Identification and characterization of finger millet *OPAQUE2* transcription factor gene under different nitrogen inputs for understanding their role during accumulation of prolamin seed storage protein

Vikram Singh Gaur¹ · Lallan Kumar² · Supriya Gupta² · J. P. Jaiswal³ · Dinesh Pandey² · Anil Kumar²

Received: 24 July 2017 / Accepted: 2 February 2018 / Published online: 7 March 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

In this study, we report the isolation and characterization of the mRNA encoding *OPAQUE2* (O2) like TF of finger millet (FM) (*Eleusine coracana*) (*EcO2*). Full-length *EcO2* mRNA was isolated using conserved primers designed by aligning O2 mRNAs of different cereals followed by 3' and 5' RACE (Rapid Amplification of cDNA Ends). The assembled full-length *EcO2* mRNA was found to contain an ORF of 1248-nt coding the 416 amino acids O2 protein. Domain analysis revealed the presence of the BLZ and bZIP-C domains which is a characteristic feature of O2 proteins. Phylogenetic analysis of *EcO2* protein with other bZIP proteins identified using finger millet transcriptome data and O2 proteins of other cereals showed that *EcO2* shared high sequence similarity with barley BLZ1 protein. Transcripts of *EcO2* were detected in root, stem, leaves, and seed development stages. Furthermore, to investigate nitrogen responsiveness and the role of *EcO2* in regulating seed storage protein gene expression, the expression profiles of *EcO2* along with an α -prolamin gene were studied during the seed development stages of two FM genotypes (GE-3885 and GE-1437) differing in grain protein content (13.8 and 6.2%, respectively) grown under increasing nitrogen inputs. Compared to GE-1437, the *EcO2* was relatively highly expressed during the S2 stage of seed development which further increased as nitrogen input was increased. The *Eca-prolamin* gene was strongly induced in the high protein genotype (GE-3885) at all nitrogen inputs. These results indicate the presence of nitrogen responsiveness regulatory elements which might play an important role in accumulating protein in FM genotypes through modulating *EcO2* expression by sensing plant nitrogen status.

Keywords Opaque2 · *Eleusine coracana* · Nutritional value · Prolamine

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-018-1150-1>) contains supplementary material, which is available to authorized users.

Vikram Singh Gaur and Lallan Kumar have contributed equally.

✉ Anil Kumar
anilkumar.mbge@gmail.com

¹ College of Agriculture, JNKVV, Waraseoni, Balaghat, Madhya Pradesh, India

² Department of Molecular Biology and Genetic Engineering, College of Basic Sciences and Humanities, G. B. Pant University of Agriculture and Technology, Pantnagar 263145, India

³ Department of Genetics and Plant Breeding, G. B. Pant University of Agriculture and Technology, Pantnagar 263145, India

Abbreviations

DAS	Day after sowing
HPG	High protein genotype,
LPG	Low protein genotype
N	Nitrogen
NUE	Nitrogen use efficiency
O2	Opaque2
PBF	Prolamine-binding factor
qPCR	Real-time PCR
RACE	Rapid amplification of cDNA ends
SSP	Seed storage proteins
UTR	Untranslated region

Introduction

Cereal seeds are important sources of carbohydrates, proteins, and other nutrients in human diet. The proteins present in the cereal seeds or the “seed storage proteins” (SSP) are specifically synthesized during seed development stages and are highly accumulated in endosperm tissue of the monocotyledonous seed. The stored proteins are then used as a nitrogen source for the germinating seedlings. SSPs are traditionally classified into albumin, globulin, prolamin, and glutelin according to their physical properties based on solubility, and the genes encoding these proteins are highly and specifically expressed in maturing seeds resulting in high levels of these proteins in the seeds. Most cereals, such as maize, barley, and wheat, produce prolamines as their major SSP (Shewry and Tatham 1999), but rice accumulates glutelins as the major component. The expression patterns of these SSP genes are primarily regulated at the transcriptional level by the promoter and several motifs have been identified as cis-regulatory elements involved in the endosperm-specific expression of cereal SSP genes (Osborne 1924; Murphy 2000; Gowda et al 2005). Among these, the endosperm box (TGTAAGTNAATNNGA/GTGAGTCAT; N: A, C, G, or T), also known as the -300 element or prolamin element, is well conserved around 300 bp upstream of the transcription start site (Kreis et al. 1985) in many SSP gene promoters (Hartings et al. 1989). It consists of two independent cis-elements, the prolamin box (P-box; TG(T/C/A)AAAG) and the GCN4 motif (TGA(G/C)TCA) (Hammond-Kosack et al. 1993). The P-box, also called as endosperm motif (E-motif), is recognized by plant-specific PBF-DOF transcription factors, and specific Dof proteins interacting with the P-box in different cereals such as in maize (ZmPBF), barley (BPBF), wheat (WPBF), and rice (RPBF) (Vicente-Carbajosa et al. 1997; Mena et al. 1998; Yamamoto et al. 2006) have been identified. The other motif of the E box, i.e., the GCN4 motif, is recognized by members of bZIP transcription factor gene family. The bZIP proteins interacting with the GCN4 box have been identified in different cereals such as O2 in maize, BLZ1 in barley, SPA1 in wheat, and RISBZ1 in rice (Albani et al. 1997; Vicente-Carbajosa et al. 1998; Wu et al. 1998; Onate et al. 1999; Onodera et al. 2001). As the binding sites of both the transcription factors are very close, direct interactions between maize O2 and PBF have also been reported (Vicente-Carbajosa et al. 1997; Marzábal et al. 2008). The GCN4 is also known as the nitrogen element (N-motif) and has been reported to act as a negative *cis*-regulator under low *N* levels (Muller and Knudsen 1993). This indicates that the bZIP transcription factors interacting with the GCN4 motif regulate the expression of seed storage

