energy	P	34	158	2.60E-03	1.22E-01
GO:0004518	Nuclease activity	F	7	12	3.26E-03	1.22E-01
GO:0003676	Nucleic acid binding	F	5	97	2.88E-02	3.46E-01
GO:0016788	Hydrolase activity, acting on ester bonds	F	7	15	8.26E-03	1.42E-01
Up-regulated in C_3 photosynthetic pathway						
Total			484	810		
GO:0009579	Thylakoid	C	35	114	2.84E-03	1.95E-01
GO:0006464	Cellular protein modification process	P	16	63	8.09E-03	1.95E-01
GO:0016301	Kinase activity	F	13	8	8.95E-03	1.95E-01
GO:0009605	Response to external stimulus	P	10	5	1.09E-02	1.95E-01
GO:0006412	translation	P	48	63	3.00E-02	3.62E-01
GO:0016788	Hydrolase activity, acting on ester bonds	F	3	19	4.20E-02	3.71E-01
GO:0005975	Carbohydrate metabolic process	P	44	58	4.30E-02	3.71E-01

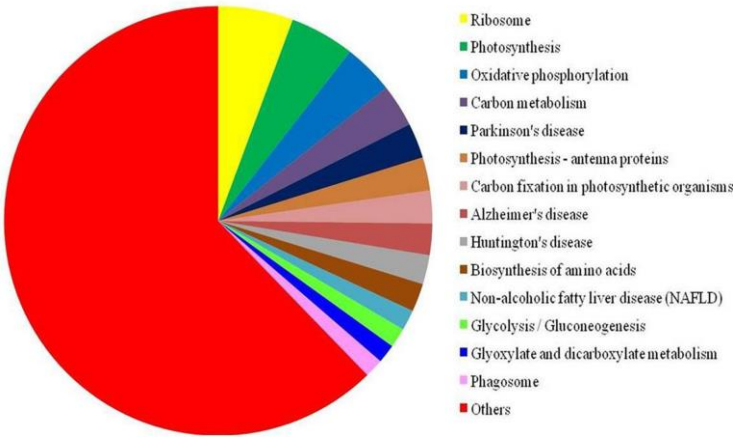

Fig 4. KEGG biochemical mappings of *H. verticillata*.

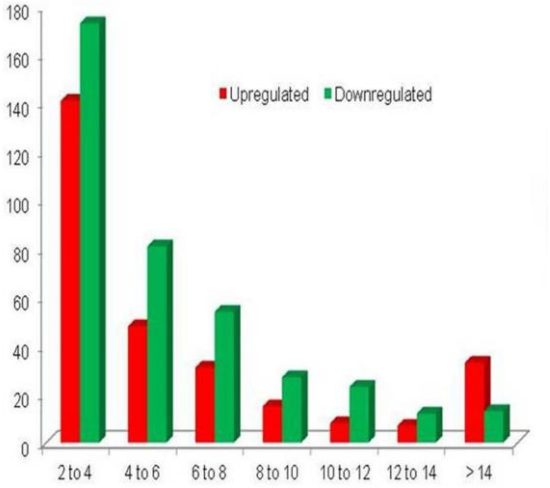

Fig 5. Overview of the differences in genes expression between the C_3 and C_4 photosynthetic pathway in *H. verticillata*. Fold change distribution of the differentially expressed genes.

