Comparative transcriptome profiling of high and low grain-iron containing Indian barnyard millet (Echinochloa frumentacea L.) genotypes during different stages of grain development

Shital M. Padhiyar
Junagadh Agricultural University
Jasminkumar Kheni
Junagadh Agricultural University
Shraddha B. Bhatt
Junagadh Agricultural University
Hiral Desai
Junagadh Agricultural University
Rukam S. Tomar (rukam@jau.in)
Junagadh Agricultural University

Article

Keywords:

Posted Date: March 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2624534/v1

Abstract

In the era of food nutritional security, the development of minerals-rich grains is an essence for fighting malnutrition. In the present study, we tried to identify the transcripts responsible for the higher accumulation of grain-Fe in Indian barnyard millet through transcriptome sequencing of genotype BAR-1433 (high Fe content) and BAR-1423 (low Fe content) during two stages of spike development i.e., spike emergence and milking stage. During the spike emergence stage, a set of 895 up-regulated and 126 down-regulated transcripts were identified between the high and low grain-Fe containing genotype, while during the milking stage, the number of up-regulated and down-regulated transcripts were 436 and 285. The transcripts which were commonly up-regulated during both the stages were having roles in nucleolar protein, metal-nicotianamine transporter, ribonucleoprotein complex, Vinorine synthase, Cellulose synthase, Auxin response factor, embryogenesis abundant protein, Cytochrome c oxidase, and Zinc finger BED domain-containing protein. Transcripts with significant differences in induction or repression between the two genotypes included genes related to ABC Transporter family proteins, Calcium dependent kinase family, Ferritin, Metal ion binding, Iron-sulfurculster binding, Cytochrome family, Zinc finger transcription factor family, Ferredoxin–NADP reductase type 1 family, Putative laccase, Multicopper oxidase family and Terpene synthase family. Six contigs representing their probable function for metal transporter, iron sulfur, metal ion binding, auxin-responsive GH3-like protein 2, and cytochrome P450 71B16 were used for designing primers to be used for validation. The result of qRT-PCR coincided with the result of the transcriptome. Thus, this study reports a repertoire of genes associated with high iron content in barnyard millet and a proof concept for deployment of transcriptome information for validation in mapping population and its use in marker-assisted selection for bio fortification of barnyard millet with iron. This is first report on a detailed transcriptome analysis to identify transcripts associated with high and low grain-iron content during panicle developmental stages in barnyard millet.

1. Introduction

Barnyard millet belongs to the family Poaceae, subfamily Panicoideae ,and tribe Paniceae comprise two different cultivated species, *Echinochloa utilis*, and *Echniochloa frumentacea*. *Echinochloa utilis* is also called Japanese barnyard millet, whereas *E. frumentacea* has several names such as sawa millet, billion-dollar grass, and also Indian barnyard millet. It is the fourth most produced minor millet, providing food security to many poor people across the world. This type of millet is considered a minor cereal and is grown widely in India, China, Japan, Pakistan, Africa, and Nepal[1]. During the last three years, India has stood as the largest producer of barnyard millet (0.147 mt), as well as the country with the largest area (0.146 m ha^{-1}) under it with average productivity of 1034 kg/ha[2].

Barnyard millet (*Echinochloa* species) is an ancient millet crop grown in warm and temperate regions of the world. It is a multi-purpose crop that is cultivated for food and fodder. It is one of the important minor millets which has a drought-tolerant capacity, fast-growing, and is cultivated over a wide array of environmental conditions and poor soils[3]. In addition to these agronomic advantages, the grains are valued for their high nutritional value and lower expense as compared to major cereals like rice, wheat,

and maize. It contains a rich source of protein (11.1%), carbohydrates (65%), fiber (9.8%)and most notably, micronutrients like iron (Fe) and zinc (Zn) that are related to numerous health benefits[4]. *E. Frumentacea* has the capability to reduce down the blood glucose in comparison to any other minor millets[5]. These extraordinary features of barnyard millet make it an ideal crop for subsistence farmers and also as a replacement crop during the failure of the major crop during the *kharif* season.

