

# Expression of Photosynthesis Pathway Gene From Rice and Maize for Understanding Role in Plant Stress and Development Using Bioinformatics Approaches

saurabh pandey (✉ [saurabhpandey28@gmail.com](mailto:saurabhpandey28@gmail.com))

International Centre for Genetic Engineering and Biotechnology <https://orcid.org/0000-0001-5504-0825>

Vaibhav Raina

International Centre For Genetic Engineering and Biotechnology New Delhi

---

## Research article

**Keywords:** Carbonic Anhydrase (CA), Phosphoenolpyruvate Carboxylase (PEPC), NADP dependent Malic Enzyme (NADP-ME), Malate Dehydrogenase (MDH), Pyruvate Orthophosphate Dikinase (PPDK), Genomic annotation, Synteny

**Posted Date:** July 14th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-38571/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

## Background

Rice and maize go to family Poaceae contains many crops of agronomic trait and also represent two carbon metabolism systems, C3 and C4. Analysis of the maize sequence provides new insights into the employment of C3 genes to the C4 mechanism which allowed us to identify more orthologs in other crops. This investigation reports comparative account of genome wide in silico identification of C4 pathway related genes from *Zea mays* (Zm) and *Oryza sativa* (Os) from the available whole genome sequence information. The annotation of gene sequences, signature motif analysis, protein phosphorylation analysis, study of upstream cis-acting elements, phylogenetic tree construction, chromosomal locations, syntenic mapping and microarray expression analysis of C4 pathway related gene family from both the genomes have been attempted.

## Results

A total of 30 and 37 C4-pathway genes have been predicted from rice and maize genome respectively. Multiple-sequence-alignment and signature motif analysis of these proteins of rice and maize revealed high conserveness. Phosphorylation analysis revealed that maize have high number than rice. The phylogenetic analysis of C4 related genes across both plant species clearly resulted in four sub-groups in both plants. In Rice, the 30 genes of C4 pathway related genes family are distributed on eleven out of twelve chromosomes, while in maize, they are randomly distributed on all the chromosomes. Most of the genes of Zm's chromosome 1 show syntenic relationship with chromosome 1. The cis-regulatory-elements of Zm and Os genes suggested its diverse functions associated with plant growth development, stress and hormone responsiveness as well as endosperm and meristem specific gene expression. This investigation of Zm and Os can now offer new insights into the role of different C4 pathway related genes and examine the comparative syntenic mapping between two monocot models and allows for better understanding about how genes evolve within monocots. Therefore, in silico investigation of C4-photosynthetic-pathway gene family needs to be supported by wet lab experimentation of the novel genes for elucidating their function in many biological courses.

## Conclusion

Results revealed that photosynthetic pathway related gene play a potent role in stress response and plant growth and development.

# Background

The prevention of food related consequences arising due to ever increasing population, rapid urbanization, extreme global climate changes and acute water shortage requires an urgent effort to increase the crop yield by a significant percentage to meet daily ends of both developing and developed nations. The solution to such a big problem can be developed by multiple approaches. One such approach which seems most promising is introduction of C4 photosynthetic pathways into C3 crops such as rice, which substantially increases the efficiency of plants to utilise nitrogen, water and radiation [1, 2], thus producing a higher yield than current C3 crops available [3]. It could be done by detailed analysis of expression pattern of C4 associated genes which influence biochemical process or produce any prominent morphological and anatomical changes in plants and their protein which could be incorporated into the C3 plants by means of genetic engineering [4, 5]. Result from this analysis could be exploited to increase the efficiency of crop yield. Some of the striking similarities that could prove beneficial in future experiments could be relationship between orthologous genes of C4 and C3 plants, changing gene expression patterns of genes in C4 in response to a stimuli which are also present in C3 plants [6]. Incorporating and regulating the expression of C4 genes in C3 plants could also be handled by various approach. C4 pathway is one of three biochemical process used to fix atmospheric CO<sub>2</sub> along with CAM & C3 pathway and is characterized by formation of first 4 carbon product in plants. Also referred as Hatch-Slack Pathway [7]. Evolution of C4 pathway has been independently observed 66 times from their C3 ancestors [8], at least in 19 families during angiosperm evolution [9]. C4 pathway has evolved in plants to prevent loss of about 40% of its efficiency which occurs due to photorespiration, an energy demanding process and further more leads to net loss of CO<sub>2</sub> [10].

Photorespiration occurs in response to remove phosphoglycolate- a by product of reaction involving addition of oxygen by Rubisco to ribulose-1,5 biphosphate resulting in generation of one molecule each of 3-phosphoglycerate and 2-phosphoglycolate [10, 11]. Also Phosphoglycolate has no known metabolic purpose and toxic in higher concentrations [11, 12]. It has evolved as an adaptation to high temperature and light intensities. C4 plants outperform C3 plants as they use less water per CO<sub>2</sub> fixed than C3 plants and hence are more suitable for hot/dry climate [13] Major differences between C4 and C3 pathway involved are:- (1) the differentiation of the two cell and chloroplast types, (2) the presence of an additional set of genes, and (3) a mechanism regulating the cell-specific expression of these additional genes [14, 15]. The enzymes involved in C3 pathway are located in the chloroplast of the bundle sheath cells while those involved in C4 pathway in mesophyll and /or bundle sheath cells for intercellular metabolic cooperation. In NADP-ME pathway it can be seen that PEPCase, NADP-dependent malate dehydrogenase (NADPMDH), and pyruvate phosphate dikinase (PPDK) function in M cells, whereas NADP-ME and RuBPCase activities are localized in BS cells [16, 17, 18]. The C4 pathway consist of 3 key steps:-A] initial fixation of CO<sub>2</sub> into cytosol of mesophyll cells by PEPC [phosphoenolpyruvate carboxylase] to form a C4 acid, Oxaloacetatae [OAA] B] decarboxylation of a C4 acid in bundle sheath cells to release CO<sub>2</sub> and carbon regeneration of primary CO<sub>2</sub> acceptor phosphoenolpyruvate [PEP] [19]. The decarboxylation is catalyzed by 3 enzymes namely NADP-malic enzyme [NADP-ME], NAD malic enzyme [NAD-ME] and phosphoenolpyruvate carboxykinase [PEP-CK]. Recent comparative studies have revealed C3 plants to contain 2 set of genes housekeeping and C4 like genes whose expression level is low in C3 plants. To have a significant effect on metabolism gene of C4 plants need to be overexpressed in C3 plants [20]. Studies have shown that the C4 photosynthetic pathway related genes viz. CA, PEPC, NADP-ME, MDH, PPDK and the regulatory protein PPDK-RP in the flag leaf during grain filling stage in different rice genotypes express in C3 rice leaves with a wide genotypic variation [21]. Further studies on the role and contribution of C4 photosynthetic pathway related genes of rice in carbon metabolism will help to enhance the photosynthetic efficiency of rice. Some of the past research indicating effect of overexpressed genes of C4 plants in C3 plants. Our aim was to check role of C4 enzymes individually and to check their role with a combination of other enzymes [22–23]. A series of experiments carried out by researchers in past experiments showed the physiological and photosynthetic effects of C4 enzymes on overexpression in C3 plants. Although PEPC is involved in initial fixation of CO<sub>2</sub> in C4

pathway, on overproduction in C3 plants, it stimulated respiration in light and synthesis of amino and organic acids instead of increasing photosynthetic efficiency of rice plants the C4 pathway [24–25]. It has been observed that in the leaves of C3 plants, PEPC has an anaplerotic role replenishing the tricarboxylic acid cycle with intermediates, which are withdrawn for nitrate assimilation and the subsequent amino acid synthesis [26]. No prominent changes were observed in physiological and photosynthetic characteristics of transgenic rice in which C4 specific PPDK was introduced [27]. Although elevation above a threshold level generated a minor response. Overproduction of C4 specific NADP-ME led to increased photo-inhibition of photosynthesis, leaf chlorophyll bleaching and serious stunting because of increase in the NADPH/NADP ratio in the chloroplast [28].

In the present study, comparative genome wide in silico identification of C3/C4 gene to protein such as CA, PEPC, PECK, NADP-ME, NADP-MDH and PPDK involved in photosynthesis pathway of rice and maize. The member further were analyzed in detail in term as chromosomal location(s), gene structure, phylogenetic tree for evolutionary relationship construction and also analyzing the cis-regulatory elements associated with these genes in the promoter region. Further, comparative phylogeny and syntenic mapping and protein phosphorylation examine with putative protein signature sequences with their function. Further, we have attempted gene expression analysis development anatomical and stress condition during various stages of panicle and seed development implies their involvement in diverse developmental processes. rice and maize C3/C4 gene family have also been attempted.

## Materials And Methods

Search of databases for the identification of photosynthetic genes/ proteins family members

In order to perform comparative analysis, the sequences of C3/C4 photosynthetic gene/ protein sequences such as CA, PEPC, PEPC, NADP-ME, NADP-MDH and PPDK from rice (*O. sativa*) and Maize (*Z. mays*) were downloaded from different sources, TIGR (The Institute for Genomic Research; <http://www.tigr.org/>) and rice genome annotation project (<http://rice.plantbiology.msu.edu/>) for rice and NCBI The National Center for Biotechnology Information as primary database (<http://www.ncbi.nlm.nih.gov/>) for maize, respectively. And check all maize gene/ protein correspond to Maize genome database MaizeGDB <http://www.maizegdb.org/> For the identification annotation of photosynthetic C3/C4 genes in rice and maize, the homology search of the C3/C4 genes photosynthetic genes proteins was performed by BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) using blastp and tblastn algorithm [29–30]. Detailed analysis C3/C4 genes photosynthetic genes proteins in whole genomes of rice and sorghum, including their identification, classification, sequence analysis, domain signature analysis, chromosomal locations phylogenetic relationships and expression analysis were carried out as described below and gene list given in Table 1a and b.

Table 1

(A) List of C3/C4 photosynthetic genes in Rice along with their chromosomal locations, alternative spliced forms, CDS and polypeptide length, localization and homologs.

	NAME	Locus ID	GENE DISCRIPTION	Chr	CDS Coordinates (5'-3'):	Nucleotide length	PPL	PMW	Ppl	pfamhit
1	PPDK	LOC_Os05g33570	Orthophosphate dikinase precursor (EC 2.7.9.1)	Chr5	19737857 – 19718506	2844	948	102788	6.3145	PF01326.12
2	PPDK	LOC_Os03g31750	Orthophosphate dikinase precursor (EC 2.7.9.1)	Chr3	18160127 – 18153143	2679	893	97212.7	5.3604	PF01326.12
3	NADP malic	LOC_Os01g09320	NADP-dependent malic enzyme 3	Chr1	4744591 – 4738905	1920	640	69865.8	7.1548	PF00390.12
4	NADP malic	LOC_Os01g52500	NADP-dependent malic enzyme 4	Chr1	30166024 – 30171659	1932	644	70482.4	5.5251	PF00390.12
5	NADP malic	LOC_Os01g54030	NADP-dependent malic enzyme 5	Chr1	31076571 – 31081377	1758	586	64269.8	6.9361	PF00390.12
6	NADP malic	LOC_Os02g44550	NADP-dependent malic enzyme 6	Chr2	26993990 – 26988678	561	187	20398	9.4713	PF00390.12
7	NADP malic	LOC_Os05g09440	NADP-dependent malic enzyme 7	Chr5	5297419 – 5293555	1713	571	62934.9	5.5253	PF00390.12
8	PEPCK	LOC_Os03g15050	Phosphoenolpyruvate carboxykinase [ATP]	Chr3	8224127 – 8218791	2001	667	73771.8	7.5064	PF01293.13
9	PEPCK	LOC_Os04g50208	Phosphoenolpyruvate carboxykinase [ATP]	Chr4	29972696 – 29969234	1104	368	39279	9.2502	PF01293.13
10	PEPCK	LOC_Os10g13700	Phosphoenolpyruvate carboxykinase [ATP]	Chr10	7444108 – 7438335	1605	535	59518.4	6.7889	PF01293.13
11	PPC	LOC_Os01g02050	PEPcase; Phosphoenolpyruvate carboxylase	Chr1	572138 – 562056	3108	1036	115840	6.9464	PF00311.10
12	PPC	LOC_Os01g11054	PEPcase; Phosphoenolpyruvate carboxylase	Chr1	5899555 – 5909595	3045	1015	114197	5.9411	PF00311.10
13	PPC	LOC_Os01g55350	PEPcase; Phosphoenolpyruvate carboxylase	Chr1	31865031 – 31859331	2775	925	105314	6.594	PF00311.10
14	PPC	LOC_Os02g14770	PEPcase; Phosphoenolpyruvate carboxylase	Chr2	8184220 – 8172419	2907	969	109933	6.1467	PF00311.10
15	PPC	LOC_Os08g27840	PEPcase; Phosphoenolpyruvate carboxylase	Chr8	16970571 – 16964694	2895	965	110056	5.5479	PF00311.10
16	PPC	LOC_Os09g14670	PEPcase; Phosphoenolpyruvate carboxylase	Chr9	8697573 – 8692191	2928	976	110474	6.1404	PF00311.10
17	CA	LOC_Os01g45274	Carbonic anhydrase	Chr1	25696613 – 25705090	819	273	29117.4	8.2758	PF00484.12
18	CA	LOC_Os09g28910	Carbonic anhydrase	Chr9	17572392 – 17567632	1002	334	36800.6	7.9867	PF00484.12
19	CA	LOC_Os11g05510	Carbonic anhydrase	Chr11	2488205 – 2488622	342	114	11814.3	7.4979	PF00194.14
20	CA	LOC_Os09g28130	Carbonic anhydrase	Chr9	17075709 – 17075190	291	97	10802.2	5.1666	PF00194.14
21	NAD MDH	LOC_Os08g33720	NAD malate dehydrogenase, chloroplastic	Chr8	21057561 – 21054659	1194	398	41538.7	7.5418	PF00056.16
22	NAD MDH	LOC_Os01g61380	malate dehydrogenase, chloroplastic	Chr1	35499017 – 35501765	1191	397	41787	7.8961	PF00056.16
23	NAD MDH	LOC_Os01g46070	malate dehydrogenase, mitochondrial	Chr1	26190752 – 26194517	1023	341	35460.9	8.8513	PF00056.16
24	NAD MDH	LOC_Os12g43630	malate dehydrogenase, glyoxysomal	Chr12	27099351 – 27094647	1071	357	37385.4	7.9869	PF00056.16

	NAME	Locus ID	GENE DISCRIPTION	Chr	CDS Coordinates (5'-3'):	Nucleotide length	PPL	PMW	Ppl	pfamhit
25	NAD MDH	LOC_Os10g33800	malate dehydrogenase, cytoplasmic	Chr10	17913818–17917850	999	333	35568.9	5.9726	PF00056.16
26	NAD MDH	LOC_Os08g44810	malate dehydrogenase [NADP], chloroplastic	Chr8	28141042–28146270	1302	434	47008.9	7.3392	PF00056.16
27	NAD MDH	LOC_Os05g49880	malate dehydrogenase, mitochondrial	Chr5	28621585 – 28617595	1023	341	35435.9	8.2991	PF00056.16
28	NAD MDH	LOC_Os07g43700	malate dehydrogenase, chloroplastic	Chr7	26155933 – 26153825	1215	405	42221.7	9.0291	PF00056.16
29	NAD MDH	LOC_Os04g46560	malate dehydrogenase, cytoplasmic	Chr4	27605166–27608347	1059	353	38297.1	7.2212	PF00056.16
30	NAD MDH	LOC_Os03g56280	malate dehydrogenase, glyoxysomal	Chr3	32089685 – 32086001	1065	355	37023	8.0614	PF00056.16

Table 1

(B) List of C3/C4 photosynthetic genes in Maize along with their chromosomal locations, alternative spliced forms, CDS and polypeptide length, localization, and homologs.