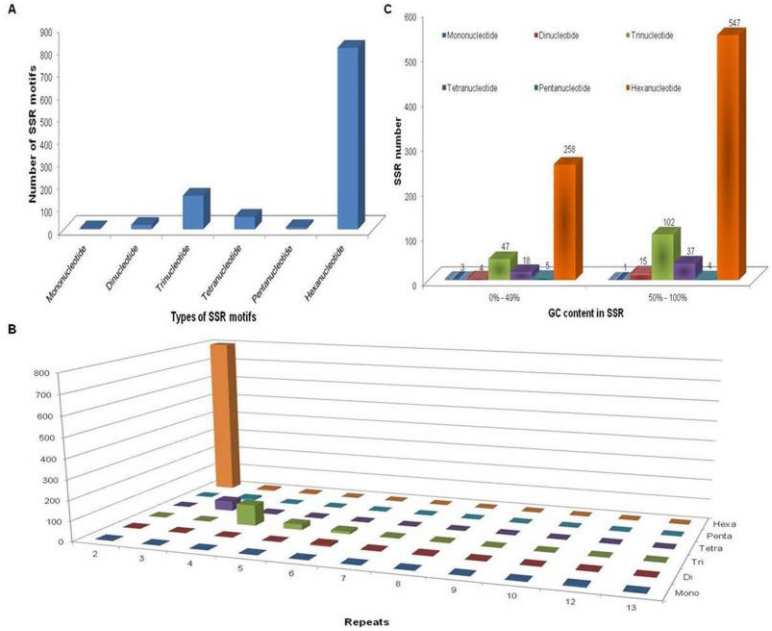

Fig 6. Distribution of SSRs. (a) Number of SSRs classified based on motif type; (b) GC content in microsatellites among different nucleotide types found in the transcriptome of *H. verticillata*; (c) Motif type and their frequency plotted as a function of the repeat number.

regulate the expression of nuclear-encoded chloroplast targeted genes and normal development of chloroplast in rice (Miyoshi et al., 2003). This Hap3/NF-YB" TF might be one of the candidates involved in the induction of photosynthetic genes. Park et al., (2001) and Guo et al., (2004) studied that AP2-EREBP TF lowers the loss of chlorophyll contents under high salt stress in tobacco. Hsieh et al., (2002) and Oh et al., (2007) observed that AP2-EREBP TF improves maximum quantum efficiency of PSII/ chlorophyll accumulation under different abiotic stresses. Kim et al., (2001) reported that C2H2 TF helps in faster recovery of chlorophyll content under cold stress.

RPKM based comparative expression profiling study (Mortazavi et al., 2008) was done to systematically characterize the mRNAs present in C_3 and C_4 state, and also to identify the transcripts of differentially expressed genes. Six hundred sixty six (666) unigenes were differentially expressed (Table S5). The photosynthesis related genes (unigenes 46, 6.91%) showed an over fourteen-fold change in expression level, which mainly included "photosystem I p750 chlorophyll a apoprotein al", "atp synthase subunit beta", "cell wall associated hydrolase" and "1-cys peroxiredoxin", while 314 unigenes (47.15%) showed a 2~4-fold change, including "malate dehydrogenase" and "phosphoenol-pyruvate carboxylase" (Fig. 5). These are the major C_4 photosynthetic genes, which show that our samples have started to shift to the C_4 photosynthetic pathway. Interestingly, of the 1,463 unigenes under investigation, 204 and 186 unigenes were uniquely expressed in HvT and HvC, respectively.

GO enrichment analysis of differentially expressed genes

To study the function of differentially expressed genes, GO terms were extracted using the Blast2GO tool and subjected to GO enrichment analysis. Annotation of the differentially expressed genes revealed that 356 unigenes belonged to 26 GO groups, while the remaining 310 unigenes could not be classified (Fig. 2). The number of up-regulated unigenes (90) with GO annotations in HvT was less than the number of down-regulated unigenes (266). GO annotation revealed that many differentially expressed genes were involved in "transferase activity", "kinase activity", "response to stimulus", "biological regulation", "antioxidant activity" and "electron carrier activity". A majority of the up-regulated unigenes in HvT were involved in "nuclease activity", "thylakoid", "generation of precursor metabolites and energy" and "hydrolase activity". In contrast, the GO groups "cellular protein modification process", "protein modification process" and "DNA binding" were under-represented in the up-regulated unigenes of the HvC (Table 4).