In recent years, micronutrient malnutrition and deficiency are one of the foremost issues globally including India, and have peaked in recent time. Anaemia caused by iron deficiency is a significant global public health issue. Worldwide, 42% of pregnant women, 30% of non-pregnant women (aged 15 to 50), 47% of preschool children (under 5 years), and 12.7% of young males (> 15 years) are anaemic, according to a World Health Organization (WHO) report[6]. Children's growth and cognitive development, non-pregnant women's cognitive, physical, and psychological health, and pregnant women's maternal and neonatal outcomes are all negatively impacted by iron deficiency anaemia. In most Asian and African nations, its prevalence among women between the ages of 15 and 49 is greater than 40%. Milled grains of rice, wheat, and maize have supplanted the traditional, nutrient-dense crops in developing nations. Starch is abundant in refined diets, but minerals, particularly micronutrients like iron and zinc, are lacking. Given that low-iron staple foods make up the majority (> 80%) of the diet in developing nations, it is impractical to consume enough iron through the remaining 20% of the diet. Therefore, it is crucial to diversify the staple food by including food crops that are naturally high in iron content, like millets. Additionally, as compared to refined rice and refined wheat, millets offer 2.3 to 4.0 times more dietary fibre (6.4 to 11. 5 g/100g), which serves as food for the gut flora, improving its abundance and changes the makeup of the gut in a positive way[7]. Indian barnyard millet is rich in Fe along with other micronutrients. A cup of 100 grams of barnyard millet has the potential to provide 100% of the daily value of iron and 67% of the daily value during pregnancy along with a significant amount of calcium. Several studies disclosed the nutritional profile of barnyard millet, particularly the high Fe and Zn content in the grains. In spite of this information, negligible research work has been carried out to tap the enormous potential of this crop. This is largely because of the growing of staple crops like rice, wheat and maize post-green revolution to increase food production. In addition to this, very limited genomic information on barnyard millet has further restricted the involvement of modern breeding and biotechnological techniques in crop improvement programs. Expressed Sequence Tag (EST) has a proven record to accelerate the research on many crops[8] and it has been hardly carried out in barnyard millet. Therefore, to accelerate the research activities of this very significant crop, it is essential to increase its genomic and transcriptomic information as early as possible. Henceforth, the identification of potential genes associated with the accretion of iron and the transfer of these identified genes to high-yielding barnyard millet cultivars or even to other major staple crops like wheat, rice, and maize need an in-depth study of these genes and factors affecting them.

In recent years, advances in next-generation sequencing (NGS) technologies have delivered exceptional results for creating genomic resources and unveiling important molecular mechanisms controlling specific biological processes. It has also paved the way for large-scale sequencing and has demonstrated

to be a valuable tool having huge potential applications in plant biology including transcriptome investigations and genome sequencing[9]. RNA-Sequencing is a whole transcriptome sequencing method that can measure gene expression at the transcriptional level thus providing immense information about non-coding regions, determining the structure of transcripts and identification of differentially expressed genes that can be applied to define alleles associated with important agronomical traits[10]. It has the potential to identify important secondary metabolic pathways, various transcripts associated with diseases and is helpful for discovering novel genes[11].

A biofortification breeding program to increase the Fe content in the barnyard millet genotypes has the ability to eradicate anaemia and to provide the daily requirement of iron to the human. The biofortification breeding program can be accelerated if advanced biotechnological tools are incorporated into conventional breeding. This needs a complete study of genes involved in Fe accumulation and its pathways. In this study, we report the transcriptome of Indian barnyard millet having high iron content during two different stages of spike development. The samples from high and low grain-iron-containing genotypes were collected to compare them for the identification of differentially expressed genes involved in iron accumulation in the grains.