Gene symbol	Ch	Exon	Location	Protein ID	RNA ID	DOMAIN	Location		
umc2175	8	8	NC_024446.1 (146085247..146092154)	NP_001168699.1	NM_001175228.1	cd00884	CH	beta CA	CARBONIC ANHYDRAS
csu869(cah)	3	7	NC_024461.1 (215513277..215518550,	NP_001151431.1	NM_001157959.1	cd00884	CH	beta CA	CARBONIC ANHYDRAS
LOC100280638	4	5	NC_024462.1 (56994835..56996762,	NP_001147028.1	NM_001153556.1	cd03124	NF	alpha CA	CARBONIC ANHYDRAS
LOC100274185	2	7	NC_024460.1 (57684848..57687212,	NP_001142031.1	NM_001148559.1	cd03124	CH	alpha CA	CARBONIC ANHYDRAS
LOC100283312	7	10	NC_024465.1 (127621625..127626283,	NP_001149686.1	NM_001156214.1	cd00884	CH	beta CA	CARBONIC ANHYDRAS
LOC100279972	9	4	NC_024467.1 (103408368..103410241)	NP_001146392.1	NM_001152920.1	cl00012	CH	alpha CA	CARBONIC ANHYDRAS
LOC100274597	3	7	NC_024461.1 (215506036..215509026,	NP_001152905.1	NM_001159433.1	cd00884	CH	beta CA	CARBONIC ANHYDRAS
LOC100272439	2	10	NC_024460.1 (187773678..187778034,	NP_001140385.1	NM_001146913.1	cl00391	CH	beta CA	CARBONIC ANHYDRAS
LOC100283752			100283752	NP_001150123.1	NM_001156651.1		NF	beta CA	CARBONIC ANHYDRAS
LOC100282652	1	5	NC_024459.1 (199145284..199147439)	NP_001149032.1	NM_001155560.1	cd03124	CH	alpha CA	CARBONIC ANHYDRAS
LOC100275493	3	14	NC_024461.1 (215547921..215555236,	NP_001105359.1	NM_001111889.1	cl00391	NF	beta CA	CARBONIC ANHYDRAS
pep1	9	10	NC_024467.1 (62306266..62311652)	NP_001154820.1	NM_001161348.1	pfam00311		PEPcase	Phosphoenolpyruvate carboxylase
pep4	7	10	NC_024465.1 (86459218..86464631,	NP_001105438.1	NM_001111968.1	pfam00311		PEPcase	Phosphoenolpyruvate carboxylase
pep7	5	10	NC_024463.1 (144847532..144854986,	NP_001105503.1	NM_001112033.1	pfam00311		PEPcase	Phosphoenolpyruvate carboxylase
IDP1621	4	10	NC_024462.1 (227810840..227818674)	NP_001130365.1	NM_001136893.1	cl21521		pepcase	Phosphoenolpyruvate carboxylase
LOC541622	1	11	NC_024459.1 (35744408..35748808)	NP_001296837.1	NM_001309908.1	cd00484		PEPCK	Phosphoenolpyruvate carboxylase
PCK2	9	11	NC_024467.1 (142082093..142086899)	NP_001146178.1	NM_001152706.1			PEPCK	Phosphoenolpyruvate carboxylase
pdk1	6	19	NC_024464.1 (146179394..146189973)	NP_001105738.2	NM_001112268.2	PRK09279	CH	PPDK	Pyruvate, phosphate dikinase
LOC100274067	10	13	NC_024468.1 (20391750..20418385)	NP_001141918.1	NM_001148446.1	pfam01326	CH	PPDK	Pyruvate, phosphate dikinase
LOC100381411	6	33	NC_024464.1 (110869370..110881829)	NP_001167723.1	NM_001174252.1		CH	PPDK	Pyruvate, phosphate dikinase
me3	3	20	NC_024461.1 (7276413..7281610)	NP_001105313.1	NM_001111843.1	PLN03129	CH	ME-NAD	NAD-dependent enzyme
me2	6	23	NC_024464.1 (139420838..139469949)	NP_001105383.2	NM_001111913.2	PLN03129	CH	ME-NAD	NAD-dependent enzyme
me4	8	19	NC_024466.1 (174612837..174617017)	NP_001105292.1	NM_001111822.1	PLN03129	CY/CH	ME-NAD	NAD-dependent enzyme
LOC100284598	6	9	NC_024464.1 (130314932..130318270)	NP_001150965.1	NM_001157493.1	PLN03129	CY/CH	ME-NAD	NAD-dependent enzyme
LOC100501486	5	19	NC_024463.1 (23916406..23923194)	NP_001183119.1	NM_001196190.1	PTZ00317	MT	ME-NAD	NAD-dependent enzyme
LOC100286036	3	20	NC_024461.1 (201756875..201761835)	NP_001152396.1	NM_001158924.1	PLN03129	CH	ME-NAD	NAD-dependent enzyme
mdh6	1	14	NC_024459.1 (203210173..203213877,	NP_001105420.1	NM_001111950.1	PLN00112	CH	MDH	NADP-Malic Dehydrogenase

Gene symbol	Ch	Exon	Location	Protein ID	RNA ID	DOMAIN	Location		
LOC100193663	4	4	NC_024462.1 (83532085..83534409)	NP_001132228.1	NM_001138756.1	PLN00106	GL	MDH	NADP-Malate Dehydrogen
mdh5	1	7	NC_024459.1 (231398321..231403647,	NP_001105603.1	NM_001112133.2	PLN00135	CY	MDH	NADP-Malate Dehydrogen
LOC100856934				NP_001241749.1	NM_001254820.1	PLN00106	CH	MDH	NADP-Malate Dehydrogen
LOC100282134	1	8	NC_024459.1 (282605052..282608688,	NP_001148518.1	NM_001155046.1	PLN00106	GL	MDH	NADP-Malate Dehydrogen
LOC100274264	6	7	NC_024464.1 (165831696..165836197,	NP_001142100.1	NM_001148628.2	PLN00106	MT	MDH	NADP-Malate Dehydrogen
LOC100280767		2		NP_001147160.1	NM_001153688.1	PLN00135	CY	MDH	NADP-Malate Dehydrogen
LOC100193743	1	2	NC_024459.1 (213144354..213146991,	NP_001132302.1	NM_001138830.1	PLN00106	CH	MDH	NADP-Malate Dehydrogen
LOC100193491	3	10	NC_024461.1 (91369952..91374547,	NP_001132077.2	NM_001138605.2	PLN00106	GL	MDH	NADP-Malate Dehydrogen
LOC100273428		9	NW_007617880.1 (516..4849)	NP_001141337.1	NM_001147865.1	PLN00106	MT	MDH	NADP-Malate Dehydrogen

Table 2

(A) Cis-regulatory elements in the upstream region of C3 photosynthetic genes of rice. Elements responsive to plant growth and development, stress response and hormones were analyzed.

GROUP	GENE ID	Plant growth and development	Biotic and abiotic stress response	Hormone response
PPDK	Os05g0405000		ARE(1),MBS(1), TC-rich repeats(1),	TCA element(2),ABRE(4),
	Os03g0432100	CAT box(1),circadian(1),	MBS(2),	CGTCA-motif(1),TGACG-motif (1),ABRE(3)
NAD-ME	Os01g0188400	Skn1-motif(1),as-2-box(1),	ARE(2),HSE(2),	TCA element(1),ABRE(1)
	Os01g0723400	CCGTCC-box(1),	MBS(2), HSE(1), TC-rich repeats(1),	CGTCA-motif(1), TGACG-motif (1), ABRE(1),TATC box(1)
	Os01g0743500	CCGTCC-box(1),O2 site(1), Skn1-motif(1),	MBS(1),	CGTCA-motif(1), TGACG-motif (1), ABRE(6)
	Os02g0665000	CCGTCC-box(1), Skn1-motif(1),	ARE(1), MBS(2),LTR(1),	GARE-motif(1), TGA-element(1),
	Os05g0186300	CCGTCC-box(1),	ARE(4), HSE(1), LTR(1),	CGTCA-motif(2), TGACG-motif (2), ABRE(1), GARE-motif(1)
PEPCK	Os03g0255500	CCGTCC-box(1), circadian(1),	ARE(1), MBS(2),	p-box(1),ERE(1)
	Os04g0592500	CAT box(1), CCGTCC-box(1), Skn1-motif(1), circadian(1),GCN4 motif(1),	ARE(1),box-w1(1), TC-rich repeats(1),	TCA element(1), TGACG-motif (1), ABRE(3), TATC box(1), AuxRR-core(1)
	Os10g0204400	CAT box(1), CCGTCC-box(1), Skn1-motif(3),	ARE(1), MBS(3), HSE(2), TC-rich repeats(1),	TGACG-motif (1), ABRE(5)
PEPC	Os01g0110700	CAT box(1), O2 site(1), circadian(1)	ARE(1),HSE(1),	TCA element(1), TGACG-motif (1), ABRE(1)
	Os01g0208700	CAT box(1), Skn1-motif(2),	ARE(1), MBS(1), TC-rich repeats(2), LTR(1),	CGTCA-motif(2), TCA element(1), TGACG-motif (2), TATC box(1),
	Os01g0758300	Skn1-motif(3),circadian(2),	MBS(1),	CGTCA-motif(1), TGACG-motif (1), GARE-motif(1)
	Os02g0244700	CCGTCC-box(1), O2 site(2), circadian(2),	ARE(1), MBS(2), box-w1(1), TC-rich repeats(1),	CGTCA-motif(2), TGACG-motif (2)
	Os08g0366000	circadian(1)	ARE(1), HSE(1), TC-rich repeats(1),WUN-motif(1),	GARE-motif(1)
	Os09g0315700	CCGTCC-box(2), Skn1-motif(1), circadian(1),	ARE(3), HSE(3),	TCA element(1)
BETA-CA	Os01g0639900	CAT box(1), MSA-like (1), Skn1-motif(1), circadian(1), GCN4 motif(1),	ARE(1),	CGTCA-motif(2), TGACG-motif (2),ABRE(5)
	Os09g0464000	CCGTCC-box(1), MSA-like (1), Skn1-motif(1), as-2-box(1), GCN4 motif(1)	ARE(1),MBS(1)	ABRE(1)
ALPHA-CA	Os11g0153200	CCGTCC-box(1), O2 site(1), circadian(1), as-2-box(2)	ARE(2),HSE(6),	ERE(1)
	Os09g0454400			
NAD-MDH	Os08g0434300	CAT box(1),	ARE(3), MBS(2), TC-rich repeats(2), WUN-motif(1),	TGACG-motif (1), GARE-motif(1), p-box(1), ERE(1)
	Os01g0829800	CAT box(1)	MBS(2),	CGTCA-motif(2), TGACG-motif (2), ABRE(1),CE3(1)
	Os01g0649100	CCGTCC-box(1), Skn1-motif(4), circadian(2), GCN4 motif(1)	ARE(2), MBS(3), TC-rich repeats(1)	CGTCA-motif(1), TGACG-motif (1), GARE-motif(1), p-box(1)
	Os12g0632700	Skn1-motif(3),	TC-rich repeats(1)	TCA element(1), TGACG-motif (2), ABRE(2), GARE-motif(1), p-box(1), TATC box(2)
	Os10g0478200	CAT box(1), O2 site(2), Skn1-motif(2),	ARE(1), MBS(1), HSE(1), TC-rich repeats(1), LTR(2)	CGTCA-motif(2), TGACG-motif (2), ABRE(2),
	Os08g0562100	O2 site(2), Skn1-motif(1),	HSE(1)	CGTCA-motif(1), TCA element(1), TGACG-motif (1), ABRE(1), GARE-motif(1),p-box(1), TATC box(1),
	Os05g0574400	CAT box(1), Skn1-motif(1), circadian(1), GCN4 motif(1),	MBS(2), WUN-motif(1), LTR(2),EIRE(1),	CGTCA-motif(1), TCA element(1), TGACG-motif (2), ABRE(1), GARE-motif(1),motif llb(2)
	Os07g0630800	Skn1-motif(1)	MBS(2), box-w1(1), TC-rich repeats(1)	CGTCA-motif(1),TCA element(3), TGACG-motif (1), ABRE(2),ERE(1), AuxRR-core(1)
	Os04g0551200	circadian(1)		TCA element(1)
	Os03g0773800	CCGTCC-box(1) O2 site(1), Skn1-motif(1)	ARE(1), MBS(1),	CGTCA-motif(3), ABRE(4),TGACG-motif (3)

Table 2(B) Cis-regulatory elements in the upstream region of C4 photosynthetic genes of maize. Elements responsive to plant growth and development, stress response and hormones were analyzed.