Marker development

Molecular markers play an important role in plant biology. These are very commonly used for the analysis of plant genomes and identification of the association between genomic variation and heritable traits. SSRs (microsatellites) markers are usually associated with phenotypic and functional variations, so they can be utilized in marker assisted selection programmes based studies in plants. These SSRs are multi allelic in nature, reproducible, highly abundant. They cover the genome extensively and are co-dominant in nature (Wei et al., 2011; Powell et al., 1996). SSRs derived from ESTs, genes or cDNA clones are referred to as EST- or genic SSRs. Genic SSRs have some intrinsic advantages over genomic SSRs, such as ease of generation, stability and cross-species transferability. Genic SSRs are used in variety of studies, such as genetic mapping and assessment of functional diversity (Varshney et al., 2005). Transcriptomic SSR markers exhibit high interspecific transferability, which can be harnessed as functional markers in various genetic studies. The transcriptomic data obtained by 454 sequencing provides an excellent source for mining and developing gene-associated markers. In total, 1041 SSRs were identified by SSR Locator in 650 (42.26%) unigenes of

H. verticillata. Hexa-nucleotide repeats (77.33%) were the most abundant SSRs in the transcriptome followed by tri-nucleotide (14.32%) and the remaining repeat motifs were mononucleotide (0.38%), dinucleotide (1.83%), tetran-ucleotide (5.28%) and pentanucleotide (0.86%) (Fig. 6a, 6b). The dominant repeat motif in all the SSRs identified in this study are AAGCGA/TCGCTT (33, 3.17%), followed by AAAATC/GATTTT (15, 1.44%), CTC/GAG (15, 1.44%), CCG/CGG (13, 1.25%) and GGA/TCC (12, 1.15%) (Table S7). Further validation in relation to phenotypic trait will be important.

Materials and Methods

Plant growth conditions

H. verticillata plants were collected from the Department of Botany, University of Delhi, New Delhi. The growth conditions were based on a modified version of Ginkel et al., (2001) and Rao et al., (2002) with changes in temperature. *H. verticillata* plants were kept in plastic tubs with irradiance of 300 µmol $m^{-2}s^{-1}$ at the National Phytotron Facility, IARI Campus, New Delhi at 34°C/22°C for 14/10 hr photoperiod's day/night to induce C_4 CCM (HvT) and at 20°C/15°C 14/10 hr to maintain C_3 state (HvC). Apical shoots with 40-45 leaves were harvested every day from the plants maintained under both conditions. These harvested shoots were flash-frozen in liquid N_2 and stored at -80°C till further use.

RNA isolation and expression analysis of PEPC 2

Total RNA was isolated from100mg plant material, HvT and HvC, using the RaFlex™ Total RNA Isolation Kit (GeNei™, Bangalore, India) as per the manufacturer's protocol. Contaminating genomic DNA was removed by DNase I treatment (Qiagen, Hilden, Germany). The quality of the total RNA was checked on a 1.2% formaldehyde agarose gel under denaturing conditions.

Complementary DNA (cDNA) was prepared using the Superscript® III First Strand cDNA Synthesis Kit (Invitrogen, Carlsbad, CA) as per the manufacturer's protocol. Semi-quantitative RT-PCR and Real-time PCR (qPCR) were performed to analyse the expression of *PEPC 2* gene using gene specific primer and the 18S rRNA primer as an internal control (Table S1).

Preparation of cDNA library and transcriptome sequencing

Messenger RNA (mRNA) was isolated from the total RNA of the samples (HvC and HvT; shifting to C_4 state) using the Oligotex® mRNA Midi Kit (Qiagen, Hilden, Germany) as per the manufacturer's protocol. The quality and quantity of RNA was determined using a Nanodrop 1000 (NanoDrop Technologies, USA) and a Bioanalyzer Chip RNA7500 series II (Agilent Technologies, USA). Equal quantities of high-quality mRNA from each sample (HvC and HvT) were used for the transcriptome study.