2. Materials And Methods

2.1 Plant genotypes and Fe concentration

Genotypes of Indian barnyard millet (*Echinochloa frumentacea* L.) were obtained from the ICAR-Indian Institute of Millets Research (ICAR-IIMR), Rajendra Nagar (Hyderabad, Telangana, India). All the permissions for carrying out this research were taken and the complete work was carried out according to the guidelines laid by the institute for working with the plant. The collection of seeds and the complete experiment was carried out according to national and institutional guidelines and also complies with international guidelines. The seeds of all 30 genotypes were sown in the multiplication plot available at the Department of Biotechnology during the *Kharif* season, 2020 (second fortnight of June) in three replications. The mature seeds were harvested individually and were analysed for Fe content in 3 replications using wet oxidation method and digesting it by diacid-HNO_3:$HClO_4$ in a ratio of 3:1. The Fe content was measured through microwave plasma atomic emission spectroscopy (MP-AES, Agilent technologies) using Emission wavelength of 259.94 nm, viewing position of zero and nebulizer gas flow rate of 0.6 L/min. Out of the thirty genotypes, the genotype BAR-1433 was having high Fe content and genotype BAR-1423 was having low Fe content and was taken as control. Genotypes BAR-1433 and BAR-1423 were sown in field conditions followed by recommended agricultural practices which include the application of NPK in ratio of 40:20:0 with the splitting of nitrogen in two doses of 20 kg each. The samples were collected at two stages for each genotype i.e., spike emergence, and milking stage for transcriptomic analysis after 35 days after sowing (DAS)and 55 DAS respectively in three replications. All samples were frozen immediately in liquid nitrogen and stored at -80°C until RNA extraction.

2.2 RNA extraction and high-throughput sequencing

Total RNA was extracted from the 12 samples (3 replication from two genotypes at two stages) using TRIzol Reagent (Invitrogen, Carlsbad, California, United States) and then followed RNeasy Plant Mini Kit (*QiaGen*, Valencia, CA) as per the manufacturer's protocol. The quality of RNA was confirmed on 1.0% agarose gel electrophoresis and concentration was measured using a Qubit® RNA HS Assay Kit (Invitrogen™) using Qubit® 2.0 Fluorometer™ (Invitrogen™). Approximately 1µg of total RNA was used for the isolation of mRNA following the manufacturer protocol of Dynabeads® mRNA DIRECT™ Kit (Thermofisher Scientific, USA). After purification mRNA was fragmented by RNase enzyme at 37°C in RNAase buffer provided in cDNA library preparation kit Total RNA-Seq Kit v2 (Thermofisher Scientific, USA). The fragments were reverse transcribed using random hexamers and superscript II reverse transcriptase (Invitrogen™). cDNA library was amplified using PCR for the enrichment of the adapter-ligated fragments. Each sample was molecularly barcoded during the library preparation to differentiate from each other during downstream analysis. The individual libraries were measured using a Qubit 2.0 Fluorometer and validated for quality in E-Gel 2% Agarose (Invitrogen™). Subsequently, these libraries of each sample were diluted up to 100pM concentration and subjected to emulsion PCR (Ion OneTouch™ 2 system, Thermofisher Scientific, USA), and then these enriched templates were sequenced using ION S5™ system (Thermofisher Scientific, USA) next-generation sequencer to generate raw data.

2.3 Pre-processing of RNA-Seq data and de novo assembly

Initially, low-quality reads were removed from all the individual sequence data files of each sample. The reads having > 50% bases with low-quality scores and/or > 10% bases unknown (N bases) were removed from each raw data for accuracy of results using CLC genomics workbench 20.0 (CLC GWB)[12] and Prinseq quality control tools (http://prinseq.sourceforge.net/). Raw reads were again processed for removing adapters and low-quality sequences (< Q30) using CLC GWB 20.2 parameters[13]. RNA-Seq read quality before and after trimming was assessed using FastQC[14].

De novo transcriptome assembly was created using Trinity (v. 2.11.0) with default settings[15,16] and CAP3[17] was employed in the redundancy reduction of the assembly. Further, CD-HIT program (v. 4.8.1) with default parameters (similarity 95%) was again used to reduce transcript redundancy and produce unique genes ("unigenes")[18]. Coding regions of the assembled transcripts were predicted using TransDecoder v. 5.5.0 (http://transdecoder.github.io).

2.4 *De novo* transcriptome profiling

The replicated reads were aligned individually and then combined together to make 4 samples which were further used for different analysis. The level of transcripts expression was analysed on the basis of number of reads mapping to each transcript. CLC RNA-Seq analysis was used for aligning the reads of each sample onto the assembled transcripts[19]. Assembler provides the annotated transcripts and their annotated length, Coverage, RPKM, and transcripts per kilobase million (TPM) values for each sample. The resulting files of aligned reads were input in differential expression in two groups, a tool for quantifying the abundances of a set of target sequences from sampled subsequence based on a model (GLM model) using the negative binomial distribution[20]. Gene expression levels were estimated by RPKM

and FDR P values. Genes with RPKM fold changes > 2 or <-2, and FDR-corrected p values < 0.05 were regarded as Differentially Expressed Genes (DEGs).