protein genes by sensing the nitrogen status of the plant during grain filling. Since a number of studies have indicated that nitrogenous fertilizers directly affect the seed protein content of cereals, this characteristic feature of GCN4 motif might be important to understand the molecular mechanisms involved in the regulation of seed storage protein accumulation under varying *N* levels, and further, providing clues for improving nitrogen use efficiency (NUE) in cereal crops. In the present situation, where it has been realized that higher application of nitrogenous fertilizers is causing detrimental effects, considerable research attention has been focused to develop cultivars that can produce high yield under low nitrogen inputs.

Interestingly, the ability to naturally accumulate high levels of seed storage proteins under low nitrogen conditions is shown by finger millet (*Eleusine coracana*) which is an annual robust grass, mainly grown as a cereal grain in the semi-arid tropics and subtropics of the world under rainfed conditions (Fakrudin et al. 2004). The seeds of finger millet are very rich in minerals such as calcium and phosphorous as compared to other cereals, and it is a good source of essential amino acids and vitamins. Finger millet is organic by default as it capitalizes on low levels of fertilizers, indicating that it has high nitrogen use efficiency (Gupta et al. 2011; Kumar et al. 2009, 2016). Probably, it might have developed unique nitrogen sensing and assimilation mechanism that allows the plant to perform well under low nitrogen conditions. However, lack of genomic information similar to rice and other cereals has restricted the researchers to understand underlying mechanisms. Finger millet accumulates prolamines as the major seed storage proteins which make it nutritionally superior over other cereal grains. High accumulation of prolamines in grains of finger millet suggests that it is driven by strong promoters and regulatory proteins. Identification of such regulatory proteins and promoter elements is necessary to understand the mechanism by which finger millet accumulates prolamines which could be further deployed to engineer new finger millet or cereal cultivars that accumulate high amounts of proteins under low nitrogen inputs. In this paper, we report the identification of finger millet O2-like transcription factor and its expression during seed developing stages of two finger millet genotypes differing in seed protein content grown under varying nitrogen inputs.

Materials and methods

Molecular cloning of *O2* mRNA of finger millet from finger millet

Primer designing for isolation of full-length *O2* transcript of finger millet

As finger millet genome sequence was not available, conserved primer approach was first used to isolate partial *O2* fragment. Coding DNA sequences of *O2* like proteins of rice (RISBZ1-BAC83055.1), wheat (SPA-A- CAR85682.1, SPA-B- CAA70216.1 and SPA-D- CAR85689.1), maize (ZmO2-NP_001105421), and barley (BLZ1-CAA56374.1) were downloaded from NCBI, and conserved regions were identified using online available local alignment tool MEME 4.11.2 (Bailey et al. 2006). Primers were manually designed from the conserved regions that could amplify maximum possible region of target finger millet *O2* transcript. The sequences of primers used to amplify partial *O2* gene of finger millet are given in the (Supplementary Table no. 1).

Plant material, RNA isolation, cDNA synthesis, and PCR conditions

The high seed protein (HPG) content finger millet genotype “GE3885” was raised in plastic pots in an autoclaved mixture (1:1) of agropeat and vermiculite at 25 ± 1 °C under 14 h photoperiod in the transgenic glass house facility of the Department of Molecular Biology and Genetic Engineering, GBPUAT, Pantnagar. In cereals, *O2* is highly transcribed during the mid-phase of the grain filling with highest expression during the mid-day (Pietro et al. 1999). To isolate full-length finger millet *O2* transcript, spikes at the S2 stage (i.e., mid-phase of seed development) of finger millet genotype (GE3885) were chosen. The spikes were excised and immediately frozen in liquid nitrogen for RNA isolation.

Total RNA was isolated using the Qiagen RNeasy plant mini kit including the on-column DNA digestion step as described by the manufacturer. RNA samples were analyzed using 1% agarose gel and subsequently quantified using Nanodrop (Thermo Fisher Scientific, USA). 2 µg of total RNA was used to synthesize first-strand cDNA by oligo (dT)₁₈ primer using Revert Aid™ H-Minus First-strand cDNA synthesis kit (Fermentas). PCR was performed in a 25-µl reaction volume containing 0.2-mM dNTPs, 30 ng of each primer, 1.5-mM MgCl₂, 0.8-U Taq DNA polymerase (Fermentas), and 100 ng of cDNA. PCR amplification of *O2* was carried out according to the

following conditions: 5-min initial denaturation at 95 °C followed by 40 cycles of 94 °C for 1 min, 58 °C for 30 s, 72 °C for 30 s, with final extension of 5 min at 72 °C, and final hold at 4 °C. The amplicons obtained by PCR were eluted from gel, cloned in pGEMT Easy cloning vector (Promega), and sequenced using Applied Biosystems 370 at University of Delhi.