HvC and HvT cDNA were synthesized from respective mRNA using a cDNA Synthesis System (Roche, Basel, Switzerland) as per the manufacturer's protocol using hexamer primers (Roche, Basel, Switzerland). The cDNA samples were sheared by nebulisation to produce random fragments of approximately 300-800 bp in length. The nebulised cDNA was recovered with the QIAquick® PCR Purification Kit (Qiagen, Hilden, Germany). Rapid Library Molecular Identifier (RL-MID) were ligated to the fragmented purified cDNA samples according to the standard procedure. One MID contains a barcode sequence that is used to discriminate between samples from different libraries. Library quantitation was performed using the TBS 380 Fluorometer (Turner Biosystems, CA, USA). The relative fluorescence unit of each dilution was recorded and checked against the RL standard curve, and the sample concentration was calculated. An Agilent Bioanalyzer High Sensitivity DNA chip was used to check the quality of the libraries. The quality-passed samples were pooled and amplified by emulsion PCR. The sequencing of the libraries was performed on a 454-GS FLX Titanium sequencer as per the manufacturer's protocol (454 Life Sciences, Roche, Basel, Switzerland).

Sequencing data analysis and assembly

All the obtained raw reads were processed to remove sequences smaller than 60 bases (including adapters and primer) based on the assumption that these small reads might represent sequencing artefacts. At the same time Q20, Q30, GC content and sequence duplication level of the clean data were collected. High quality reads were assembled into unique putative transcripts (including contigs and singletons) using the Newbler ESTs assembler (V 2.6). The assembly was performed using the default parameters.

Unigene function annotation, GO classification and metabolic pathway analysis

Blast homology searches and sequence annotations were carried out using Blast2GO software v.2.7.1 (http://www.blast2 go.org) (Conesa et al., 2005; Conesa and Gotz, 2008; Gotz et al., 2008). The assembled sequences were compared against the NCBI nr protein database via BLAST X using an E-value cut-off of $1.0E^{-3}$. The searches were limited to the first 10 significant hits for each query to increase the computational speed. For gene ontology mapping (GO; http://www.geneontology.org), the program extracts the GO terms associated with homologies identified with NCBI's QBLAST and shows a list of GO annotations represented as hierarchical categories of increasing specificity. GO terms were modulated using the annotation augmentation tool, ANNEX (Myhre et al., 2006), followed by GO Slim (Harris et al., 2004). The data presented herein represent a GO analysis at level 2, demonstrating general functional categories. Enzyme classification (EC) codes were generated from direct mapping of GO terms to the enzyme code equivalents. GO term enrichment analysis was performed using Blast2GO software. This enrichment analysis was performed to evaluate the enrichment of various GO categories for the unigenes having expression values 2-fold or above in both the samples.

The unigenes sequences were also aligned to the COG database to predict and classify proteins. KEGG pathways were assigned to the unigene sequences using the single-directional best hit (SBH) method on the online KEGG Automatic Annotation Server (KAAS) (http://www.genome. jp/kegg/kaas) (Moriya et al., 2006). All the assembled unigenes were analysed against the PlantTFcat online tool (http://plantgrn.noble.org/PlantTFcat) for identification of the transcription factor in the *H. verticillata* transcriptome.

GC content analysis and Simple Sequence Repeat (SSR) discovery

Emboss GeeCee tool (http://emboss.bioinformatics.nl/cgi-bin/emboss/geecee) was used to measure the GC content of

the sequences and SSR markers. SSR markers were identified using SSR Locator V.1 software (da Maia et al., 2008). Mononucleotides repeats of more than 12, dinucleotide repeats of more than 6, trinucleotide repeats of more than 4, tetranucleotide repeats of more than 3, pentanucleotide repeats of more than 3 and hexanucleotide repeats of more than 2 were utilised as search criteria for SSRs.