2.5 Functional annotation

Basic Local Alignment Search Tool (BLAST) and BLAST2GO were used for assigning functional annotations to the unigenes[21,22] with an e-value of 1E-6. For Pathway analysis, DEGs were annotated[23]. GO classification and KEGG pathways enriched for DEGs were performed for the up- and down-regulated genes of the significant difference in gene proportion between the two genotypes. Finally, the Pathway enrichment analysis of DEGs was carried out by feeding gene id into Shiny GO[24]. The detailed workflow of barnyard millet transcriptome analysis has been depicted in Supplementary Fig. 1.

2.6 Validation of DEGs through qRT-PCR analysis

To confirm the transcripts involved in Fe accumulation in the grain during spike developmental stages, stored RNA samples were utilised and cDNA was synthesized from an aliquot of total RNA using QuantiTect Reverse Transcription Kit (QIAGEN, USA) and served as the template for qRT-PCR (Quantitative Real-Time Polymerase Chain Reaction). Based on the gene ontology terms (GO terms) given to the contigs showing differential gene expression, six of these contigs were selected that were either involved in metal/ion transport or were directly involved in Fe function. Primer 3 software was used for primer designing from six contigs involved in the accumulation of Fe[25]. The qRT-PCR was performed using QuantiFast SYBR Green PCR Master Mix (QIAGEN, USA) on ABI-7300 Real-Time PCR detection system, (Applied Biosystem) using standard 40 cycles along with melt curve step (average annealing temp 56^0C). The elongation factor (EF1) transcript was used as an endogenous reference for normalization. To obtain a linear relationship, PCR conditions were optimized for each set of genes. Finally, differential gene expressions were computed in terms of $\Delta\Delta^{CT}$ fold change value[26].

3. Results

3.1 Concertation of Fe in barnyard millet

The seeds obtained from ICAR-IIMR were sown at the Department of Biotechnology, JAU in three replications to multiply the seeds and to analyse them for Fe content. The Fe concentration in the seeds of 30 barnyard millet genotypes was measured using Microwave Plasma Atomic Emission Spectroscopy (MP-AES) using diacid-HNO_3:$HClO_4$ digestion method in three replications. Significant genetic variation in Fe concentration was found among the set of 30 genotypes with the range varying from 3.83 mg/100g to 13.14 mg/100g. The maximum Fe content was found in genotype BAR-1433 while the minimum was found in genotype BAR-1423 (Table 1). These two genotypes with contrasting Fe content were used for transcriptome sequencing to find out the differentially expresses genes in terms of Fe content.

Table 1
Sample details and concentration of Fe in the grains of barnyard millet genotypes

Sr No	Genotypes	Fe (mg/100g)	Sr No	Genotypes	Fe (mg/100g)	Sr No	Genotypes	Fe (mg/100g)
1	BAR-1406	5.52 ± 0.13	11	BAR-1416	5.46 ± 0.08	21	BAR-1426	6.76 ± 0.05
2	BAR-1407	6.82 ± 0.09	12	BAR-1417	5.4 ± 0.05	22	BAR-1427	4.22 ± 0.04
3	BAR-1408	7.69 ± 0.13	13	BAR-1418	5.38 ± 0.09	23	BAR-1428	5.48 ± 0.12
4	BAR-1409	5.79 ± 0.11	14	BAR-1419	4.98 ± 0.04	24	BAR-1429	10.18 ± 0.17
5	BAR-1410	5.07 ± 0.06	15	BAR-1420	5.98 ± 0.04	25	BAR-1430	5.97 ± 0.04
6	BAR-1411	6.14 ± 0.11	16	BAR-1421	8.09 ± 0.09	26	BAR-1432	4.94 ± 0.08
7	BAR-1412	6.54 ± 0.06	17	BAR-1422	4.34 ± 0.06	27	BAR-1433	13.14 ± 0.09
8	BAR-1413	6.49 ± 0.07	18	BAR-1423	3.83 ± 0.02	28	BAR-1434	8.92 ± 0.04
9	BAR-1414	5.53 ± 0.03	19	BAR-1424	4.8 ± 0.13	29	BAR-1329	5.74 ± 0.08
10	BAR-1415	5.8 ± 0.06	20	BAR-1425	4.61 ± 0.06	30	BAR-1489	5.9 ± 0.12
	Mean	6.18						
	Minimum	3.83						
	Maximum	13.14						
	S.Em.	0.050						
	C.D. at 5%	0.142						
	C.V. %	1.405						