RACE for isolating full-length *O2* mRNA

The total RNA isolated from the S2 stage of GE3885 (as described earlier) was used to perform 3' and 5' RACE to isolate the full-length *O2* mRNA. The primers used for amplification of 5' and 3' cDNA (Supplementary Table no. 1) were designed from the previously isolated partial sequence of finger millet *O2* transcript. Both 3' and 5' RACE were carried out using the 3' and 5' RACE Roche Kit and the procedure were followed according to the manufacturer's instructions. However, briefly, the procedure followed is as follows. For carrying out both 3' and 5' RACE, 2 µg of total RNA was first used to synthesize first strand of cDNA. The gene-specific primer 5' GSP1 and oligo (dT)₁₈ anchor primer were used for 5' and 3' RACE cDNA synthesis, respectively. The cDNA, hence, prepared for 5' RACE was purified using high pure PCR product purification kit (Roche) and was poly-A tailed using dATP and terminal transferase (Roche) according to the manufacturer's instructions. For carrying out PCR, the cDNAs were diluted ten times and used as template with internal Gene-Specific Primers as given in the supplementary table no. 1. After first round of amplification, a series of nested PCRs were performed using the PCR product of each step diluted 100 times as template. The first round of cDNA amplification and subsequent rounds of nested PCRs in both 5' and 3' RACE PCR were performed in 25-µl PCR reaction volume containing 1 × KCl buffer (Fermentas) containing .2-mM dNTPs, 30 ng of each primer, 1.5-mM MgCl₂, 0.8-U Taq DNA polymerase (Fermentas), and 100 ng of template. Amplification was carried out according to the following conditions: 5-min initial denaturation at 95 °C followed by 40 cycles of 94 °C for 1 min, annealing temperature of 58 °C for 1 min, 72 °C for 2 min with final extension of 10 min at 72 °C; and, finally, hold at 4 °C. The 5' and 3' RACE PCR products were eluted from gel with the help of gel elution kit (Qiagen) as per manufacturer's instructions and were subsequently cloned into pGEMT cloning vector (Promega). The integrity of the cloned fragments was confirmed through colony PCR and restriction digestion. Nucleotide sequences of the clones were determined through the Applied Biosystems 370 sequencer at the University of Delhi.

Identification and phylogenetic analysis of bZIP proteins from finger millet developing seed transcriptome data

We recently sequenced the developing spike transcriptome of two finger millet genotypes differing in seed calcium content (Kumar et al. 2014). Although the genotypes used to generate the transcriptome sequence data were different, but it provided an excellent resource to identify and characterize all the transcripts encoding bZIP domain proteins expressed during finger millet seed development stages. A local installation of NCBI-BLAST (version 2.2.27) was used to create custom BLAST database from the assembled contigs fasta file of the high and low calcium finger millet genotypes. The Maize O2 protein sequence (GenBank accession no. NP001105421) was used as a query to retrieve its homologous transcript and the transcripts encoding bZIP domain.

Protein sequences of O2 mRNAs reported in other crops (downloaded from NCBI), protein sequence of the isolated full-length finger millet O2 mRNA, and the protein sequences of the transcripts identified using the developing seed transcriptome data were aligned using the Clustal W multiple sequence alignment tool available online (<http://www.genome.jp/tools/clustalw/>, Kyoto University Bioinformatics Centre). The resultant tree file (with the extension-.dnd) was downloaded and a phylogenetic tree was constructed using the neighbourhood joining method (Tamura et al. 2007) using the software MEGA Ver. 7. To identify conserved motifs, the O2 sequences of other cereal crops were analyzed using the online tool MEME Version 4.11.2 (<http://meme-suite.org/tools/meme>) (Bailey et al. 2006). The maximum number of motifs was set to 20 with minimum and maximum width 6 and 50, respectively, while the other factors were of default selections. The deduced amino acid sequence of O2 mRNA isolated from finger millet along with O2 of other cereal crops was analyzed for the presence of putative phosphorylation sites using NetPhos2.0 (Blom et al. 1999).