Expression analysis

Clean reads from each of the libraries were used for mapping the sequences of each unigene and to calculate the unigene expression according to the "RPKM" method (Mortazavi et al., 2008). The significances of differences in gene expression between the C_3 and C_4 photosynthesis pathways were estimated using DEGseq, an R package [56]. The threshold of the P-value in multiple tests was identified using the False discovery rate (FDR), and the unigenes were considered as differentially expressed when the FDR was less than 0.05 and the log ratio was greater than 1 (two-fold change) between samples (Benjamini and Hochberg, 1995).

Conclusion

To the best of our knowledge, this is the first report to study the transcriptome of *H. verticillata*. A total of 533,595 high quality transcriptomic reads were obtained. A significant number of putative metabolic pathways and functions associated with the unique sequences were identified. Moreover, 1,041 SSRs were predicted among the unigenes, which can be used for marker development. These SSRs can also be utilized in studies involving genetic linkage and quantitative trait loci analysis. Overall, our transcriptome analysis of *H. verticillata* provides a valuable resource that will lead to a better understanding of the transition from the C_3 to C_4 photosynthetic pathway. This will aid in future studies related to engineering C_4 photosynthetic pathway in C_3 crop plants that will perform better in the event of rise in temperature and CO_2 level due to global warming. Further investigations will be necessary to fully elucidate the role of putative genes identified in this study, during transition of C_3 to C_4 photosynthetic pathway in *H. verticillata*.

Acknowledgment

We would like to thank the staff of NGS Sequencing Laboratory, ICAR-NRCPB, New Delhi, India for providing high throughput sequencing facility.

Author's contribution

EG and RSS carried out experiment, prepared cDNA library for Roche sequencing, sequencing run and expression analysis. AKM performed read generation, process of assembling, expression analysis. EG and AKS carried out homology searching, annotation, GO, COG, KEGG pathway analysis, SSR markers discovery. KK conceived the study, designed the experiment, associated with wet lab work and coordinated the study. EG, AKS and KK did result interpretation, analysis and integration of results, drafted and finalized the manuscript. KK checked the final manuscript. All authors have read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate a practical and powerful approach to multiple testing. J Roy Stat Soc Bl. 57: 289–300.

Bowes G, Salvucci ME (1989) Plasticity in the photosynthetic carbon metabolism of submersed aquatic macrophytes. Aquat Bot. 34: 233-266.

Brandner SJC, Salvucci ME (2000) Rubisco activase constrains the photosynthetic potential of leaves at high temperature and CO_2. P Natl Acad Sci USA. 97: 13430–13435.

Brautigam A, Gowik U (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biol. 12: 831-841.

Brautigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D et al (2011) An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. Plant Physiol. 155: 142-156.

Brautigam A, Schliesky S, Kulahoglu C, Osborne CP, Weber AP (2014) Towards an integrative model of C_4 photosynthetic subtypes insights from comparative transcriptome analysis of NAD-ME, NADP-ME and PEP-CK C_4 species. J Exp Bot. 65: 3579-3593.

Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics. 2008: 619832.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al (2005) Blast2GO a universal tool for annotation visualisation and analysis in functional genomics research. Bioinformatics. 21: 3674-3676.

da Maia LC, Palmieri DA, de Souza VQ, Kopp MM, de CarvalhoFIF, et al (2008) SSR Locator Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. Int J Plant Genomics. 4: 363-374.

Edwards GE, Franceschi VR, Voznesenskaya EV (2004) Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. Annu Rev Plant Biol. 55: 173–196.

Estavillo GM, Rao SK, Reiskind JB, Bowes G (2007) Characterisation of the NADP malic enzyme gene family in the facultative single-cell C_4 monocot *Hydrilla verticillata.* Photosynth Res. 94: 43-57.

Flicek P, Birney E (2009) Sense from sequence reads methods for alignment and assembly. Nat Methods. 6: 6-12.

Freitag H, Stichler W (2000) A remarkable new leaf type with unusual photosynthetic tissue in a central asiatic genus of chenopodiaceae. Plant Biol. 2: 154-160.