3.2 RNA sequencing and *denovo* assembly

Transcriptome sequencing of both the genotypes (BAR-1433 and BAR-1423) was carried out in three replications during the two developmental stages of panicle. The samples were collected during spike emergence and milking stage from both the genotypes. An average of 25.6 million raw reads were generated in high Fe-containing genotype while 43.4 million reads were generated in the low Fe-containing genotype (Table 2). The raw data were subjected to adaptor trimming and bases quality check, it was found that more than 87% reads of the total reads were of high quality and were subjected to mapping to the barnyard millet *de novo* assembly. CLC, SOAP *denovo* trans and Trinity assembler were used to perform *denovo* assembly. The total number of transcripts including singletons was 225035,

177466 and 488689 in CLC, SOAP *denovo* trans and trinity assembler, respectively. The detailed description of the assemblers employed and the output is briefed in Table 3.

Table 2
Reads, de-novo assembly and mapping statistics of different stage of High and Low-Fe containing genotypes of barnyard millet

Raw reads Name	**Total Number of raw reads**	**Total number of high quality reads after quality control &** **Percentage trimmed (%)**
HFe_Spike _R1	49,58,173	44,90,055 (90.56%)
HFe_Spike _R2	60,36,437	49,62,450 (82.21%)
HFe_Spike _R3	20,12,050	16,64,695 (82.74%)
HFe_milking_R1	40,81,195	36,54,762 (89.55%)
HFe_milking_R2	58,74,780	51,69,355 (87.99%)
HFe_milking_R3	27,30,263	24,16,438 (88.51%)
LFe_Spike_R1	47,77,732	42,24,165 (88.41%)
LFe_Spike _R2	48,44,317	40,49,722 (83.6%)
LFe_Spike _R3	22,42,520	17,68,627 (78.87%)
LFe_milking_R1	70,91,421	65,43,373 (92.27%)
LFe_milking_R2	1,27,79,824	1,15,87,914 (90.67%)
LFe_milking_R3	1,16,80,499	1,08,24,581 (92.67%)

Table 3
De Novo assembly statistics of master assembly

	BM_CLC	BM_SOAP denovo_Trans	BM_TRINITY	BM_CAP3	BM_Unigenes
# contigs (> = 0 bp)	225035	177466	488689	27228	20849
# contigs (> = 1000 bp)	10288	180	94963	12544	9474
# contigs (> = 5000 bp)	14	0	652	127	111
# contigs (> = 10000 bp)	0	0	16	5	5
Total length (> = 0 bp)	99530747	31851533	340205016	31871464	24424427
Total length (> = 1000 bp)	14804104	211325	158717873	22881994	17550212
Total length (> = 5000 bp)	80262	0	3964192	775697	684999
Total length (> = 10000 bp)	0	0	177893	56199	56199
# contigs	57231	3932	237463	22277	16848
Largest contig	9642	1933	12255	12254	12254
Total length	45978422	2588014	259145907	29978933	22897258
GC (%)	46.05	46.33	46.21	46.75	46.76
N50	786	636	1202	1570	1604
N90	539	521	611	723	719
L50	19619	1614	68319	6273	4669
L90	48367	3426	190569	17344	13065
# N's per 100 kbp	0.00	0.00	0.00	0.00	0.00