Detection of EcO2 transcripts in different tissues and expression analysis of EcO2 and α -Prolamin gene in developing stages of two finger millet genotypes differing in seed protein content under different nitrogen inputs

Two FM genotypes, namely, GE 3885 (HPG) and GE 1437 (LPG), that have contrasting grain protein (13.8 and 6.2%) but nearly similar carbohydrate contents (64.8 and 75.6%), respectively, were raised in earthen pots (10 kg) in poly-house conditions at 25 °C/18 °C temperature over a 14-h/12-h (day/night) schedule. The details of the genotypes under study are given in the supplementary table no. 2. The plants were irrigated daily and exogenous application of three

doses of nitrogen fertilizer in the form of urea were given at the rate of 20 kg/ha (low), 40 kg/ha (normal), and 60 kg/ha (high) in three split doses at the three critical stages, viz., one-third at the time of sowing, one-third at the vegetative growth (60 DAS), and one-third at the onset of grain filling stage (90 DAS). A control experiment of no nitrogen dose was also included in the study. Developing seed tissue samples of the stages—S₁ (booting or inflorescence emergence), S₂ (anthesis), S₃ (grain filling), and S₄ (grain ripening or maturation) (Mirza et al. 2015)—were collected in the forenoon (between 0800 and 0900 h) and were immediately frozen in liquid nitrogen and stored at – 80 °C till further use. At least three independent biological replicates of each tissue sample were collected. Total RNA isolation and cDNA synthesis was carried out as described previously. Real-time PCR (qPCR) was carried out using the 5' Real Master Mix SYBR ROX (Eppendorf India Limited, Chennai, India) according to the manufacturer's instructions. To further understand the expression profiles of the isolated O2 gene, an EST of an α -prolamin gene of *Eleusine coracana* (GenBank accession no. CX265222) was identified and gene-specific primers were subsequently designed for carrying out real-time PCR. Since both the genes are members of multigene families, to ascertain the specificity of the primers, and to rule out the possibility of amplifying non-specific targets, the amplicons produced by both the gene-specific primers produced expected size bands which were subsequently cloned, sequenced and verified. The list of gene-specific primers used for O2 and prolamin gene (GenBank Acc. no. CX265222) expression analysis is given in the supplementary table no. 1. To study the tissue-specific expression levels of EcO2, total RNA was isolated from roots, stem, leaves, and all the stages of developing spikes of finger millet genotype GE 3885; and was converted to cDNA as described earlier. Using EcO2 gene-specific primers, expression profiling was carried out in all the tissues. The tubulin gene primer (GenBank accession no. CX265249) was used as an internal control. The temperature profiles used for the qPCR were 95 °C for 2-min initial denaturation followed by 40 cycles of 95 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s. All samples were amplified in triplicate, and the mean value was considered. The normalized transcript expression was quantified using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001) and presented as fold change over control. In both the cases (detecting expression levels of EcO2 throughout the plant and under different nitrogen inputs), the fold change in expression was calculated relative to the S1 stage of seed development.

Statistical analysis of expression data

All samples were amplified in triplicate and the data were analyzed using the statistical software “Statistica” version

8.0. Three factor completely randomized design comprising two genotypes as first level, four treatments (control, low, normal, high—nitrogen doses) as the second level and the four tissues as (S_1 -, S_2 -, S_3 -, and S_4 -developing seed stages) as the third level was considered for analyzing and generating interaction graphs of the real-time gene expression data.

Results

Isolation of partial and full-length O2 transcript through 3' and 5' RACE

The conserved primer set designed to amplify the target finger millet O2 transcript resulted in two distinct amplicons of 585- and 500-bp sizes, respectively (Supplementary Fig. 1). Both the bands were eluted, cloned, and subsequently sequenced. By performing different NCBI-BLAST searches using the nucleotide sequence of both the bands as query, only nucleotide sequence of 585 bp was found to contain the characteristic bZIP domain and showed significant similarity to the known O2 proteins. The pairwise alignment of this putative finger millet O2 mRNA fragment with full-length O2 mRNA of maize revealed that the partial fragment lacked both 5' and 3' end. Therefore, this fragment was considered to design outer and nested primers for performing 3' and 5' RACE to obtain the full-length sequence. The first round of 5' RACE PCR using the outer primers produced some unclear bands with smear in the background. However, intense amplicons were produced in the first and second nested PCRs with the expected reduction in band size according to the distance between the nested primers (Supplementary Fig. 2). Similarly, the first round of 3' RACE PCR using the outer primers produced smeared products which subsequently produced clear and intense bands using the first and second nested primers with expected reduction in band sizes (Supplementary Fig. 3). The nested 3' and 5' RACE PCR products were subsequently eluted, cloned, and sequenced. 548 and 632 nucleotides were recovered through 5' and 3' RACE, respectively. The assembled sequenced

showed that the putative finger millet *Opaque2* mRNA contains an ORF of 1248 nt encoding 416 amino acids containing the characteristic BRLZ and bZIP-C domain which is present in all opaque2 proteins. The ORF was flanked by 181- and 130-nt 5' and 3' untranslated region (UTR), respectively (Fig. 1). The full-length finger millet O2 (*EcO2*) mRNA was submitted to the GenBank database and is available with the accession no. KX440959.