Furbank RT, von Caemmerer S, Sheehy J, Edwards G (2009) C_4 rice a challenge for plant phenomics. Funct Plant Biol. 36: 845–856.

Ginkel LC, Bowes G, Reiskind JB, Prins HBA (2001) A CO_2-flux mechanism operating via pH-polarity in *Hydrilla verticillata* leaves with C_3 and C_4 photosynthesis. Photosynth Res. 68: 81-88.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36: 3420-3435.

Gowik U, Westhoff P (2011) The path from C_3 to C_4 photosynthesis. Plant Physiol. 155: 56–63.

Guo ZJ, Chen XJ, Wu XL, Ling JQ, Xu P (2004) Overexpression of the AP2/EREBP transcription factor OPBP1 enhances disease resistance and salt tolerance in tobacco. Plant Mol Biol. 55: 607–618.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, et al (2004) The gene ontology (GO) database and informatics resource. Nucleic Acids Res. 32: 258-261.

Hatch MD (1988) C_4 photosynthesis a unique blend of modified biochemistry anatomy and ultrastructure. Biochim Biophys Acta. 895: 81-106.

Hausler RE, Hirsch HJ, Kreuzaler F, Peterhansel C (2002) Overexpression of C_4-cycle enzymes in transgenic C_3 plants a biotechnological approach to improve C_3-photosynthesis. J Exp Bot. 53: 591–607.

Hsieh TH, Lee JT, Yang PT, et al (2002) Heterology expression of the Arabidopsis *C-repeat/dehydration response element binding factor 1* gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. Plant Physiol. 129: 1086–1094.

Kim JC, Lee SH, Cheong YH, et al (2001) A novel cold-inducible zinc finger protein from soybean SCOF-1 enhances cold tolerance in transgenic plants. Plant J. 25: 247–259.

Kobza J, Edwards GE (1987) Influences of leaf temperature on photosynthetic carbon in metabolism in wheat. Plant Physiol. 83: 69-74.

Kuhl JC, Cheung F, Yuan Q, Martin W, Zewdie Y, et al (2004) A unique set of 11008 onion expressed sequence tags reveals expressed sequence and genomic differences between the monocot orders asparagales and poales. Plant Cell. 16: 114–125.

Li W, Godzik A (2006) Cd-hit a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22: 1658–1659.

Magnin NC, Cooley BA, Reiskind JB, et al (1997) Regulation and localization of key enzymes during the induction of Kranz-less C_4 photosynthesis in *Hydrilla verticillata.* Plant Physiol. 115: 1681-1689.

Matsuoka M, Furbank RT, Fukayama H, Miyao M (2001) Molecular engineering of C_4 photosynthesis. Annu Rev Plant Phys. 52: 297–314.

Metzker ML (2010) Applications of next-generation sequencing technologies the next generation. Nat Rev Genet. 11: 31–46.

Miyao M (2003) Molecular evolution and genetic engineering of C_4 photosynthetic enzymes. J Exp Bot. 54: 179–189.

Miyao M, Masumoto C, Miyazawa S, Fukayama H (2011) Lessons from engineering a single-cell C_4 photosynthetic pathway into rice. J Exp Bot. 62: 3021-3029.

Miyoshi K, Ito Y, Serizawa A, Kurata N (2003) OsHAP3 genes regulate chloroplast biogenesis in rice. Plant J. 36: 532-540.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2006) KAAS an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35: 182–185.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 5: 621–628.

Myhre S, Tveit H, Mollestad T, Laegreid A (2006) Additional gene ontology structure for improved biological reasoning. Bioinformatics. 22: 2020–2027.

Oaks A (1994) Efficiency of nitrogen utilization in C_3 and C_4 cereals. Plant Physiol. 106: 407–414.

Oh SJ, Kwon CW, Choi DW, Song SI, Kim JK (2007) Expression of barley HvCBF4 enhances tolerance to abiotic stress in transgenic rice. Plant Biotechnol J. 5: 646–656.