3.3 Differential expression of transcripts

The replicated transcripts were combined together to identify differentially expressed transcripts. While comparing the sample, at spike emergence stage between the High Fe (HFe) and Low Fe containing genotype a set of 895 up-regulated and 126 down-regulated transcripts were identified (Supplementary Fig. 2 and Supplementary Table S1). At the milking stage, the number of up-regulated and down-regulated transcripts were 436 and 285, respectively (Supplementary Fig. 3 and Supplementary Table S2).When the transcripts of both the stage i.e. spike emergence and milking were combined and

compared, a set of 957 transcripts were either up or down-regulated during the spike emergence stage and 657 transcripts were either up or down-regulated during the milking stage, however, 64 transcripts were commonly either up or down-regulated (Fig. 1A) in HFe. The total number of transcripts that were up-regulated during the spike emergence and milking stages were 895 and 436, respectively of which 27 were commonly up-regulated (Fig. 1B). Among the down-regulated transcripts, 22 transcripts were commonly down-regulated during both the stages, while the uniquely down-regulated transcripts during spike emergence were 104 and during milking stage were 263 (Fig. 1C).

3.4 Functional annotation of differentially expressed transcripts

Annotation of 64 differentially expressed transcripts (commonly expressed during both the stage), was performed by aligning the transcripts with NR and Uniport database. The transcripts which were commonly up-regulated during both the stages were having role in nucleolar protein, metal-nicotianamine transporter, ribonucleoprotein complex, Vinorine synthase, Cellulose synthase, Auxin response factor, embryogenesis abundant protein, Cytochrome c oxidase and Zinc finger BED domain-containing protein. The transcripts which were down-regulated during both the stages were having role in Chitinase, bZIP transcription factor, Light-independent protochlorophyllide reductase, Nitrate reductase, Phenylalanine ammonia-lyase, Lysine-specific histone demethylase and Bifunctional pectinesterase (Supplementary Table S3).

Furthermore, Go-based classification was implemented for the up- and down-regulated transcripts in both the spike emergence and milking stage. Out of the total 895 up-regulated transcripts during the spike emergence, 811 transcripts were significantly enriched with the GO terms. The Biological Process (BP) GO terms that were associated with up-regulated transcripts during the spike emergence stage were biological regulation, multicellular organismal process, developmental process, immune system process, response to stimulation, cellular process, metabolic process, locomotion, biological adhesion, detoxification, etc. The molecular function (MF) GO terms that were significantly enriched were binding, transporter activity, transcription regulator activity, molecular carrier activity, nutrient reservoir activity, etc. The Cellular components (CC) were cell, organelle, protein-containing complex, extracellular region, etc (Fig. 2A). The up-regulated transcripts with BP GO terms during the milking stage were biological regulation, developmental process, metabolic process, localization, cellular component organization, locomotion, etc while MF GO terms were catalytic activity, binding, transporter activity, transcription regulation activity, etc (Fig. 2B). Among the down-regulated transcripts during the spike emergence and milking stage, 111 and 272 transcripts were enriched with GO terms, respectively (Fig. 2C and 2D).

3.5 DEGs involved in transportation and uptakes of mineral

The uptake, transportation and accumulation of minerals (Zn, Fe, Cu, Ca, Cu, Cd etc.) from root to different parts of the plant is a complex regulatory process controlled by group of genes. The protein families involved in the transportation and accumulation of minerals in different parts of the plants from root/shoot have been identified in the current experiment of barnyard millet which includes ABC

Transporter family proteins, Calcium-dependent kinase family, Ferritin, Metal ion binding, Iron-sulfur cluster binding, Cytochrome family, Zinc finger transcription factor family, Ferredoxin–NADP reductase type 1 family, Putative laccase, Multicopper oxidase family and Terpene synthase family (Table 4). The numbers of up-regulated transcripts belonging to ABC transporter family were 4 and 5 during spike emergence and milking stage respectively. Transcripts belonging to the metal ion binding protein family were highly up-regulated during both the stages i.e. 218 transcripts during spike emergence and 105 during milking stage indicating their role in metal ion binding activity. The number of transcripts in other protein family-like calcium-dependent kinase (8 each in spike emergence and milking stage), Ferritin (1 in spike emergence stage), iron-sulfur cluster binding (27 each in spike emergence and milking stage), cytochrome (26 in spike emergence and 8 in milking stage), Zinc finger transcription factor (8 in spike emergence and 6 in milking stage), Ferredoxin–NADP reductase type 1 family (3 in spike emergence and 2 in milking stage), Putative laccase multi copper oxidase family (1 in spike emergence and 3 in milking stage) and Terpene synthase (6 in spike emergence and 1 in milking stage). The number of up-regulated transcripts related to the family of transportation and uptake of minerals was high during the spike emergence stage compared to the milking stage. This indicates the spike emergence stage is more critical for the accumulation of the minerals in the seeds as compared to the other latter stages and hence during the spike emergence stage the plant should be provided maximum minerals so that they can be accumulated in the seeds.