Identification of finger millet transcripts encoding bZIP domain proteins and in silico characterization of the isolated full-length finger millet O2 mRNA

Using the local standalone BLAST programme, a total of 155 and 126 transcripts from the developing seed transcriptome data of high calcium (GP-45) and low calcium (GP-1) genotypes, respectively, were identified encoding the characteristic bZIP domain. Phylogenetic analysis of all the finger millet proteins containing bZIP domains, O2 of other cereals, and the full-length O2 protein isolated through 3' and 5' RACE showed that O2 of all the cereals and the isolated finger millet O2 protein sequence clustered into a distinct group, while all other proteins were grouped separately. Corresponding protein sequence homologs of the isolated full-length *EcO2* mRNA (identified in the genotype GE-3885) were also identified in the genotypes GP-1 and GP-45 and were 100% identical to each other (Fig. 2). Finger millet O2 protein showed highest similarity to the barley BLZ1 (75% similarity), followed by rice RISBZ1 (57% similarity), wheat SPA-A (49%), SPA-B (50.3%), SPA-D (48.9%), and maize ZmO2 (51%).

Protein sequences of different cereal O2 mRNAs along with the isolated finger millet O2 mRNA were further analyzed to find conserved motifs in the coding region using the online local alignment tool MEME ver. 4.11.1 (Bailey et al. 2006). The distribution of the conserved motifs in different O2 proteins is shown in Fig. 3. As depicted in the figure, the composition and spatial distribution of motifs in the *EcO2* was similar to the barley

Fig. 1 Pictorial representation of isolated full-length finger millet *OPAQUE2* mRNA

Fig. 2 Phylogenetic relationship of the isolated full-length finger millet OPAQUE 2 protein with known OPAQUE 2 proteins of other cereals and other finger millet bZIP proteins expressed during seed development stages

Fig. 3 Identification of conserved motifs among finger millet OPAQUE 2 protein (EcO2) along with the O2 proteins of other cereals

BLZ1 except that one motif (motif no. 13) present in EcO2 was absent in BLZ1, while one motif (motif 16) present in BLZ1 was absent in EcO2.

Analysis of full-length O2 proteins for putative phosphorylation sites using Net Phos 2.0 revealed the presence of 30 phosphorylation sites in the isolated finger millet O2 protein, while 32, 31, 29, 23, 31, 27, and 30 phosphorylation sites were found to be present in SPA-A, SPA-D, SPA-B, RISBZ1, BLZ1, BLZ2, and ZmO2, respectively (Supplementary Table 3).

Detection of *EcO2* transcripts in different tissues

The *EcO2* transcripts were detected in roots, stem, leaves, and in all the stages of developing finger millet seeds/spikes. However, relative to the S1 stage of seed development highest induction of *EcO2* was observed in the S2 stage, indicating that *EcO2* dominantly expresses during the seed development (Fig. 4).

Fig. 4 Tissue-specific expression profile of the *OPAQUE2* gene in vegetative tissue (root, stem, and leaf) and developing spikes (S1, S2, S3, and S4) of the finger millet genotype (GE-3885). Total RNA was isolated from roots, stem, leaves, and during all the stages of developing finger millet seeds, and was converted to cDNA. *EcO2* gene-specific primers were used to detect the expression in all the tissues. Fold change in expression was calculated relative to the S1 stage of seed development

Expression profiling of finger millet *EcO2* and α -prolamin gene in response to different nitrogen inputs in two genotypes differing in seed protein contents

In the control experiment where no nitrogen was added, the expression of *EcO2* in both the genotypes (GE1437 and GE3885) was low and did not change during subsequent seed developmental stages. However, in case of

high protein genotype GE3885, slightly higher expression was detected at the S₂ stage. As the nitrogen dose was increased the expression of *EcO2* was differential. In both the genotypes, *EcO2* transcripts were induced from S₁ to S₂ stage with increase in nitrogen dose before steeply declining in the subsequent (S₃ and S₄) stages. Compared to GE1437, at all the nitrogen doses, *EcO2* was more strongly induced in GE3885 at the S₂ stage of seed development and highest expression of (28-fold relative

Fig. 5 Gene expression profile of *O2* gene at different stages of finger millet genotypes (GE-1437 and GE-3885) grown under different nitrogen inputs. The *O2* gene expression patterns were studied using real-time PCR as described in the materials and methods. Fold change in expression was calculated relative to the S1 stage of seed development. The expression data were statistically analyzed using

the statistical software 'Statistica'. The figure shown is the graphical representation of the interactions between the independent variable used in the present study, i.e., genotype $\times$ treatment $\times$ stages; LS Means, Wilks lambda = 0.19090, $F(18, 126) = 9.0213$, $p = 0.00000$. Control-(without N), low-(20 kg/ha), normal-(40 kg/ha), and high-60 kg/ha)

to the S1 stage) was detected at high nitrogen dose treatment (Fig. 5).