Gene ID	Gene	Growth and Development	Biotic/Abiotic Stress	Hormonal
GRMZM2G097457	ZmPPDK	O2-site (1), Skn-1_motif (2)	ARE (4), MBS (1)	
GRMZM2G306345		circadian(1), Skn-1_motif(2)	TC-rich repeats(1), ARE(1), HSE(1), MBS(2)	TGACG-motif(1), CGTCA-motif(1)
GRMZM2G122479	ZmNADP-ME	O2-site (2), GCN4_motif (2), Skn-1_motif (3)	MBS (2), ARE (1), Box-W1 (1), TC-rich repeats	ABRE (4), CGTCA-motif (1), TGACG-motif (1)
GRMZM2G159724		Skn-1_motif(2), circadian(1)	MBS(1),LTR(1),HSE(1),Box-W1(1),ARE(1),MNF(1)	TGACG-motif(1) TGA-box(1) TATC-box(1)
GRMZM2G404237		Skn-1_motif(1) circadian(1)	ARE(1) TC-rich repeats(1) LTR(1)	CGTCA-motif(3) TCA-element(1) TGACG-motif(3) TATC-box(1)
GRMZM5G886257		circadian (1) Skn-1_motif (1)	MNF(1) ARE(1) LTR(1)	CGTCA-motif (1) TGACG-motif (1) ABRE(5)
GRMZM2G118770		Skn-1_motif(2)	MBS(1) HSE(2)	
GRMZM2G001696	ZmPEPCK	Skn-1_motif (1)	ARE (2) MBS(3)	ABRE (1) AuxRR-core (1) TCA-element (1)
GRMZM5G870932		Skn-1_motif(1) O2-site (1) circadian(3)	ARE (3) 5UTR Py-rich stretch(2) MBS (1) TC-rich repeats (1)	ABRE(1) CGTCA-motif(1) ERE(1) TGA-element(1) TGACG-motif(1) TCA-element (1)
GRMZM2G082780	ZmPEPC	circadian(2) Skn-1_motif(2) GCN4_motif(1)	ARE(1)	TGA-element(1)
GRMZM2G069542		Skn-1_motif(1) GCN4_motif(1)	MBS(2) ARE(1)	TGACG-motif(1) TCA-element(1) ABRE(1)
GRMZM2G074122		circadian(1) O2-site(1)	TC-rich repeats(2) MBS(1)	

			Box-W1(1) 5UTR Py-rich stretch(1)	
GRMZM2G083841		circadian (2) Skn-1_motif(1) GCN4_motif (4) O2-site (1) as-2-box(1)	MBS (1) MNF(2)	TGACG-motif(1) TGA-element(1) GARE-motif(1)
GRMZM2G110714		O2-site(1) circadian(1) Skn-1_motif (3)	ARE (1) Box-W1 (1) MBS(1)	TGACG-motif (2) ABRE (1) CGTCA-motif (2)
GRMZM2G473001		circadian(2) Skn-1_motif(5)	ARE(1)	CGTCA-motif(3) TGACG-motif(3) TATC-box(1)
GRMZM2G121878	ZMCA		5UTR Py-rich stretch(2) MBS(1)	ABRE(1) GARE-motif(1) P-box(1) motif lib(2)
GRMZM2G348512		circadian(1) Skn-1_motif(2)	TC-rich repeats(1) EIRE(1),MNF(2)	TCA-element(1) GARE-motif(2)
GRMZM2G145101		circadian(1) GCN4_motif(1) Skn-1_motif(4)	ARE(3) TC-rich repeats(1)	CGTCA-motif(1) TCA-element(1) TGACG-motif(1) ERE(3)
GRMZM2G414528		Skn-1_motif(3)	5UTR Py-rich stretch(2) ABRE(7) MBS(2) <b>MNF(1)</b>	CGTCA-motif(1) TCA-element(1) TGA-element(1) TGACG-motif(1)
GRMZM2G068455	ZmNADP-MDH	circadian(1), CCGTCC-box(1)	ARE(1) Box-W1(1) LTR(8)	
GRMZM2G141289		CCGTCC-box(1) Skn-1_motif(2) GCN4_motif(1)	ARE(1) TC-rich repeats(1) LTR(1)	TCA-element(3) GARE-motif(1) TGA-element(1)
GRMZM2G161245		Skn-1_motif (1) GCN4 motif (1) as-2-box (10)	MBS (1) Box-W1 (1)	ABRE (4)
GRMZM2G154595		circadian(3)	TC-rich repeats(3)	motif lib(1)

	Skn-1_motif(3) GCN4_motif(2)	MBS(2) Box-W1(1) ARE(1)	TGA-element(3) TCA-element(2) GARE-motif(1)
GRMZM2G466833	GRMZM2G466833 Skn-1_motif(1) O2-site(1)	ARE(1)	TGACG-motif(1) TCA-element(1) TATC-box(1) GARE-motif(1) ABRE(1) CGTCA-motif(1)
GRMZM2G072744	Skn-1_motif (3)	ARE (2) LTR (1) MBS (2) G-box (2)	CGTCA-motif (1) TATCCAT/C-motif (1) TGACG-motif (1)
GRMZM2G415359	circadian(3) Skn-1_motif(5)	TC-rich repeats(2) MBS(1) L-box(1) EIRE(1)	CGTCA-motif(4)
GRMZM2G035767	O2-site(1) Skn-1_motif(2)	MBS(1)	TGACG-motif (2) CGTCA-motif(2) GARE-motif (1)
GRMZM2G101290	Skn-1_motif(3) CCGTCC-box(1)	ARE(1) MBS(1)	CGTCA-motif(1) TCA-element(1) TGACG-motif(1) TGA element(2)

Table (3 A) Multilevel consensus signature sequences of motifs of rice defined by Motif scan software

PROTEIN ID	SIGNATURE SEQ 1	SIGNATURE SEQ 2
Os05g0405000	EFFSFGTNDLTQMTFGYSR PEP-utilizing enzymes signature 2.	GGMHAAVGILTARGGMTSHAAVVAR PEP-utilizing enzymes phosphorylation site signature.
Os03g0432100	EFFSFGTNDLTQMTFGYSR PEP-utilizing enzymes signature 2.	GGMNAAAGILTARGGMTSHAAVVAR PEP-utilizing enzymes phosphorylation site signature.
Os01g0188400	FNDDIQGTASVVLGALL Malic enzymes signature.	KERDAHY TYR_PHOSPHO_SITE Tyrosine kinase phosphorylation site.
Os01g0723400	FNDDIQGTASVVLGSLV Malic enzymes signature.	TWSKGRAVFA Aminotransferases class-II pyridoxal-phosphate attachment site.
Os01g0743500	FNDDIQGTAAVVLAGLI Malic enzymes signature.	KERDAHY Tyrosine kinase phosphorylation site.
Os02g0665000	NOT FOUND	NOT FOUND
Os05g0186300	FNDDIQGTASVVLAGLI Malic enzymes signature.	RPDDLVKY Tyrosine kinase phosphorylation site.
Os03g0255500	LIGDDEHCWSDTGVSN Phosphoenolpyruvate carboxykinase (ATP) signature.	
Os04g0592500	RDQEVSY Tyrosine kinase phosphorylation site.	
Os10g0204400	LIGDDEHCWSDNGISN Phosphoenolpyruvate carboxykinase (ATP) signature	
Os01g0110700	VLTAHPTQINRR Phosphoenolpyruvate carboxylase active site 1.	VMVGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
Os01g0208700	VLTAHPTQSVRR Phosphoenolpyruvate carboxylase active site 1.	IMIGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
Os01g0758300	VLTAHPTQSVRR Phosphoenolpyruvate carboxylase active site 1.	VMIGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
Os02g0244700	VFTAHTQSVRR Phosphoenolpyruvate carboxylase active site 1.	VMIGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
Os08g0366000	VLTAHPTQSVRR Phosphoenolpyruvate carboxylase active site 1.	VMIGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
Os09g0315700	VLTAHPTQSVRR Phosphoenolpyruvate carboxylase active site 1.	VMIGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
Os01g0639900	CADSRVCP Prokaryotic-type carbonic anhydrases signature 1.	EYAVCALVELIVVIGHSRCG Prokaryotic-type carbonic anhydrases signature 2.
Os09g0464000	CADSRVCP Prokaryotic-type carbonic anhydrases signature 1.	EFAVNTLEVENVLVVGHSRCG Prokaryotic-type carbonic anhydrases signature 2.
Os11g0153200	NOT FOUND	NOT FOUND
Os09g0454400	SEHTINGTRFDAEMHMV Alpha-carbonic anhydrases signature.	
Os08g0434300	VTTLDVVRANTFV Malate dehydrogenase active site signature.	
Os01g0829800	VTTLDVVRANTFI Malate dehydrogenase active site signature.	RNCDITSY Tyrosine kinase phosphorylation site.
Os01g0649100	VTTLDVVRAKTFY	

	Malate dehydrogenase active site signature	
Os12g0632700	VTTLDVVRANTFV	
	Malate dehydrogenase active site signature.	
Os10g0478200	LTRLDHNRALGQI	
	Malate dehydrogenase active site signature.	
Os08g0562100	LTRLDENRAKQQL	RSKGDGDY
	Malate dehydrogenase active site signature.	Tyrosine kinase phosphorylation site.
Os05g0574400	VTTLDVVRAKTFY	
	Malate dehydrogenase active site signature.	
Os07g0630800	VTTLDVVRANTFV	
	Malate dehydrogenase active site signature.	
Os04g0551200	LTRLDHNRALGQV	
	Malate dehydrogenase active site signature.	
Os03g0773800	VTTLDVARANTFV	
	Malate dehydrogenase active site signature.	

Table (3 B) Multilevel consensus signature sequences of motifs of maize defined by Motif scan software

Protein ID	Signature seq 1	Signature seq 2
NP_001168699.1 (CA)	CSDSRVCP carbonic anhydrases signature 1.	EYAVCALKVEV LVVIGHSCCG Prokaryotic-type carbonic anhydrases signature2
NP_001151431.1 (CA)	CSDSRVC Prokaryotic-type carbonic anhydrases signature 1.	EYAVCALKVEVLVVIGHSCCG Prokaryotic-type carbonic anhydrases signature 2.
NP_001147028.1 (CA)	Not found	Not found
NP_001142031.1 (CA)	SEHTVDGRRFPVELHVM Alpha-carbonic anhydrases signature.	
NP_001149686.1 (CA)	CADSRVCP Prokaryotic-type carbonic anhydrases signature 1.	EFAVNTLQVENVLVIGHSRCG Prokaryotic-type carbonic anhydrases signature 2.
NP_001146392.1 (CA)	Not found	Not found
NP_001152905.1 (CA)	CSDSRVCP Prokaryotic-type carbonic anhydrases signature 1.	EYAVCALKVEVIVVIGHSRCG Prokaryotic-type carbonic anhydrases signature 2.
NP_001140385.1 (CA)	CADSRVCP Prokaryotic-type carbonic anhydrases signature 1.	
NP_001150123.1 (CA)	CSDSRVCP Prokaryotic-type carbonic anhydrases signature 1.	EYAVCALKVEVIVVIGHSRCG Prokaryotic-type carbonic anhydrases signature 2
NP_001149032.1	SEHTVDGRRYAMELHVM Alpha-carbonic anhydrases signature	
NP_001105359.1	CSDSRVCP Prokaryotic-type carbonic anhydrases signature 1.	EYAVCALKVQVIVVIGHSCCG Prokaryotic-type carbonic anhydrases signature 2.
NP_001154820.1	VFEALKNQTVDLVFTAHTQSARR Phosphoenolpyruvate carboxylase active site 1.	VMVGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
NP_001105438.1	VLTAHTQSVRR Phosphoenolpyruvate carboxylase active site 1.	VMIGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
NP_001105503.1	VFTAHTQSVRR Phosphoenolpyruvate carboxylase active site 1.	VMIGYSDSGKDAG Phosphoenolpyruvate carboxylase active site 2.
NP_001105808.1	IGRGRFGTVRRCYAAATGEPFALK Protein kinases ATP-binding region signature.	VAHRDVKPDNLLF Serine/Threonine protein kinases active-site signature.
NP_001105774.1	IGRGRFGVVRRCVSRATGEFAVK Protein kinases ATP-binding region signature.	VVHRDVKPDNVLL Serine/Threonine protein kinases active-site signature.
NP_001105773.1	IGRGRFGVVHRCTSRATGEFAVK Protein kinases ATP-binding region signature	VAHRDVKPDNILI Serine/Threonine protein kinases active-site signature.
NP_001105772.1	IGRGRFGVVHRCTSRATGEFAVK Protein kinases ATP-binding region signature	VAHRDVKPENILI Serine/Threonine protein kinases active-site signature.
NP_001130365.1	KDVLEGDPY Tyrosine kinase phosphorylation site.	
NP_001296837.1	LIGDDEHCWSDNGVSN Phosphoenolpyruvate carboxykinase (ATP) signature.	
NP_001296837.1	LIGDDEHCWSDNGVSN Phosphoenolpyruvate carboxykinase (ATP) signature.	
NP_001146178.1	LIGDDEHCWSENGVSN	

	Phosphoenolpyruvate carboxykinase (ATP) signature.	
NP_001105738.2	EFFSFGTNDLTQMTFGYSR PEP-utilizing enzymes signature 2.	GGMHAAVGILTERGGMTSHAAVVAR PEP-utilizing enzymes phosphorylation site signature.
NP_001141918.1	Not found	Not found
NP_001167723.1	KWNERAY Tyrosine kinase phosphorylation site.	
NP_001105313.1	FNDDIQGTASVVLAGLL Malic enzymes signature.	KAYELGLATRLPPPSDLVKYAENCMY Tyrosine kinase phosphorylation site.
NP_001105383.2	FNDDIQGTASVVLAGLL Malic enzymes signature.	KLLNDEFY Tyrosine kinase phosphorylation site.
NP_001105292.1	FNDDIQGTAAVVLAVGIV Malic enzymes signature.	KERDTHY Tyrosine kinase phosphorylation site
NP_001150965.1	FNDDIQGTASVVLAGLL Malic enzymes signature.	TWTKGRAVFA Aminotransferases class-II pyridoxal-phosphate attachment site.
NP_001183119.1	FNDDVQGTAGVALAGLL Malic enzymes signature.	KLLEDKLYL Tyrosine kinase phosphorylation site.
NP_001152396.1	FNDDIQGTASVVLAGLI Malic enzymes signature.	RERETHY Tyrosine kinase phosphorylation site.
NP_001147966.1	KITGEFGY Tyrosine kinase phosphorylation site.	
NP_001105420.1	LTRLDENRAKQCL MDHactive site signature.	RSKGDGDY Tyrosine kinase phosphorylation site.
NP_001132228.1	VTTLDVVRANTFV MDHactive site signature.	
NP_001105603.1	LTRLDHNRALGQI MDHactive site signature	
NP_001241749.1	VTTLDVVRANTFV MDHactive site signature.	
NP_001148518.1	VTTLDVARANTFV MDHactive site signature.	
NP_001142100.1	VTTLDVVRAKTFYA MDHactive site signature.	
NP_001147160.1	LTRLDHNRALGQI MDHactive site signature.	
NP_001132302.1	VTTLDVVRANTFV MDHactive site signature.	
NP_001132077.2	VTTLDVVRANTFV MDHactive site signature.	
NP_001141337.1	VTTLDVVRAKTFY NADP-MDHactive site signature.	
NP_001307728.1	Not found.	

The maize genome size is greater than rice genome but has two fewer chromosomes relative to rice. Chromosomal map for rice genes was constructed by using chromosome map tool available in Oryza base Integrated Science Database (<http://viewer.shigen.info/oryzavw/maptool/MapTool.do>) and mapviewer [http://www.ncbi.nlm.nih.gov/projects/mapview/map\\_search.cgi?taxid=4577&build=100.0](http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=4577&build=100.0) and maize bin viewer of MaizeGDB for maize.