Park JM, Park CJ, Lee SB, Ham BK, Shin R, Paek KH (2001) Overexpression of the tobacco *Tsi1* gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. Plant Cell. 13: 1035–1046.

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1: 215-222.

Rao SK, Fukayama H, Reiskind J, Miyao M, Bowes G (2006a). Identification of C_4 responsive genes in the facultative C_4 plant *Hydrilla verticillata.* Photosynth Res. 88: 173–183.

Rao SK, Magnin NC, Reiskind JB, Bowes G (2002) Photosynthetic and other phosphoenolpyruvate carboxylase isoforms in the single cell facultative C_4 system of *Hydrilla verticillata*. Plant Physiol. 130: 876–886.

Rao SK, Reiskind JB, Bowes G (2006b) Light regulation of the photosynthetic phosphoenolpyruvate carboxylase (PEPC) in *Hydrilla verticillata.* Plant Cell Physiol. 47: 1206–1216.

Rao SK, Reiskind JB, Bowes G (2008) Kinetic analyses of recombinant isoforms of phosphoenolpyruvate carboxylase from *Hydrilla verticillata* leaves and the impact of substituting a C_4-signature serine. Plant Sci. 174: 475–483.

Reiskind JB, Berg RH, Salvucci ME, Bowes G (1989) Immunogold localization of primary carboxylases in leaves of aquatic and a C_3 C_4 intermediate species. Plant Sci. 61: 43-52.

Reiskind JB, Maberly SC (2014) A tribute to george bowes linking terrestrial and aquatic botany. Aquat Bot. 118: 1-3.

Reiskind JB, Madsen TV, Van Ginkel LC, Bowes G (1997) Evidence that inducible C_4-type photosynthesis is a chloroplastic CO_2-concentrating mechanism in *Hydrilla* a submersed monocot. Plant Cell Environ. 20: 211-220.

Reynolds M, Bonnett D, Chapman SC, Furbank RT, Mane Y, Mather DE, Parry MAJ (2010) Raising yield potential of wheat. I. overview of a consortium approach and breeding strategies. J Exp Bot. 62: 439-52.

Sage RF (1990) A model describing the regulation of ribulose-1-5-bisphosphate carboxylase electron transport and triose phosphate use in response to light intensity and CO_2 in C_3 plants. Plant Physiol. 94: 1728-1734.

Saibo NJM, Lourenco T, Oliveira MM (2009) Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. Ann Bot-London. 103: 609-623.

Sheehy JE, Mitchell PL, Hardy B (2000) Redesigning Rice Photosynthesis to Increase Yield. Proceedings of the Workshop on the Quest to Reduce Hunger: Redesigning Rice Photosynthesis, 30 November - 3 December 1999, Los Baños, Philippines. Makati City (Philippines): International Rice Research Institute and Amsterdam (The Netherlands): Elsevier Science B.V. pp-293.

Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28: 33-36.

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants features and applications. Trends Biotechnol. 23: 48-55.

Voznesenskaya EV, Franceschi VR, Kiirats O, Freitag H, Edwards GE (2001) Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. Nature. 414: 543–546.

Vu JCV, Allen LH, Boote KJ, Bowes Gb (1997) Effects of elevated CO_2 and temperature on photosynthesis and rubisco in rice and soybean. Plant Cell Environ. 20: 68-76.

Wang L, Feng Z, Wang X, Wang X, Zhang X (2010) DEGseq an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 26: 136–138.

Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X (2011) Characterization of the sesame (*Sesamum indicum* L) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics. 12: 451-464.

Zhang X, Wollenweber B, Jiang D, Liu F, Zhao J (2008) Water deficits and heat shock effects on photosynthesis of a transgenic *Arabidopsis thaliana* constitutively expressing ABP9 a bZIP transcription factor. J Exp Bot. 59: 839–848.