In both the genotypes, the prolamin gene expression continuously increased and reached at maximum levels at the S₄ stage of seed development. However, compared to GE1437 in all the nitrogen treatments, the prolamin gene was more strongly induced in the genotype GE3885. In the genotype GE3885 grown under control conditions, prolamin transcripts increased to almost 100-fold in the S₄ stage (relative to the S1 stage), while 350-, 400-, and 470-fold increase was detected in the plants grown under low, normal, and high nitrogen doses, respectively. In the genotype GE1437 grown under control conditions, no apparent change in expression of prolamin gene expression was detected up to the S₃ stage. However, at the S₄ stage, prolamin gene was induced to almost 50-fold. A similar delayed induction of prolamin gene was observed GE1437 plants grown under low, normal, and high nitrogen conditions (Fig. 6). Highest induction of prolamin gene in GE1437 of about 100-fold (relative to S1 stage) was detected at the S₄ stage of seed development when the plants were grown under high nitrogen dose.

Discussion

The O2 proteins belonging to the bZIP family of transcription factors regulate the cereal seed storage protein genes during seed development. Since the identification of first O2 gene in maize, O2 homologs in other cereals such as rice, barley, and wheat have been identified and characterized. Although finger millet is regarded as crop with high NUE and is emerging as a nutritionally important food crop, the O2 homolog of finger millet has not yet been identified and characterized. Therefore, the present investigation was undertaken to identify and characterize the O2 mRNA of finger millet that would aid in our understanding the regulation of seed storage protein accumulation during seed development under nitrogen-limiting conditions.

Isolation of full-length finger millet OPAQUE 2 transcript

Since OPAQUE 2 regulates seed storage protein genes during seed development, the high seed protein genotype GE3885 that might have high O2 expression levels during

Fig. 6 Gene expression profile of *PROLAMIN* gene at different stages of finger millet genotypes (GE-1437 and GE-3885) grown under different nitrogen inputs. The *PROLAMIN* gene expression patterns were studied using real-time PCR as described in the materials and methods. Fold change in expression was calculated relative to the S1 stage of seed development. The expression data were statistically

analyzed using the statistical software 'Statistica'. The figure shown is the graphical representation of the interactions between the independent variable used in the present study, i.e., genotype $\times$ treatment $\times$ stages; with LS Means, Wilks lambda = 0.19090, $F(18, 126) = 9.0213$, $p = 0.00000$. Control-(without N), low-(20 kg/ha), normal-(40 kg/ha), high-(60 kg/ha)

the seed development was chosen to isolate the full-length finger millet *O2* transcript. A conserved primer approach was used to first isolate the partial sequence of finger millet *O2* and then use the 3' and 5' RACE technique to isolate the remaining part of the mRNA. The conserved primers amplified a partial region of 585 nucleotides, while the 3' and 5' RACE amplified 632 and 548 nucleotides, respectively. Assembly of the fragments resulted in the identification of an ORF of 1248 nucleotides encoding 416 amino acid protein containing the characteristic “BRLZ” and “bZIP-C” domain found in all the *O2* like proteins characterized in other cereals (Hurst 1995; Nakase et al. 1997).

Finger millet OPAQUE 2 protein shows similarity to the barley BLZ1

To further characterize the isolated *O2* transcript, a phylogenetic analysis was carried out using *O2* of rice, maize, wheat, and barley and all the transcripts encoding bZIP proteins expressed during seed development stages identified using the developing seed transcriptome data of two finger millet genotypes. Since all *O2* proteins share common functions, as expected, all the known *O2* proteins along with the isolated finger millet *O2* protein clustered together indicating sequence and functional conservedness among themselves. However, finger millet *O2* protein showed a higher overall sequence similarity to barley BLZ1 (~ 68% identical amino acid residues; (Supplementary Table 3) than to maize *O2* and wheat SPA (~ 36–39% identity). The motif analysis carried out using the online tool MEME also showed that the isolated finger millet *O2* is closer to the barley *BLZ1* gene than other *O2* genes. The phylogenetic analysis also identified the corresponding *O2* transcripts in the two genotypes (GP45 and GP1) which were used for generating the developing seed transcriptome data (Fig. 2). Both the proteins were 100% identical to the *O2* of the genotype GE3885 identified through 3' and 5' RACE. Another finger millet transcript was also found clustered with the *O2* homologs. Protein BLAST of this sequence revealed that it is 86% similar to the maize *ohp1* (OPAQUE 2 heterodimerization protein1, GenBank accession no. DAA519421).

Identification of phosphorylation sites in *O2* proteins revealed 30 phosphorylation sites in the finger millet *O2* protein which in contrast to rice RISBZ1 is almost similar to that found in SPA-A, SPA-D, SPA-B, BLZ1, BLZ2, and ZmO2. It has been reported earlier that *O2* proteins are post-translationally regulated by phosphorylation. The hyper phosphorylated form has low activity which leads to low-level expression of alpha prolamin (Ciceri et al. 1997). Hence, further experiments need to be done to determine the phosphorylation–dephosphorylation cascade of finger millet *O2* protein that might be crucial to understand the activity of *EcO2* under the influence of varying N-input.