#### Phylogenetic analysis

Amino acid sequences of all the identified C3/C4 photosynthetic genes from rice and maize were aligned separately using ClustalW and the phylogenetic tree was constructed using N-J method of MEGA version 4.0.02 [31–32]. Each node was tested using the bootstrap approach by taking 5000 replications to ascertain the reliability of nodes. The number indicated percentages against each node.

#### Analysis of Signature and Phosphorylation site prediction

To identify the signature sequences within the protein sequences of C3/C4 photosynthetic genes, the deduced protein sequences of all the C3/C4 photosynthetic genes in rice and maize were analyzed using online SMART motifscan ([http://myhits.isb-sib.ch/cgi-bin/motif\\_scan](http://myhits.isb-sib.ch/cgi-bin/motif_scan)) and fingerPRINTscan (<http://www.ebi.ac.uk/Tools/pfa/fingerprints/>). Prediction of Serine, threonine and tyrosine specific Phosphorylation site prediction of C3/C4 protein sequences (Supp Table 1) is done by employing tools NetPhos 2.0 server [33].

#### Cis- acting regulatory elements/ promoter analysis

For promoter analysis, sequences 1,000 bp upstream of the initiation codon of the putative all C3/C4 genes such as CA, PEPC, PEPCK, NADP-ME, NADP-MDH and PPDK were retrieved and subjected to search using CARE program (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) of Plant CARE database to identify cis-regulatory elements [34].

#### Expression Analysis using publically available Microarray data

Data expression data of rice C3 C4 genes were retrieved from project Web site (<http://www.ricearray.org/expression/expression.php>) and Gene expression omnibus (<http://www.ricearray.org/expression/expression.php>). The signature was considered to be significant if it uniquely identified an individual gene and showed perfect match (100% identity over 100% length of the tag). The probe ID of C3/C4 genes of maize were retrieved from the Maize genome database (<http://www.plantgdb.org/ZmGDB/>) using under different developmental stages anatomical, developmental and stress condition after that expression result visualized by Geneinvestigator. The Expression Omnibus platform accession numbers GSE6901 using for rice. We applied the R package for Quantile normalization of data for removing array to array variation and Z-score conversion for finding out significantly expressed gene in a particular tissue. We selected those genes which have z-score  $\geq 1.64$  means 90% confidence level. We remove all genes that have standard deviations from the local mean values less than 1.64 for filtering the data. For the cluster analysis of genes related to calcium transporters, we used hierarchical clustering technique using Cluster 3.0. This is a reliable and useful way for analyzing all sorts of large genomic datasets using calculate the distance matrix.

## Result And Discussion

#### Chromosomal Mapping of Maize photosynthetic pathway genes

All of the 6 genes [37 isoforms viz CA-11, PEPC-4, PEPCK-2, PPDK-3, NADP-ME-6 and NADP-MDH-11 isoforms] of C4 photosynthetic pathway were mapped across all ten chromosomes of Zea mays. While for one gene of PEPCK [NM\_001309908.1] no matches were found, out of 11 genes of NADP-MDH two genes [NM\_001309908.1 and NM\_001147865.1] were not placed while for one gene [NM\_001320799.1] no matches were found. On further analysis it was found that only a single gene of beta Carbonic Anhydrase [NM\_001111889.1] were present as two copies at a position on larger arm of chromosome 3 and 8 shown connected by a blue line, while rest of them were present as single copies. Genes of carbonic anhydrase were present in 4 alpha and 7 beta form. Alpha forms were concentrated on Chromosome 1, 2, 4 and 9 while beta forms were concentrated on 2, 3, 7 and 8. Most of the genes were scattered across 10 chromosomes except Malate dehydrogenase, 5 isoforms of which were present on Chromosome 1 and rest were distributed across chromosome 1, 3, 4 and 6., while out of 2 genes of PEPCK only one [NM\_001152706.1] is found on longer arm of Chromosome 9. Out of three genes of PPDK two were present on chromosome 6 while Chromosome 10 had only single gene. 6 isoforms of NADP-ME were distributed across Chromosome 3, 5, 6 and 8. All four isoforms of PEPC genes were scattered respectively on different chromosome namely 4, 5, 7 and 9.

#### Chromosomal Mapping of Rice photosynthetic pathway genes

All of 6 genes [30 isoforms viz CA-4, PEPC-6, PEPCK-3, PPDK-2, NADP-ME-5 and NADP-MDH-10 isoforms] of C3 photosynthetic pathway were mapped across all twelve chromosomes of Oryza sativa and it was found that all genes existed as single copy. All genes were scattered but most of isoforms were concentrated on Chromosome 1 and no gene was present on Chromosome 6, while single isoforms were concentrated on Chromosome 7, 11 and 12. Out of four isoforms of Carbonic anhydrase two alpha forms were present on Chromosome 9 and 11 while two beta forms were present on Chromosome 1 and 9. Two isoforms of gene PPDK were present on Chromosome 3 and 4. Out of four isoforms of gene NADP-ME, three were present on Chromosome 1 and one was present on Chromosome 2. Isoforms of PEPC were present on Chromosome 1, 2, 8 and 9. Isoforms of NADP-MDH were present on 1, 3, 4, 5, 7, 8, 10 and 12. No gene was present on Chromosome 6.

#### Chromosomal distribution and the synteny between rice and maize

There is some co-linearity between rice and maize at whole genome level. The orthologs from photosynthetic gene families in both rice and maize were mapped and their corresponding chromosome locations are summarized (Fig. 1A & B). Expansion and inversion of some chromosome region were also revealed during comparative synteny in rice. Most of the genes present on Chromosome 1 of rice showed synteny with genes present on Chromosome 1, 3, 6, 8

and 9. Distribution of photosynthetic genes across chromosome of rice and maize is not uniform but phylogenetic analysis has revealed macro colinearity in between the isoforms of gene to a great degree. Isoforms of genes present on Chromosome 1 of rice showed a higher synteny with genes of maize with beta form of Carbonic Anhydrase [Os01g0639900] of rice present on chromosome 1 depicting synteny with 4 other beta isoforms (NM\_001175228.1) present on Chromosome 8 and (NM\_001156651.1, NM\_001159433.1 and NM\_001157959.1) present on Chromosome 3 respectively. 2 isoforms of NADP-ME (Os01g0723400 and Os01g0743500) on rice showed synteny with NADP-ME (NM\_001158924.1) on Chromosome 3 of maize. Alpha form of Carbonic Anhydrase of rice (Os11g0153200) on Chromosome 11 showed synteny with beta form of Carbonic anhydrase (NM\_001155560.1 and NM\_001148559.1) on Chromosome 2 of maize. Alpha form of Carbonic anhydrase (Os09g0454400) present on Chromosome 9 showed synteny with alpha Carbonic anhydrase (NM\_001155560.1 and NM\_001156214.1) present on Chromosome 1 and Chromosome 2. Two genes of PPKK present on Chromosome 3 and 5 (Os03g0432100 and Os05g0405000) of rice showed synteny with PPKK Gene (NM\_001112268.2) of maize on Chromosome 6. Also PPKK gene of maize (NM\_001174252.1) on Chromosome 6 showed synteny with PPKK gene (NM\_001148446.1) on Chromosome 10 of maize.

#### Comparative phylogenetic analysis between rice and maize individual C3/C4 genes

Phylogenetic analysis of photosynthetic genes of C4 and C3 plants [*Zea mays* and *Oryza sativa*] was done and evolutionary trees were constructed using MEGA 4. These evolutionary trees were used to develop relationship among isoforms of same gene among C4 and C3 plants [*Zea mays* and *Oryza sativa*] depicted through chromosomal map. Amino acid sequences of C3 and C4 photosynthetic genes of rice and maize were aligned using Clustal W to study their phylogenetic relationship by constructing rooted (Fig. 2a,b,c,d,e & f) tree using MEGA4. Rooted trees for 6 genes were constructed and analysed for synteny. Phylogenetic trees of 6 genes showed variance to a great degree. In carbonic anhydrase, Phylogenetic analysis revealed tree to be constructed of two big clusters A (only beta form) and B (only alpha form), where cluster A was subdivided into two small clusters a1 and a2. a1 was further dividing into x and y, while a2 into m and n clusters. Cluster B was dividing into two small cluster b1 and b2. In cluster a1 Beta isoform of CA (Os01g0639900) of rice present on chromosome 1 showed higher degree of synteny with 4 other beta isoforms (NM\_001175228.1) present on Chromosome 8 and (NM\_001156651.1, NM\_001159433.1 and NM\_001157959.1) present on Chromosome 3 respectively and NM\_001156651.1, NM\_001159433.1 also showed synteny with each other (100%). In cluster a2 beta form of CA of rice (Os09g0464000) present on chromosome 9 showed synteny with CA of maize NM\_001146913.1 and NM\_001156214.1 present on chromosome 2 and 7 respectively. In cluster B alpha form of CA of rice (Os09g0454400 and Os11g0153200) present on Chromosome 9 and 11 showed synteny with CA of maize (NM\_001148559.1 and NM\_001155560.1) present on Chromosome 2 and 1 respectively. In PPKK, phylogenetic analysis of PPKK gene showed tree to be dividing into 2 clusters A and B. In cluster A gene of PPKK present of rice (Os03g0432100 and Os05g0405000) present on Chromosome 3 and 5 showed synteny with PPKK gene of maize (NM\_001112268.2) present on Chromosome 6. Cluster B showed relationship between PPKK gene of maize NM\_001174252.1 and NM\_001148446.1 present on Chromosome 6 and 10 respectively. In PEPCK, Phylogenetic analysis revealed tree to be dividing into two clusters A and B respectively. In cluster A gene of PEPCK of rice (Os03g0255500, Os10g0204400) present on Chromosome 3 and 10 respectively showed synteny with PEPCK gene of maize (NM\_001152706.1) present on Chromosome 9 and (NM\_001309908.1) for which no matches were found. Cluster B showed a single gene of PEPCK (Os04g0592500) present on Chromosome 4 in relation with Cluster A. In PEPC, Phylogenetic analysis revealed tree to be dividing into two big clusters A and B respectively. Cluster A was further divided into two cluster big a and small b, where cluster a was again divided into one big a1 and one small cluster a2. a1 was further divided into x(x1 and x2) and y clusters. Cluster y revealed PEPC gene of rice (Os02g0244700) present on chromosome 2 in synteny with two isoforms of PEPCK gene of maize (NM\_001136893.1, NM\_001112033.1) present on Chromosome 4 and 5. Cluster x1 revealed PEPC gene of rice (Os09g0315700) present on Chromosome 9 in synteny with PEPC gene of maize (NM\_001111968.1) present on Chromosome 7. In NADP-ME, phylogenetic analysis revealed tree to be dividing into two clusters big A and small B respectively. Cluster A was further divided into one big cluster a1 and one small cluster a2. Cluster a1 into x and y and x into x1 and x2. x1 was further divided into m and n. Cluster x2 revealed relationship between isoform of NADP-ME gene of rice (Os01g0743500) present on Chromosome 1 with isoform of NADP-ME gene of maize (NM\_001111822.1) present on Chromosome 8. Cluster x1 revealed 3 relationships involving cluster m1, m2 and n. Cluster m1 revealed relationship between isoform of NADP-ME gene of rice (Os01g0723400) present on Chromosome 1 with isoform of NADP-ME gene of maize (NM\_001158924.1) present on Chromosome 3. Cluster m2 revealed relationship between isoform of NADP-ME gene of rice (Os05g0186300) present on Chromosome 5 with isoform of NADP-ME gene of maize (NM\_001157493.1) present on Chromosome 6. Cluster n revealed relationship between isoform of NADP-ME gene of rice (Os01g0188400) present on Chromosome 1 with two isoforms of NADP-ME gene of maize (NM\_001111843.1, NM\_001111913.2) present on Chromosome 3 and 6 respectively. In NADP-MDH, phylogenetic analysis revealed tree to be dividing into two big clusters A and B respectively. Cluster A was further divided into 4 clusters a, b, c and d and cluster B into two x and y respectively. Cluster a1 was further divided into x and y and x into x1 and x2. x1 was further divided into m and n. Cluster a1 revealed relationship between two isoforms of NADP-ME gene of rice (Os01g0649100, Os05g0574400) present on Chromosome 1 and 5 with two isoforms of NADP-ME gene of maize (NM\_001148628.2) present on Chromosome 6 and NM\_001147865.1 which was not placed. Cluster b1 revealed relationship between isoform of NADP-ME gene of rice (Os12g0632700) present on Chromosome 12 with isoform of NADP-ME gene of maize (NM\_001138605.2) present on Chromosome 3. Cluster b2 revealed relationship between isoform of NADP-ME gene of rice (Os03g0773800) present on Chromosome 3 with isoform of NADP-ME gene of maize (NM\_001155046.1) present on Chromosome 1. Cluster c revealed relationship between isoform of NADP-ME gene of rice (Os07g0630800) present on Chromosome 7 with isoform of NADP-ME gene of maize (NM\_001320799.1) for which no matches were found. Cluster d consisting of d1 and d2 revealed relationship between two isoforms of NADP-ME gene of rice (Os01g0829800, Os08g0434300) present on Chromosome 1 and 8 with three isoforms of NADP-ME gene of maize (NM\_001138830.1, NM\_001254820.1 and NM\_001138756.1) on Chromosome 1, 1 and 4 respectively. Cluster x revealed relationship between isoform of NADP-ME gene of rice (Os08g0562100) present on Chromosome 8 with isoform of NADP-ME gene of maize (NM\_001111950.1) present on Chromosome 1. Cluster y1 revealed relationship between isoform of NADP-ME gene of rice (Os08g0562100) present on Chromosome 8 with isoform of NADP-ME gene of maize (NM\_001111950.1) present on Chromosome 1. Cluster y2 revealed relationship between isoform of NADP-ME gene of rice (Os10g0478200) present on Chromosome 10 with isoform of NADP-ME gene of maize (NM\_001112133.2) present on Chromosome 1 and NM\_001153688.1 which was not placed.