Finger millet OPAQUE2 gene shows nitrogen responsiveness

In cereals, regulation of seed storage protein predominantly starts from transcriptional level and the *O2* and PBF transcription factor plays an important role in activating the expression of seed storage protein genes (Hartings et al. 1989; Vicente-Carbajosa et al. 1997; Marzábal et al. 2008). However, during the seed development, the *O2* genes are induced earlier than the *PBF* and the seed storage protein genes, indicating that it might be a critical activator in controlling seed storage protein genes (Yamamoto et al. 2006). In the present investigation, *O2* expression was also found to be induced early during finger millet seed development as it was sharply induced in the *S*₂ stage but subsequently declined in *S*₃ to *S*₄ stages. However, the expression of the *Ecα-prolamin* gene continuously increased from *S*₁ to *S*₄ stage and reached maximum expression levels in the *S*₄ stage. This prolamin gene was included in the study, because its promoter sequence was found to contain *O2*-binding sites (GCN4 box) (unpublished data). Interestingly, the expression of both *O2* and *Ecα-prolamin* was found to be higher in high protein genotype (GE-3885) than the low protein genotype (GE-1437) for each treatment of nitrogen, indicating that the expression of *EcO2* is regulated according to nitrogen conditions or nitrogen status of plant. Probably, the promoter of *EcO2* might have unique regulatory elements which are the targets of regulatory proteins that sense plant nitrogen status or nitrogen nutrient availability during the vegetative growth period. Identification and characterization promoter region of *EcO2* in both genotypes might shed more light about the nitrogen responsive regulation of *EcO2* gene. Our results are also supported by the previous study on maize *α-zein* by in vitro culture of maize endosperm in which expression of *α-zein* increased seven fold as compared to control when both organic, i.e., amino acids asparagine and inorganic source ammonium nitrate (NH₄NO₃) were supplied in vitro culture of maize endosperm (Balconi et al. 1993). Furthermore, mutational study at *O2*-binding site of prolamin promoter at GCN₄ site has also shown that increased amount of nitrogen has no effect when *α-zein* promoter is mutated at *O2* binding site. While in wild type of construct as nitrogen increases, the expression of alpha zein increases (Balconi et al. 1993; Muller et al. 1993). This indicates that *O2* transcription factor is under both transcriptional and post-transcriptional regulatory control and warrants further experiments to decipher this dual regulation that might bolster our understanding how finger millet accumulates high amounts of seed storage proteins under nitrogen-limiting conditions.

Conclusions

In the present study, induction of prolamin and *O2* genes under increased nitrogen inputs indicates that although finger millet crop is grown under marginal soils with almost no nitrogen inputs and considered as low nitrogen requiring plant, but the finger millet crop is nitrogen responsive and increasing nitrogen fertigation may improve yield. Further studies in this direction using field applications of nitrogen in selected genotypes might provide clues for better understanding the differential responsiveness of nitrogen in finger millet genotypes. Since there exists a wide variation in seed protein content in finger millet germplasm, allelic variations in the *OPAQUE 2* gene might be studied to identify possible DNA markers for use in finger millet crop improvement programme. Furthermore, site-directed mutagenesis of *O2* protein and understanding the protein–protein interaction of PBF and *O2* on *Ec*α-prolamin promoter might provide unique insights about the nutritional regulation of seed storage protein genes under low nitrogen conditions.

Acknowledgements This work was supported by Science and Engineering Research Board (SERB), DST, New Delhi, India, in the form of “Fast Track Scheme for Young Scientist” awarded to VSG and SG (SR/FT/LS-33/2011 and YSS/2015/00536). The logistic support provided by the Director, Experiment Station, G.B. Pant University of Agriculture and Technology, Pantnagar is also thankfully acknowledged.

Author contributions AK conceived the idea and conceptualized and designed the manuscript. VSG and LK are involved in the lab work of the presented R & D data. VSG, LK, and SG wrote the manuscript. JPI, DP, and AK contributed in critically revising the draft and updating the manuscript for publication.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Albani D, Hammond-Kosack MC, Smith C, Conlan S, Colot V, Holdsworth M, Bevan MW (1997) The wheat transcriptional activator SPA a seed-specific bZIP protein that recognizes the GCN4 like motif in the bifactorial endosperm box of prolamin genes. *Plant Cell* 9:171–184
- Bailey TL, Williams N, Misleh C, Li WW (2006) MEME discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34:W369–W373
- Balconi C (1993) The accumulation of zein polypeptide and zein mRNA in cultured endosperms of maize is modulated by nitrogen supply. *Plant J* 3:325–334
- Blom N (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294:1351–1362
- Ciceri P, Gianazza E, Lazzari B, Lippoli G, Genga A, Hoschek G, Schmidt RJ, Viotti A (1997) Phosphorylation of Opaque-2 changes diurnally and impacts its DNA binding activity. *Plant Cell* 9:97–108
- Fakrudin B, Shashidhar HE, Kulkarni RS, Hittalmani S (2004) Genetic diversity assessment of finger millet (*Eleusine coracana*, Gaertn) germplasm through RAPD analysis. *PGR Newslett* 138:50–54
- Gowda J, Haider ZA, Gupta A, Venkateshwara T, Mahadevaiah C, Somu G, Seetharam A (2005) Finger millet [*Eleusine coracana*(L.) Gaertn] Core germplasm for utilization in crop improvement. Published by the All India Coordinated Small Millet Improvement Project ICAR University of Agricultural Sciences Bangalore India, 171
- Gupta N, Gupta AK, Singh NK, Kumar A (2011) Differential expression of PBF Dof transcription factor in different tissues of three finger millet genotypes differing in seed protein content and color. *J Plant Mol Biol Rep* 29:69–76
- Hammond Kosack MC, Holdsworth MJ, Bevan MW (1993) *In vivo* footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *EMBO J* 12:545–554
- Hartings H, Maddaloni M, Lazzaroni N, Fonzo ND, Motto M, Salamini F, Thompson R (1989) The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *EMBO J* 8:2795–2801
- Hurst HC (1995) Transcription factors 1 bZIP proteins. *Protein Profile* 2:101–168
- Kreis M, Forde BG, Rahman S, Mifflin BJ, Shewry PR (1985) Molecular evolution of the seed storage proteins of barley, rye and wheat. *J Mol Biol* 183:499–502
- Kumar R, Taware R, Gaur VS, Guru SK, Kumar A (2009) Influence of nitrogen on the expression of TaDof1 transcription factor in wheat and its relationship with photo synthetic and ammonium assimilating efficiency. *Mol Biol Rep*. <https://doi.org/10.1007/s11033-008-9436-8>
- Kumar A, Gaur VS, Goel A, Gupta AK (2014) *De Novo* Assembly and characterization of developing spikes transcriptome of finger millet (*Eleusine coracana*): a minor crop Having Nutraceutical Properties. *Plant Mol Biol Rep* 1:18
- Kumar A, Metwal M, Kaur S, Gupta AK, Puranik S, Singh S, Singh M, Gupta S, Babu BK, Sood S, Yadav R (2016) Nutraceutical value of Finger Millet [*Eleusine coracana* (L.) Gaertn] and their improvement using omics approaches. *Front Plant Sci* 7:934
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) method. *Methods* 25:402–408
- Marzábal P, Gas E, Fontanet P, Vicente-Carbajosa J, Torrent M, Ludevid MD (2008) The maize Dof protein PBF activates transcription of gamma-zein during maize seed development. *Plant Mol Biol* 67:41–54
- Mena M, Vicente CJ, Schmidt RJ, Carbonero P (1998) An endosperm specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin box of a native B-hordein promoter in barley endosperm. *Plant J* 16:53–62
- Mirza N, Sharma N, Srivastava S, Kumar A (2015) Variation in popping quality related to physical, biochemical and nutritional properties of finger millet genotypes. In: Proceedings of the National Academy of Sciences, India section B: biological sciences, vol 85, no 2. pp 507–515
- Muller M, Knudsen S (1993) The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. *Plant J* 4:343–355
- Murphy DJ (2000) Shewry PR, Casey R, eds. 1999. *Seed proteins*. 883 pp. Dordrecht: Kluwer. £315 (hardback). *Ann Bot* 86(2):434–435. <https://doi.org/10.1006/anbo.2000.1212>
- Nakase M, Aoki N, Matsuda T, Adachi T (1997) Characterization of a novel rice bZIP protein which binds to the alpha-globulin promoter. *Plant Mol Biol* 33:513–522

- Onate L, Vicente CJ, Lara P, Diaz I, Carbonero P (1999) Barley BLZ2 a seed-specific bZIP protein that interacts with BLZ1 in vivo and activates transcription from the GCN4 like motif of B-hordein promoters in barley endosperm. *J Biol Chem* 274:9175–9182
- Onodera Y, Suzuki A, Wu CY, Washida H, Takaiwa F (2001) A rice functional transcriptional activator RISBZ1 responsible for endosperm specific expression of storage protein genes through GCN4 motif. *J Biol Chem* 276:14139–14152
- Osborne TB (ed) (1924) *The Vegetable proteins* (monographs on biochemistry), 2nd edn. Longmans, Green and Co, London, p 154
- Pietro C et al (1999) The maize O2 and PBF proteins act additively to promote transcription from storage protein gene promoters in rice endosperm cells. *Plant Cell Physiol* 45:1509–1518
- Shewry PR, Tatham AS (1999) The characteristics, structures and evolutionary relationships of prolamins. In: Shewry PR, Casey R (eds) “Seed Proteins”, pp. 11–36. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4 molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Vicente CJ, Onate L, Lara P, Diaz I, Carbonero P (1998) Barley BLZ1 a bZIP transcriptional activator that interacts with endosperm specific gene promoters. *Plant J* 13:629–640
- Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ (1997) A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc Natl Acad Sci USA* 94:7685–7690
- Wu CY, Adachi T, Hatano T, Washida H, Suzuki A, Takaiwa F (1998) Promoters of rice seed storage protein genes direct endosperm-specific gene expression in transgenic rice. *Plant Cell Physiol* 39:885–889
- Yamamoto MP, Onodera Y, Touno SM, Takaiwa F (2006) Synergism between RPBF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol* 141:1694–1707