#### Evolutionary relationship of C3/C4 photosynthetic pathway genes in rice and maize

A rooted tree was constructed to examine evolutionary relationship among C3 and C4 photosynthetic genes of rice and maize respectively by aligning their amino acids through Clustal-W method and tree was constructed using MEGA 4.0 Minimum Evolution method (Fig. 3A). Evolutionary phylogenetic tree analysis revealed tree to be constructed of two big clusters:- A] Cluster A consists of gene of NADP-MDH, Carbonic Anhydrase and PPDK. B] Cluster B consists of gene of NADP-ME, PEPC and PEPCK. Cluster A consist of 4 small clusters. Cluster a reveal a relationship among NADP-MDH gene of rice and maize. Evolutionary analysis shows NADP-MDH gene of glyoxysomal and mitochondrial origin of rice and maize to have evolved together and also revealing a relation between glyoxysomal and chloroplastic NADP-MDH gene of rice and maize respectively. Cluster b also revealed relation between cytosolic and chloroplastic genes of NADP-MDH of rice and maize. Cluster c revealed a relation between chloroplastic origin of beta CA of rice and maize. Cluster d revealed a unique relationship between single isoform of PEPCK gene of rice (Os04g0592500) and two isoforms of PPDK gene of maize (NP\_001141918.1, NP\_001167723.1). Cluster B consist of three clusters. Cluster e revealed relation between alpha CA of rice and maize and that of NADP-ME of maize with PEPC gene of rice. Cluster f and g revealed relation between PEPC and NADP-ME of rice and maize respectively Also individual analysis of rice and maize genes was also done by constructing tree respectively by aligning their amino acids through Clustal-W method showed in (Fig. 3B & 3C). Figure 3B depicts evolutionary tree of photosynthetic genes of rice and depict common origin of all photosynthetic genes respectively except that of PEPCK gene (Os04g0592500). Figure 3C depicts evolutionary tree of photosynthetic genes of maize and depict common origin of all photosynthetic genes respectively except that of PPDK gene (NP\_001141918.1).

#### Cis-Regulatory Elements Analysis

A total of 25 CARE [14 in rice and 11 in maize] viz ARE, MBS, TC-rich repeats, 5 UTR Py-rich Strech, GC-motif, HSE, LTR, L-Box, Box-W1, GCC-Box, Pc-CMA2a, WUN motif, MNF, EIRE and ABRE were observed in rice (-1 kb) and maize (-5 kb). Upstream regions were analyzed using Plant CARE to study correlation of transcriptional regulation of C3 and C4 photosynthetic genes among rice and maize. These elements were mainly associated with plant growth and development, stress response and hormone responsiveness, meristem-specific expression but here we focus on stress responsiveness element which frequency distribution both rice and maize shown in Fig. 4 & listed in supplementary Table 1). A total number of 25 cis-regulatory elements (CREs) [14 in rice and 11 in maize] which had stress responsive role were identified namely anoxia-response element (ARE), MYB-binding site (MBS), defense and stress responsive elements (TC-rich repeats), wound-responsive element (WUN-motif), heat shock element (HSE), ABA-response element (ABRE), 5 UTR Py-rich Strech, LTR, L-Box, Box-W1, GCC-Box, Pc-CMA2a, WUN-motif, MNF and EIRE. Name and frequencies of identified CARE with respect to C4 photosynthetic genes of rice and maize is listed in respectively Table 2A and 2B. Presence of the stress related motifs showed altered gene expression of photosynthetic genes under stress. In Rice, Anoxia-response elements (ARE) involved in anaerobic induction was most commonly observed in almost all genes in moderate frequency with higher expression in NADP-MDH gene Os08g0434300). MYB-binding site (MBS) was also present in moderate frequency in almost all genes with higher expression in PEPCK gene (Os10g0204400) and NADP-MDH gene (Os01g0649100). TC-rich repeats present in moderate frequency in almost all genes. 5 UTR Py-rich Strech were observed in moderate frequency in almost all genes with most expression in NADP-MDH gene (Os03g0773800). GC-motif present in moderate quantity in few genes with most expression in PPDK gene (Os05g0405000) and in Carbonic Anhydrase (Os01g0639900). HSE heat shock element factor involved in heat stress response were present in moderate frequency in few genes with most expression observed in Carbonic Anhydrase (Os11g0153200). LTR and L-box were present in low frequency in only 4 different genes. Box- W1- fungal elicitor responsive element (TTGACC) were present in low frequency in only 3 genes. GCC-Box, WUN-motif and Pc-CMA2a present in only 3, 3 and 2 genes with low frequency. MNF and EIRE were present in low frequency in only single genes. In Maize, Anoxia-response elements (ARE) involved in anaerobic induction was most commonly observed in almost all genes in moderate frequency with higher expression in PPDK gene (ZM2G097457). MYB-binding site MBS) were also present in moderate frequency in almost all genes with higher expression in PEPCK gene (ZM2G001696). TC-rich repeats present in moderate frequency in few genes with most expression in NADP-MDH gene (Zm2G154595). 5 UTR Py-rich Strech were observed in moderate frequency in 4 genes. HSE involved in heat stress response were present in low frequency in only 3 genes. LTR were present in low frequency with exception of NADP-MDH gene (Zm2G068455) with higher expression. L-box was present in low frequency in only single gene. Box- W1- fungal elicitor responsive element (TTGACC) was present in low frequency in only 7 genes. EIRE was present in low frequency in only single genes. MNF was present in low frequency in only 5 genes. ABRE present in higher frequency in only single gene of Carbonic Anhydrase (ZM2G414528).

#### Post translational modification (phosphorylation) analysis of C3/ C4 protein sequences of rice and maize

Graphs of phosphorylation site of 6 photosynthetic genes of C3 and C4 plants were examined and data was analysed and summarized in (Supp Table 2A &, Fig. 5A & 5B) In Os-CA Serine specific phosphorylation sites were present in abundance in all isoforms of genes of rice while that of Threonine and Tyrosine varied in all isoforms with most Threonine specific sites present in Os09g0454400 and most Tyrosine specific sites were present on Os01g0639900. In general more phosphorylation sites were present on beta CA than alpha CA in rice. 2] Zm-CA In maize Serine specific phosphorylation sites were present in abundance in all isoforms of genes with highest being in NP\_001105359.1. In general Threonine specific phosphorylation sites were less present with their absence in NP\_001151431.1. Tyrosine specific phosphorylation sites were also present in moderate quantity in all isoforms of genes of CA. In general NP\_001105359.1 isoform of beta CA gene of maize showed higher number of phosphorylation sites. Here clearly revealed that in comparison of both maize and rice more phosphorylation active sites were present on beta CA isoforms of gene and more present on maize especially in NP\_001105359.1. 3] Os-PEPC Almost equivalent number of Serine, Threonine, and Tyrosine specific phosphorylation sites present in all isoforms of PEPC genes. 4] Zm-PEPC, Almost equivalent number of Serine, Threonine, and Tyrosine specific phosphorylation sites present in all isoforms of PEPC genes except NP\_001130365.1 where less sites are present. Note that, In comparison of both maize and rice more phosphorylation active sites were present on rice. 5] Os-PEPCK Almost equivalent number of Serine and Threonine specific sites was present on all isoforms of PEPCK genes of rice. Tyrosine specific sites were present in moderate number in all isoforms except Os04g0592500 where number was less. 6] Zm-PEPCK, Almost equivalent number of Serine, Threonine, and Tyrosine specific phosphorylation sites present in all isoforms of PEPCK gene of maize. It to be note that in comparison of both maize and rice more phosphorylation active sites were present on rice especially Os03g0255500. 7] Os-NADP-MDH Only equivalent number of low Tyrosine specific phosphorylation active sites was present on all isoforms of genes while that of Serine and Threonine specific sites number varied. Serine specific sites were present more in Os01g0649100 of glyoxysomal origin

while Threonine specific sites were present in higher number in Os03g0773800 of rice. 8] Zm-NADP-MDH, Number of sites for Serine, Threonine and Tyrosine specific phosphorylation sites varied in number for all isoforms of genes. In general number of Threonine specific sites was present in higher quantities in Zm-NADP-MDH compared to all other genes in maize and rice. More number of phosphorylation sites was present on NP\_001132228.1 of glyoxysomal origin. Here clearly revealed that In comparison of both maize and rice more phosphorylation active sites were present on maize especially NP\_001132228.1. 9] Os-NADP-ME Moderate number of Serine, Threonine and Tyrosine specific sites were present on all isoforms while more number of sites was present on Os05g0186300. 10] Zm-NADP-ME Almost equivalent number of Serine, Threonine, and Tyrosine specific phosphorylation sites present in all isoforms of NADP-ME gene of maize with most expression in NP\_001105313.1. In comparison of both maize and rice more phosphorylation active sites were present on maize especially NP\_001105313.1. 11] Os-PPDK Almost equivalent number of Serine, Threonine, and Tyrosine specific phosphorylation sites present in all isoforms of PPDK gene of rice. 12] Zm-PPDK, Number of Serine, Threonine, and Tyrosine specific phosphorylation sites varied with most number of sites present in NP\_001105738.2. In comparison of both maize and rice more phosphorylation active sites were present on maize especially NP\_001105738.2.

#### Identification of protein signature

A table of signature sequences of rice and maize were listed and following observations were made and showed in (Table 3A & 3B), isoforms of PPDK gene revealed two conserved signature sequences. Isoforms of NADP-ME gene revealed two sequences:- one conserved and one variable among all isoforms. Two isoforms of PEPCK gene showed common signature [ATP signature] while one Os04g0592500 showed varied signature [Tyrosine kinase phosphorylation site]. Six isoforms of PEPC gene showed two common signature sequences. Beta Carbonic anhydrase isoforms showed two common signature sequences while only one alpha Carbonic Anhydrase showed one common signature while no signature sequences were found in Alpha CA(Os11g0153200). NADP-MDH isoforms showed one common signature sequences to all while only a second tyrosine kinase phosphorylation site was present in (Os01g0829800 and Os08g0562100). While maize, All alpha and beta isoforms of Carbonic Anhydrase had different conserved sequences while no sequences were found for alpha CA (NP\_001146392.1). All PEPC isoforms have two signature sequences varying for each gene. All NADP-MDH isoforms have almost single conserved sequence with no signature sequences found for NP\_001307728.1. All NADP-ME isoforms have almost two conserved signature sequences with single sequence present in NP\_001147966.1. All 9 PEPC isoforms of gene have almost same signature sequences. All two PEPCK isoforms of gene have same signature sequences.

Expression profiling during plant anatomical stage and across different plant tissues shows differential transcriptional regulation in rice and maize.

In rice, we observed a much diverse pattern of expression of C4 genes during anatomical stage (Fig. 6A). The genes which showed very low expression throughout the anatomical stage were OsPEPCK-2, OsPEPCK-3, OsPEPC-1 and OsPEPC-2. It was further revealed that while several C4 genes show near constant expression pattern (at either a down regulated or up regulated, a few members showed tissue specific expression and few are ubiquitously expressed across all plant tissues. OsPEPC-3 showed down regulated expressions and moderate up regulated expressions were observed at Suspension, stroma, ovary, developing anther, mature anther, embryo sac and embryo stage. Cytosolic enzyme, OsNADP-MDH- 5, uniquely showed predominantly high up regulated expression at all stages. Further observations revealed that chloroplastic OsNADP-ME-1, OsPEPC-4, chloroplastic OsCAAlfa-1, glyoxysomal OsNADP-MDH-3 and glyoxysomal OsNADP-MDH-7 were found to be fairly consistent up regulated. OsPPDK-1 was found to be high up regulated at stages which include internode pith parenchyma, root tip, spikelet, embryo and endosperm, which highlights its significant role in plant nutrition and protection and down regulated at stoma and ovary stage. Chloroplastic OsPPDK-2 was found to give active up regulated expression at callus, suspension cell and dry seed stage. Chloroplastic, NADPME-2 showed overall moderate up regulated expression. Cytoplasmic OsNADPME-3, Chloroplastic OsNADPME-4 and OsNADPME-5 showed overall down regulated expression. Interestingly, OsNADPME-3 exhibit up regulated expression specifically at dry seed, embryo-sac and endosperm stage. OsPEPCK-1 showed tissue specific function and was high up regulated for leaf, root, flag leaf and endosperm. OsPEPC-5 was not expressed and show down regulated expression for coleoptiles and germination seed. Chloroplastic OsCABeta-1 showed varied expression from down regulated to moderate up regulated expression. OsCA Beta-2, a chloroplastic enzyme, exhibit down regulated expression for developing anther which is crucial reproductive structure. For all other stages the expressions were up regulated. OsCA Alfa-2 exhibit down regulated expression at all anatomical stage. Both glyoxysomal OsNADP MDH-1 and OsNADPMDH-2 showed moderate up regulated expression except for down regulated expression for dry leaf and flag leaf stage respectively. Glyoxysomal OsNADP MDH-4 expressed fairly up regulated but expressed down regulated at whole plant stage. Further study revealed that chloroplastic OsNADP MDH-6 was down regulated at dry seed, coleoptiles and root tip stage where as it showed high up regulated expression for all other specific tissues. Glyoxysomal OsNADP MDH-8 and cytosolic OsNADP MDH-9 showed no significant role across all anatomical stage. Glyoxysomal OsNADP MDH-10 was found to be differentially expressed across the various anatomical stage. It showed high expression for specific tissues including whole plant, seedling shoot, leaf, flag leaf, spikelet and down regulated expression for suspension cell, root tip, SAM and developing anther. It showed moderate up regulated expression for other specific tissues.

To know some facts, functions and diverse expressions of C4 genes in *Zea mays*, we analyzed the micro-array based expression pattern at different anatomical stage (Fig. 6B)

It was observed that ZmCA Beta-1 showed up regulated expression for juvenile leaf, blade (lamina), adult leaf, and seedling and leaf stage and fairly up regulated expression for shoot. Both the chloroplastic enzymes, ZmCA Beta-2 and ZmCA Beta-3 exhibit regulated expression for adult leaf, juvenile leaf and foliar leaf and shoot stage. ZmCA Beta-2 showed up regulated expression for ovary and seedling stage while ZmCA Beta-3 expressed fairly up regulated in mesocotyl, leaf and seedling stage. The overall expression of chloroplastic ZmCA Alpha-1 was down regulated across all anatomical stage. Chloroplastic enzymes, ZmMDH1 and ZmNADPME-1, showed nearly same expressions at all stages. Both expressed up regulated at adult leaf, blade (lamina), juvenile leaf, foliar leaf and seedling stage which holds important function in complex structure of leaf. Both are fairly up regulated at shoot stage. ZmPEPC1 exhibit fairly up regulated expression for giant cell and root cell stage and is down regulated at all other stages. ZmCA Alpha-2, a chloroplastic enzyme showed gross expression of down regulated except for pollen stage which is an important reproductive structure and was highly up regulated. Zm MDH-2, an enzyme localised in glyoxysome, expressed at axis, embryo, glume and root cell stage and showed moderate up regulated expression and showed moderate down

regulated expression for root, root tips and ear inter-node. ZmPEPC-2 exhibit varied expression. It showed up regulated expression at adult, leaf, blade (lamina), juvenile leaf, foliar life, shoot, leaf and seedling stage whereas it is fairly down regulated at shoot apex, ear internode, internode, culm (main stem) and pollen stage. Further observations revealed that cytosolic enzyme Zm NADP-2 uniquely up regulated expressed across all stages and highly up regulated at ear internode, pulvinus and glume stage. Notably, Chloroplastic ZmNADPME-3 showed up regulated expression at embryo and endosperm stage. It confirms that this enzyme played a vital role in fertilization and providing nutrition. ZmPEPCK-1 showed up regulated expression for adult leaf and blade (lamina) and fairly upregulated at juvenile stage, foliar leaf, shoot, leaf and seedling stage and for rest of the stage it showed either no expression or negligible expression. ZmPPDK-1, a chloroplast localised enzyme expressed high up regulated at endosperm stage which clearly shows that it holds nutritive function in plants and its overall expression was down regulated. Mitochondrial Zm MDH-3 showed overall fairly up regulated expression except at root stage where it showed up regulated expression. Zm MDH-4, a cytosolic enzyme showed upregulated expression except at pollen, caryopsis, shoot apex and lateral root stage. Mitochondrial ZmNADPME-4 showed high up regulated expression for pollen and gave no expression that is, down regulated at endosperm and caryopsis stage. Glyoxysomal ZmMDH-5 showed up regulated expression for root tip and fairly up regulated for endosperm and embryo stage. Chloroplastic enzymes ZmCA Beta 4 and ZmCA Beta-5 all down regulated expression across all stages except at pollen stage where ZmCA Beta 4 expressed and predominantly up regulated. Glyoxysomal ZmMDH-6, ZmPEPC-3 and chloroplastic ZmPPDK-2 were expressed at down regulation in most of the stages, with moderate and up regulated expression in some specific stage that indicated their stage specific role.

Expression profiling during plant development and across different plant tissues shows differential transcriptional regulation in rice and maize.

To gain some insights into the possible function of C4 genes in rice, we analyzed the micro-array based expression pattern at different developmental stage (see method). During analysis, we observed a much diverse pattern of expression of C4 genes during all development stages. (Fig. 7A). The enzymes which showed overall down regulated expression throughout the developmental stage studies were OsPEPCK-2, OsPEPCK-3, OsPEPC-1, OsPEPC-2 and OsPEPC3. Interestingly, OsPEPCK-3 showed major involvement only at Panicle 6 and had up regulated expression. OsPEPC-1, OsPEPC-2 and OsPEPC3 genes were expressed in most of the stages with moderate up regulation expression in some specific stage. Amongst the glyoxysomal enzymes, OsNADP MDH- 1, OsNADPMDH- 2, OsNADPMDH- 3 & OsNADPMDH-4, OsNADPMDH- 3 was highly up regulated. OsNADP MDH-1 was moderately down regulated at pre-germination stage where as OsNADP MDH- 1, OsNADPMDH- 2 and OsNADP MDH-4 were found to be moderate up regulated for P1, P2, P3 and P4 stage. OsNADPMDH- 5, a cytosolic enzyme showed most predominantly up regulated expression across all the stages. The expression for OsNADP MDH-6, which is localised in chloroplast, was found to be down regulated at pre-germination stage and moderate down regulated at S5 stage. Other gene enzymes were found to be differentially expressed across the various developmental stages. Among glyoxysomal OsNADP MDH-7 & OsNADP MDH 10, OsNADP MDH - 7 showed up regulated expression at every stage except for S4 and S5 where moderate up regulated expression was found while OsNADP MDH- 10 showed moderate down regulated expression for P1, P2, P3 stage and moderated up regulated expression for callus suspension and germination seedling stage which is crucial for initiating development of plant. In Glyoxysomal OsNADP MDH- 8 and Cytosolic OsNADP MDH- 9, overall expression at all stages is down regulated. Os PEPC-4, OsPEPC- 5 showed up regulated expression except for down regulated expression at germinating seedling stage in OsPEPC-5. Chloroplastic OsCA Beta-2 and OsCA Alfa-1 expressions varied from moderate up regulated to high upregulated. P1, P2 and P3 stage was found to be down regulated in OsCA Alfa-1. Further studies revealed that, OsCABeta-1 and Os CA Alfa-2 showed gross down regulated expression. Chloroplastic OsCABeta-1 showed moderate up regulated expression for stages including pre-germination, tillering stage, S1 and S5. It was further observed that, among chloroplast localised enzyme, OsPPDK1 and OsPPDK2, the later showed down regulated expression for 1st leaf, 2nd leaf, 3rd leaf, tiller initiation, tillering stage, P1, P2, P3,P4, P5 stages whereas OsPPDK1 showed fairly high upregulated expression for the same stages. Chloroplastic enzymes, OsNADP ME-1 & OsNADP ME-2, were found to be consistently upregulated except for the down regulated expression at pre-germination stage in OsNADP-ME-2. Further expressions revealed that, cytosolic OsNADP-ME-3 and chloroplast localised OsNADP-ME-4 and OsNADP-ME-5 showed overall expression of down regulated. At Pregermination, S4 and S5 stage upregulated expression was observed in OsNADP- ME-3. OsPEPCK-1 has shown varied expressions. At 2nd leaf, S4 & S5 stage, expressions were observed most up regulated while at callus suspension, pregermination, 3rd leaf, tiller initiation, tillering stage P5, P6, S1, S2 and S3 moderate up regulated expressions were observed.

To gain insights into potential physiological function in *Z. mays*, we have studied their expression at different stages of plant development and across various plant tissues showed in (Fig. 7B). The enzymes that showed the down regulated expression throughout were chloroplastic enzymes, ZmCA Alpha-2, ZmCA Beta-4, ZmCA Beta-5 and ZmPPDK-2. ZmMDH-4 which is localised in cytosol uniquely showed predominantly up regulated at all developmental stage except for dough stage which is down regulated. ZmCABeta-1 and chloroplastic enzymes ZmCA Beta-2 and ZmCA Beta-3 showed moderate up regulated expression for seedling stage which is important to understand the cotyledon patterns. ZmCABeta-2 exhibit fairly up regulated expression at anthesis and inflorescence formation. ZmCA Alpha-1, a chloroplastic enzyme exhibit overall down regulated expression at all developmental stage. Further, chloroplastic ZmMDH-1 was found to exhibit down regulated expression at dough stage, fruit formation and germination stage while ate up regulated expressions were observed for inflorescence formation and seedling stage, which holds many important functions during reproduction and provide nutrients to fruits and flowers. Further studies revealed that, ZmNADP-ME-1 showed overall down regulated expression except for seedling stage, for which moderate up regulated expression was observed. ZmPEPC-1 and glyoxysomal ZmMDH-2 showed overall moderate down regulated expression at all stages. ZmPEPC-2 expressed up regulated only for seedling stage and its overall expression is down regulated. Further it was observed that ZmNADP ME-2, localised in cytosol, showed varied expressions from down regulated to high regulated. For dough stage and t formation, down regulated expression was observed where as fairly up regulated expression was observed for anthesis and seedling stage. The enzyme expressed most up regulated for inflorescence and stem elongationstage. ZmNADPME-3, a chloroplastic enzyme, exhibits high expression for dough stage and fruit formation and was up regulated. ZmPEPCK, glyoxysomal ZmMDH-6 and ZmPEPC-3, showed overall down regulated expression except for seedling stage, dough stage and germination stage, where respectively these enzymes expressed up regulated. The expression exhibit by chloroplastic ZmPPDK-1 was up regulated for dough stage and fruit formation whereas at rest of the stages, its expression was down regulated. The expression of mitochondrial ZmMDH-3, was overall fairly up regulated except for germination stage, where its expression was up regulated the most. Mitochondrial ZmNADP ME-4 was found to be differentially expressed across the various developmental stages. The enzyme did not expressed for dough stage and fruit formation stage and fairly up regulated at stem elongation, seedling stage and germination stage. Notably, it was

highly expressed for anthesis and inflorescence formation and thus expressed up regulated. ZmMDH-5 which is localised in glyoxysomes consistently expressed down regulated across all developmental stage.

#### Regulation of Transcription of photosynthetic genes in Rice and Maize under Stress

Analysis of expression of C3 and C4 photosynthetic genes under stress conditions was done using publically available microarray databases result of which is showed in Fig. 8A & 8B.

#### Expression of rice genes for stress (Seedling, Drought, Salt and Cold)

NADP-MDH gene isoforms (6 and 12) were down regulated, NADP-MDH (1, 2, 3, 4, 5, 8, 9 and 11) was upregulated while moderate expression was observed in NADP-MDH (7 and 10). PEPC gene isoforms (1, 3, and 6) were upregulated while PEPC isoforms (2, 4 and 5) were downregulated. PEPC gene isoforms (2 and 3) were downregulated while isoform (1) was upregulated. PPDK gene isoform (2) was upregulated and isoform (1) was downregulated. NADP-ME gene isoform (1 and 3) was upregulated while (2, 4 and 5) was downregulated. Carbonic Anhydrase gene isoform alpha (1 and 2) was downregulated while beta (1 and 2) was upregulated.

#### Expression of maize genes for stress (Water and drought)

All 6 isoforms of NADP-MDH gene were downregulated as well as moderate for almost all stress factors. All three isoforms of PEPC gene isoforms were generally downregulated for almost all stress factors with exception for Water (6, 7, 13 and 14) respectively. Single PEPC gene isoform was downregulated for stress factors (Water1-7) moderate for stress factor (Water 8, 9, 10, 11 and 12) while upregulated for stress factor (Water 13 and 14). Two isoforms of PPDK gene showed moderate and downregulation for various stress factors. All 4 isoforms of NADP-ME gene were downregulated as well as moderate for almost all stress factors. All five isoforms of beta Carbonic Anhydrase were almost down regulated for all stress factors except Water (13 and 14) while all both isoforms of alpha Carbonic Anhydrase were also down regulated for all stress factors. Comparative analysis of regulation of photosynthetic genes of rice and maize under stress reveal a total of down regulation of genes of maize in response to stress factors while that of rice was generally upregulated and moderate. This result is corresponding to CARE analysis given in Fig. 4A and B which showed number and frequency of Cis acting regulatory elements high in rice (14 in rice and 11 in maize). CARE such as ARE, MBS, TC-rich repeats, GC motif, HSE were present in higher frequency in rice thus corresponding to the fact of upregulation and moderate expression of C3 photosynthetic genes of rice under stress.

## Conclusion

In this investigation, we conducted a comprehensive insilico investigation and identification C3/C4 photosynthesis gene/protein sequences in rice and maize. A complete overview of CA, PEPC, PEPC, NADPH-ME, NADP-ME and PPDK genes/ protein sequences in rice and maize is presented, including the chromosomal mapping, evolutionary phylogeny relationship, serine, threonine and tyrosine specific phosphorylation site and protein signature for assessment of posttranslational modification and their cis acting regulatory elements analysis. Analysis of protein signature revealed some conserve protein motif present in each protein sequences between rice and maize. Protein phosphorylation analysis clearly revealed role in posttranslational modification in C3/C4 genes in both crops but maximum present in maize. The cis-regulatory element analysis of the C3/C4 photosynthesis gene revealed the major putative functions as regulation of genes associated with plant growth development, abiotic and biotic stress, growth hormone and light response. Presence of high number of stress responsive cis acting element in upstream suggests that these proteins might be unregulated in plant stress tolerance. The expression data needs to be correlated in revealing the function of these protein and role in plant stress. A comprehensive analysis on gene expression may provide a better assessment of the prospective genes/ protein for crop improvement. Genome editing based cisgenic approach may be utilized to validate its potential candidature for crop improvement.

## Declarations

#### Acknowledgements

SP thanks SERB for the 000230/LS grant. This work was financially supported by Department of Science and technology, The facilities provided by DST-funded for carrying out this investigation is also acknowledged.

#### Authors' contributions

SP conceived the idea and design the all experiment, SP and VR performed experiments and wrote the manuscript critically evaluated the manuscript and approved the final version of the manuscript.

#### Funding

This work was financially supported by the Ministry of Science and Technology and SERB, New Delhi India.

#### Availability of data and materials

Most data supporting the results are included in the article. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

#### Ethics approval and consent to participate

This experiment does not involve human experiments and animal experiments.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

## References

1. Karki S, Rizal G, Quick WP. Improvement of photosynthesis in rice (*Oryza sativa* L.) by inserting the C4 pathway. *Rice*. 2013 Oct;28(1):28. 6(.
2. Schuler ML, Mantegazza O, Weber AP. Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. *Plant J*. 2016 Jul;1(1):51–65. 87(.
3. Christin PA, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J, Osborne CP. Adaptive evolution of C 4 photosynthesis through recurrent lateral gene transfer. *Current Biology*. 2012 Mar 6;22(5):445-9.
4. Yerramsetty P, Agar EM, Yim WC, Cushman JC, Berry JO. An *rbc L* mRNA-binding protein is associated with C3 to C4 evolution and light-induced production of Rubisco in *Flaveria*. *Journal of experimental botany*. 2017 Jul 20;68(16):4635–49.
5. Miyao M, Masumoto C, Miyazawa SI, Fukayama H. Lessons from engineering a single-cell C4 photosynthetic pathway into rice. *J Exp Bot*. 2011;62(9):3021–9. doi:10.1093/jxb/err023.
6. Christin PA, Boxall SF, Gregory R, Edwards EJ, Hartwell J, Osborne CP. Parallel recruitment of multiple genes into C4 photosynthesis. *Genome biology and evolution*. 2013 Oct 31;5(11):2174-87.
7. Slack CR, Hatch MD. Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochemical Journal*. 1967 Jun;103(3):660.Sage RF, Sage TL, Kocacinar F. Photorespiration and the evolution of C4 photosynthesis. *Annual review of plant biology*. 2012 Jun 2;63:19–47.
8. Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U. The role of photorespiration during the evolution of C4 photosynthesis in the genus *Flaveria*. *elife*. 2014 Jun;16;3:e02478.
9. Hutchinson S, Gomes C, Alleyne D, Phillips W. An assessment of the economic and social impacts of climate change on the agriculture sector in the Caribbean.
10. Peterhansel C, Horst I, Niessen M, Blume C, Kebeish R, Kürkcüoglu S, Kreuzaler F. Photorespiration. *The Arabidopsis Book*. 2010 Mar:e0130.
11. Singh J, Pandey P, James D, Chandrasekhar K, Achary V, Kaul T, Tripathy BC, Reddy MK. Enhancing C3 photosynthesis: an outlook on feasible interventions for crop improvement. *Plant biotechnology journal*. 2014 Dec 1;12(9):1217-30.
12. Gowik U, Westhoff P. The path from C3 to C4 photosynthesis. *Plant Physiol*. 2011;155(1):56–63.
13. Anderson LE. Chloroplast and cytoplasmic enzymes II. Pea leaf triose phosphate isomerases. *Biochimica et Biophysica Acta (BBA)-Enzymology*. 1971 Apr 14;235(1):237–44.
14. Schuler ML, Mantegazza O, Weber AP. Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. *Plant J*. 2016 Jul;1(1):51–65. 87(.
15. Wang C, Guo L, Li Y, Wang Z. Systematic comparison of C3 and C4 plants based on metabolic network analysis. *BMC systems biology*. 2012 Dec 12;6(2):S9.
16. Furbank RT, Taylor WC. Regulation of photosynthesis in C3 and C4 plants: a molecular approach. *Plant Cell*. 1995 Jul;7(7):797.
17. Langdale JA. C4 cycles: past, present, and future research on C4 photosynthesis. *The Plant Cell*. 2011 Nov 1;23(11):3879-92.
18. Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant DW, Zhou W, Xu J. Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. *Nature biotechnology*. 2014 Nov 1;32(11):1158-65.
19. Wang L, Peterson RB, Brutnell TP. Regulatory mechanisms underlying C4 photosynthesis. *New Phytol*. 2011 Apr;190(1)(1):9–20.
20. Hatch MD. C4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics*. 1987 Jan 1;895(2):81–106.
21. Shen Z, Dong XM, Gao ZF, Chao Q, Wang BC. Phylogenic and phosphorylation regulation difference of phosphoenolpyruvate carboxykinase of C3 and C4 plants. *Journal of Plant Physiology*. 2017 Jun 30;213:16–22.
22. Lian L, Wang X, Zhu Y, He W, Cai Q, Xie H, Zhang M, Zhang J. Physiological and photosynthetic characteristics of indica Hang2 expressing the sugarcane PEPC gene. *Molecular biology reports*. 2014 Apr 1;41(4):2189-97.
23. Alvarez CE, Detarsio E, Moreno S, Andreo CS, Drincovich MF. Functional characterization of residues involved in redox modulation of maize photosynthetic NADP-malic enzyme activity. *Plant Cell Physiol*. 2012 Apr;17(6):1144–53. 53(.
24. Badia MB, Arias CL, Tronconi MA, Maurino VG, Andreo CS, Drincovich MF, Wheeler MC. Enhanced cytosolic NADP-ME2 activity in *A. thaliana* affects plant development, stress tolerance and specific diurnal and nocturnal cellular processes. *Plant Science*. 2015 Nov 30;240:193–203.
25. Fukayama H, Hatch MD, Tamai T, Tsuchida H, Sudoh S, Furbank RT, Miyao M. Activity regulation and physiological impacts of maize C 4-specific phospho enol pyruvate carboxylase overproduced in transgenic rice plants. *Photosynthesis Research*. 2003 Aug 1;77(2):227 – 39.
26. Ludwig M. The Roles of Organic Acids in C4 Photosynthesis. *Frontiers in plant science*. 2016;7.

27. Masumoto C, Miyazawa SI, Ohkawa H, Fukuda T, Taniguchi Y, Murayama S, Kusano M, Saito K, Fukayama H, Miyao M. Phosphoenolpyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. *Proceedings of the National Academy of Sciences*. 2010 Mar 16;107(11):5226-31.
28. Taniguchi Y, Ohkawa H, Masumoto C, Fukuda T, Tamai T, Lee K, Sudoh S, Tsuchida H, Sasaki H, Fukayama H, Miyao M. Overproduction of C4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C4-like photosynthetic pathway into rice. *Journal of Experimental Botany*. 2008 Mar 2;59(7):1799 – 809.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990 Oct 5;215(3):403–10.
30. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 1997 Sep 1;25(17):3389 – 402.
31. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*. 1994 Nov 11;22(22):4673-80.
32. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution*. 2007 May 7;24(8):1596-9.
33. Blom N, Kreegipuu A, Brunak S. PhosphoBase: a database of phosphorylation sites. *Nucleic acids research*. 1998 Jan 1;26(1):382-6.
34. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic acids research*. 2002 Jan 1;30(1):325-7.

## Figures

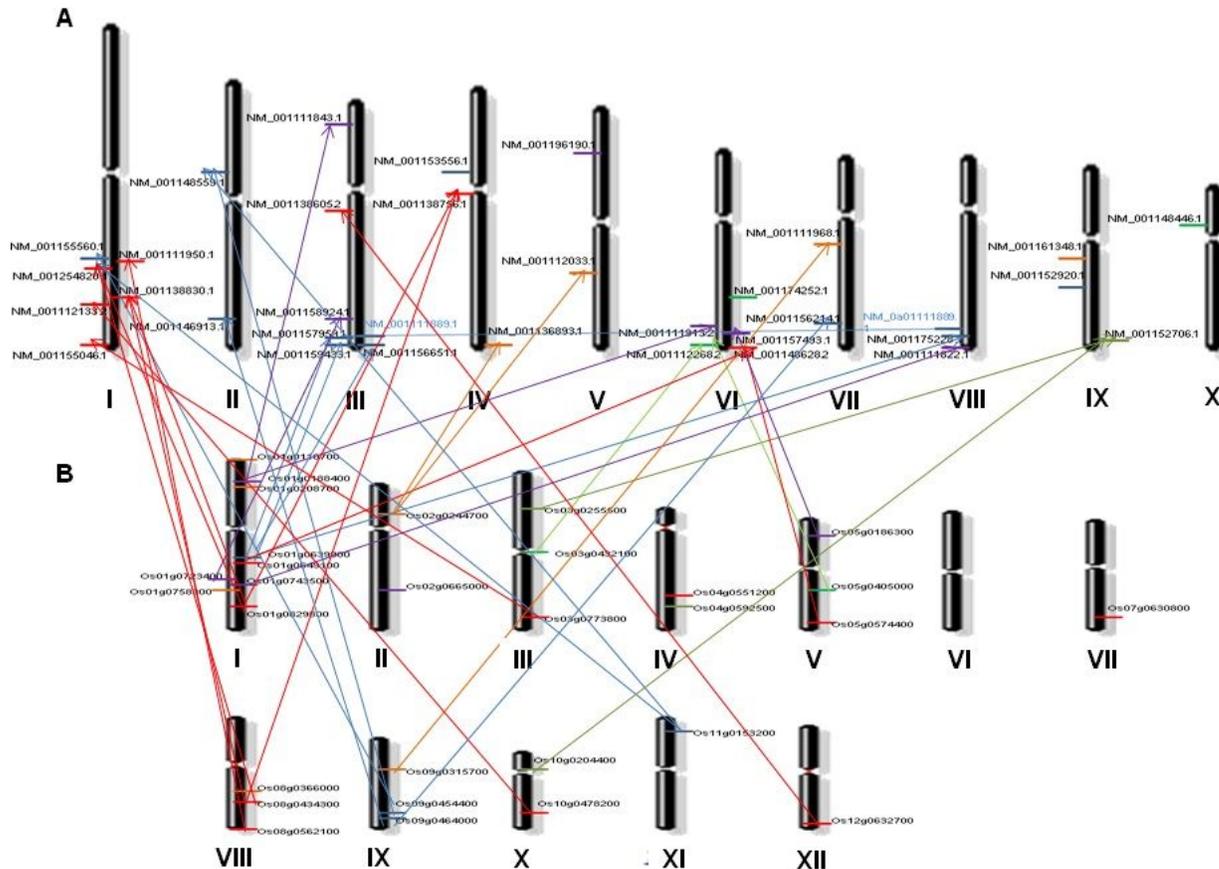
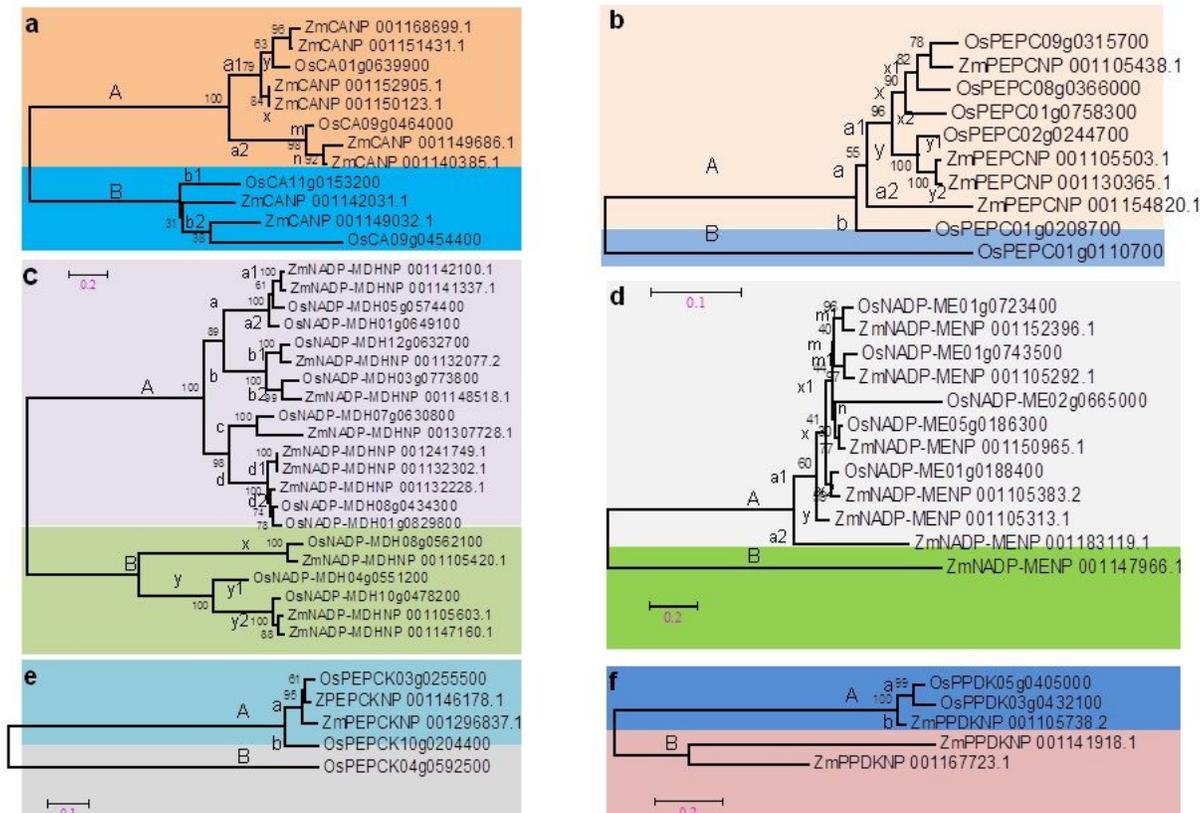


Figure 1

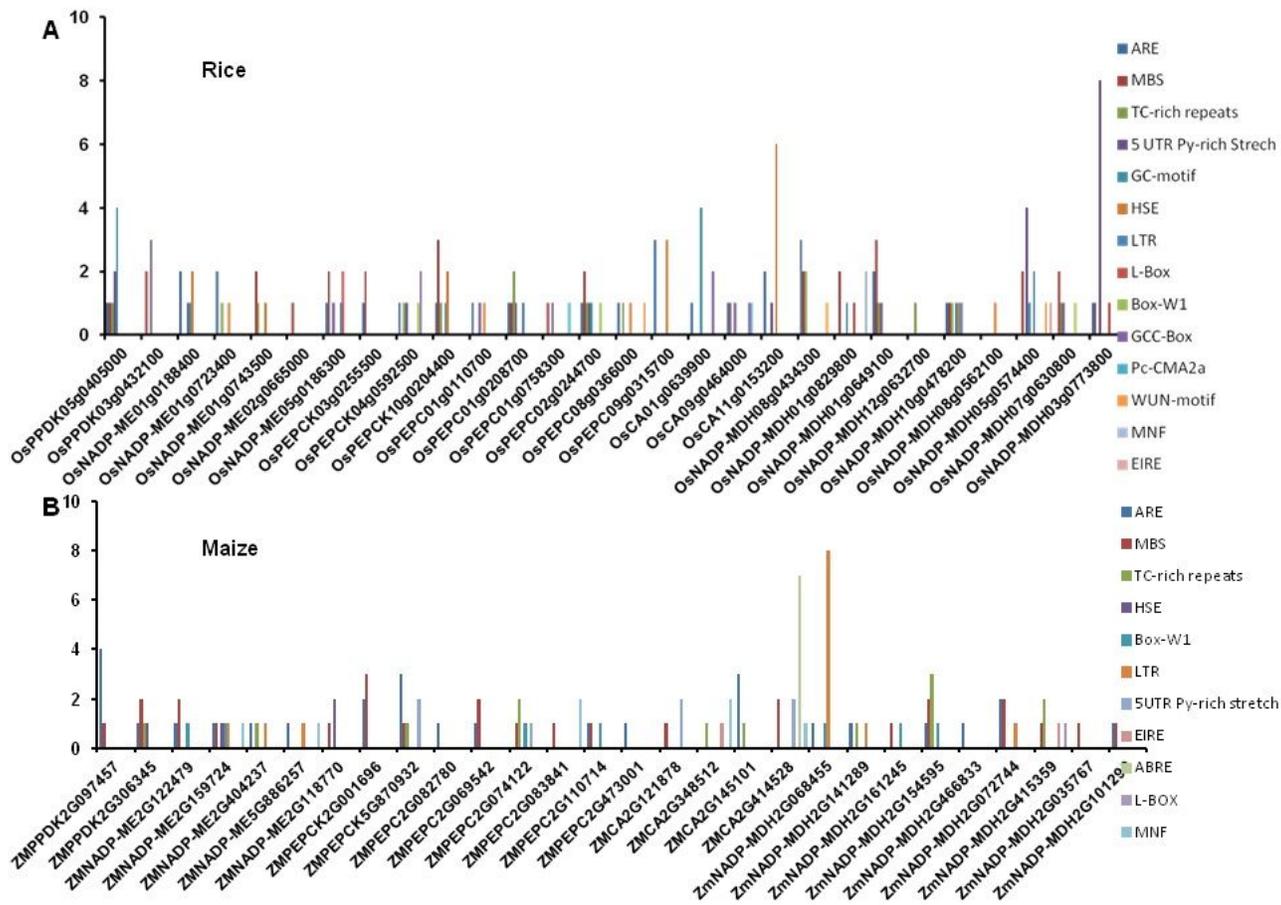
Genomic distribution and synteny mapping of C3/C4 photosynthetic genes from rice and maize



**Figure 2**

Phylogenetic tree of all CA, PEPC, NADP-MDH, NADP-ME, PEPCK, PPK protein sequences of rice and maize All genes shoed individual tree Fig 1. [A] for CA, [B] for PEPC, [C] NADP-MDH, [D] for NADP-ME, [E] for PEPCK & [F] for PPK, The phylogenetic tree was constructed by N-J method using MEGA version 4.0.02.





**Figure 4**  
 Prediction and frequency distribution of Cis-regulatory elements in the upstream region of CA, PEPC, NADP-MDH, NADP-ME, PEPC, PPDK genes rice and maize

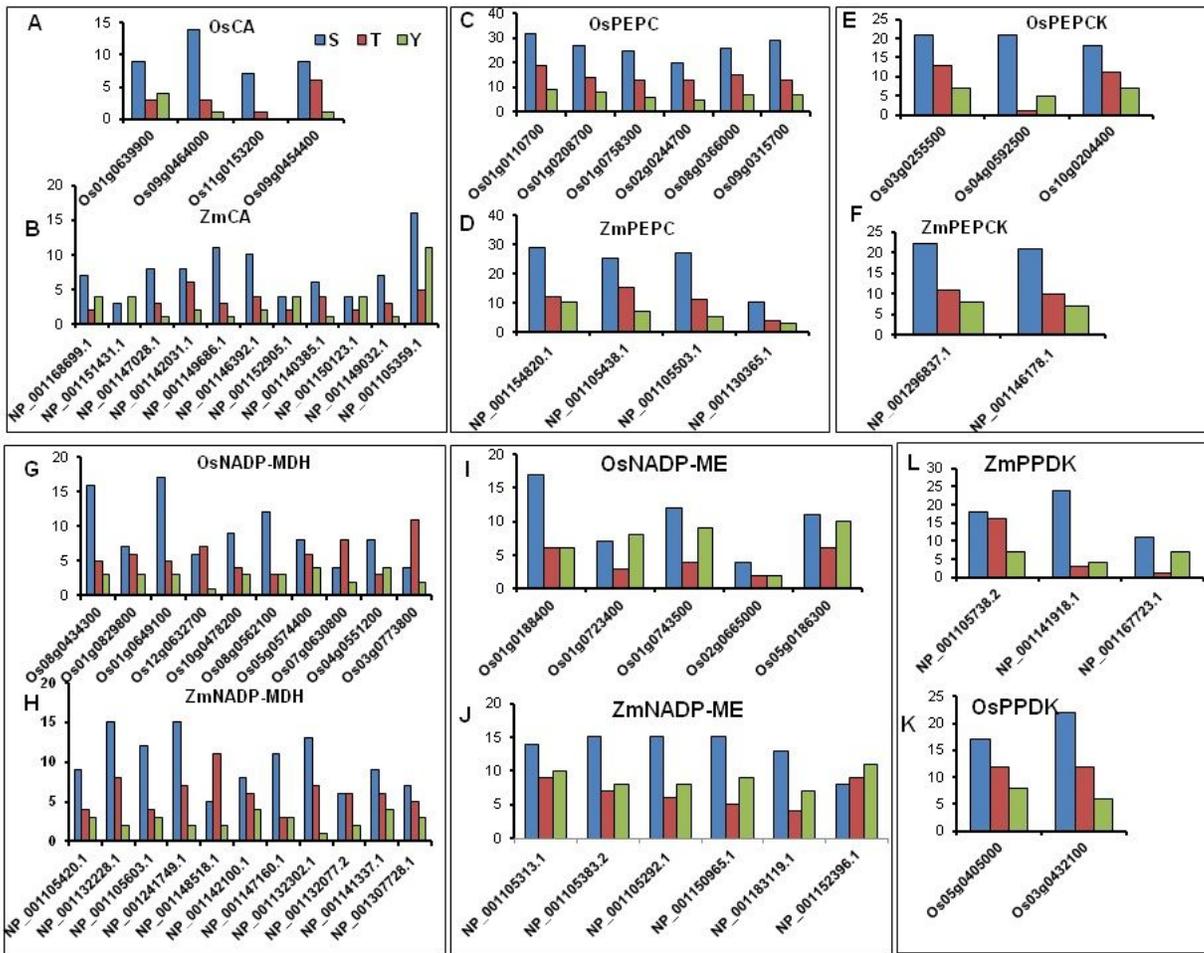


Figure 5

Protein phosphorylation frequency prediction (Serine, threonine & tyrosine specific) using Netphos (CBS Server)

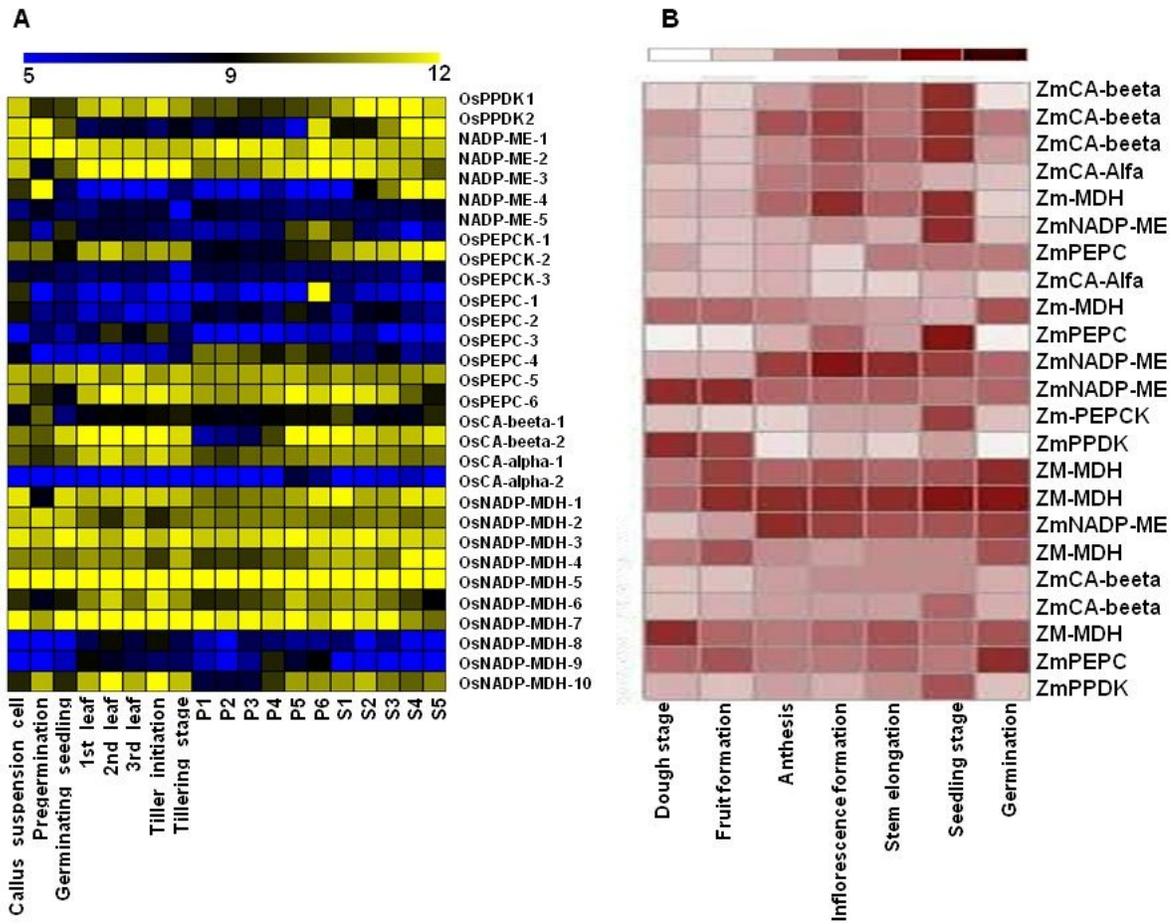


Figure 6

Expression profiling of C3/C4 photosynthesis genes of Rice and Maize at different developmental stages using microarray data, Heat map and hierarchical cluster display differential expression profile for above genes. Various stages are listed in the temporal order of development. The colour bar on top represents relative expression values.

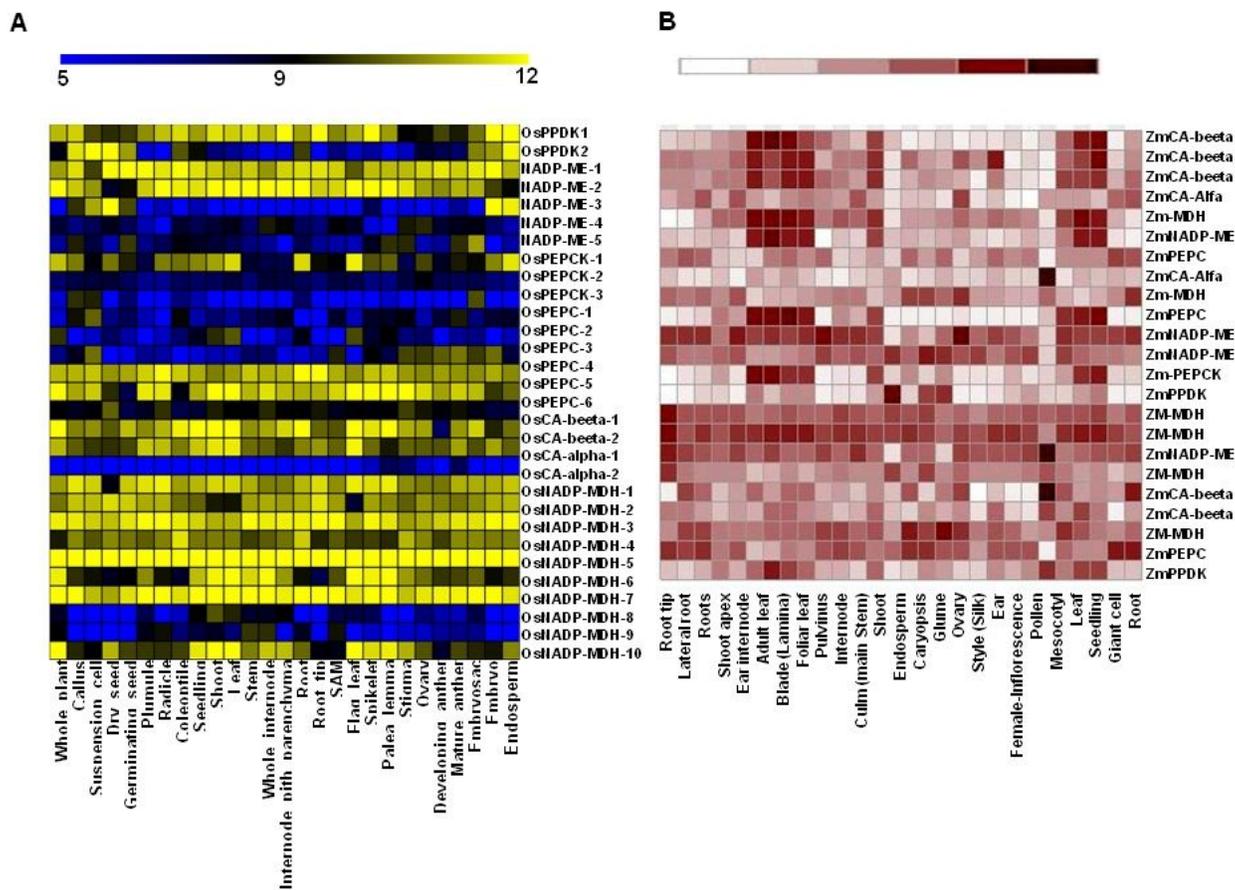


Figure 7

Expression profiling of C3/C4 photosynthesis genes of Rice and Maize at different Anatomical (tissue specific) using microarray data, Heat map and hierarchical cluster display differential expression profile for above genes. Different tissue are listed in the temporal order of development. The colour bar on top represents relative expression values.

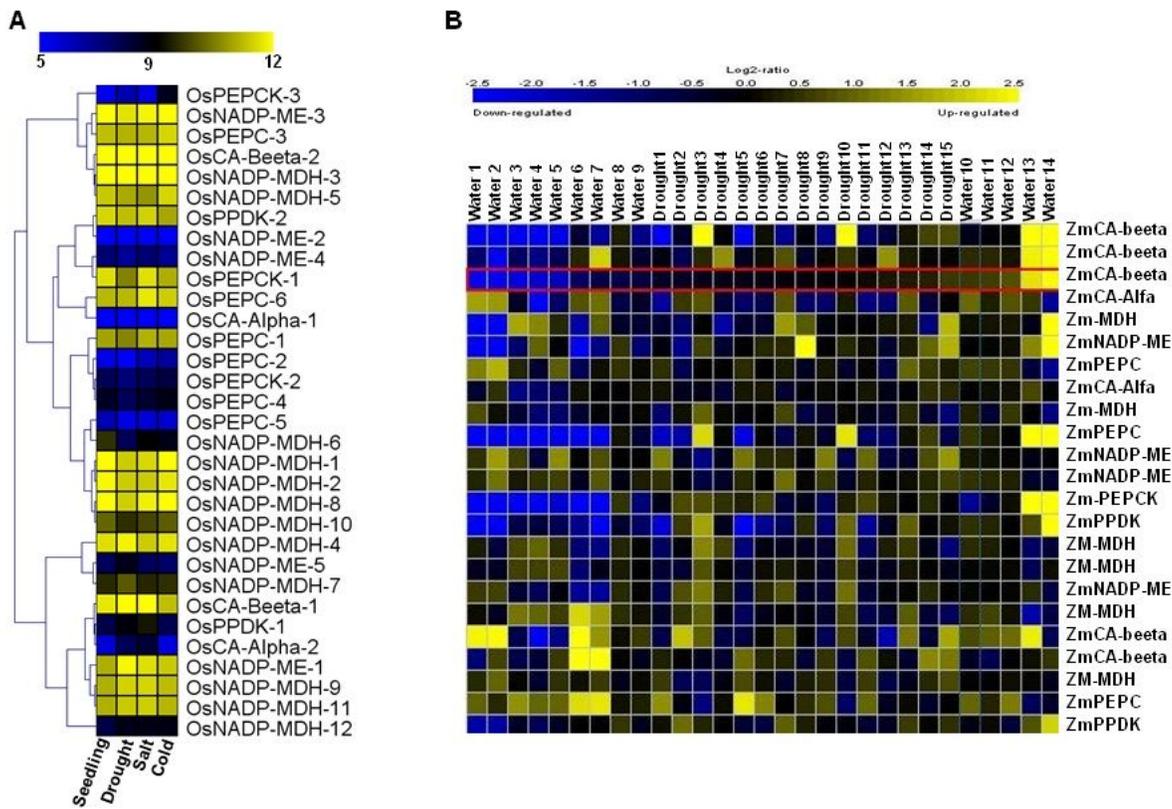


Figure 8

Expression profile of C3/C4 photosynthesis genes under abiotic stress conditions. Heat maps show the microarray-based expression pattern of C3/C4 genes from Rice (A) and Maize (B) under conditions of seedling, salt, drought and cold as indicated at the bottom of the heat map. Color bars at the bottom of each of the heat maps show the corresponding scale for log<sub>2</sub> fold change in expression. Heat maps were generated using hierarchical clustering for which weighted average linkage method and Pearson correlation distance metric were used.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S2B.xls](#)
- [S2A.xls](#)
- [S1.xls](#)