Table 4
Transcripts having dynamic association with Iron binding and metal-ion transfer during different stages of Barnyard Millet

No	Protein family	Predicted Functions	#Up Regulated in Spike Stage	#Up Regulated in Milking stage
1	ABC Transporter family proteins	ATP binding; ATPase activity, coupled to transmembrane movement of substances	04	05
2	Calcium dependent kinase family	Calcium ion binding; F: protein binding	08	08
3	Ferritin	Primary intracellular iron-storage protein	01	00
4	metal ion binding	Metal ion binding activity	218	105
5	Iron-sulfur cluster binding	Iron-sulfur cluster binding	27	27
6	Cytochrome family	Iron ion binding; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen,	26	08
7	Zinc finger transcription factor family	Transcription factor activity, sequence-specific DNA binding; metal ion binding; transcription regulatory region DNA,	08	06
8	Ferredoxin–NADP reductase type 1 family	Oxidoreductase activity, Ferredoxin reductase catalyzes the final step of electron transfer to make NADPH and ATP in plant chloroplasts during	03	02
9	Putative laccase Multicopper oxidase family	Ferroxidase activity; copper ion binding; plasma membrane; iron ion transport; lignin catabolic process	01	03
10	Terpene synthase family	Lyase activity; metal ion binding; magnesium ion binding	06	01

3.6 Genes involved in high Fe uptake

The transportation of Fe to different parts of the plant and their accumulation in the grains are affected by its own genes and many other external factors like pH, water availability, organic substances, etc.

Many studies have indicated that stress tolerance response (STR) is also responsible for the accumulation of Fe and Zn in the grains or other parts in addition to the genes involved directly in the accumulation of Fe and Zn. In the present study, we found that the further analysis of binding-molecular function revealed the presence of cation binding, iron ion binding, metal cluster binding, organic cyclic compound binding, protein binding, small molecule binding, and zinc ion binding molecular function (Fig. 3). Except for carbohydrate binding, cation binding, chromatin binding, metal cluster binding, and organic cyclic compound binding all the molecular functions related to ion binding were more prominent in the spike emergence stage rather than the milking stage.

3.7 Metabolic process involved in uptakes of mineral transportation

Enhanced metabolic processes involved in uptakes and transportation of minerals during spike emergence and milking stages of high Fe genotypes were explored through KEGG pathway enrichment analysis. During spike emergence and milking stages, about 176 and 56 KEGG pathways were found enriched for up-regulated transcripts. Pathway results infer that Purine metabolism, Thiamine metabolism, Pyrimidine metabolism, Arginine and proline metabolism, Drug metabolism - other enzymes, Tryptophan metabolism, Glycolysis / Gluconeogenesis, phenylpropanoid biosynthesis were highly enriched as compared to other pathways during spike emergence stage of high Fe barnyard millet genotypes. While moderately enriched pathways such as Pyruvate metabolism, Citrate cycle (TCA cycle), Aminoacyl-tRNA biosynthesis, Steroid hormone biosynthesis, and galactose metabolisms were commonly enriched during both stages. Highly enriched enzymes encoded by the expressed genes during the spike emergence stage were responsible for the genes and enzymes to generate hydroxyl cinnamic acids, esters, guaiacyl, syringyl, and lignin. These pathways involved different enzymes i.e., EC:4.1.1.32 - carboxykinase (GTP), EC:1.2.1.3 - dehydrogenase (NAD+), EC:4.1.1.49 - carboxykinase (ATP), EC:2.7.1.90–1-phosphotransferase, EC:2.3.1.12 - acetyltransferase, EC:6.2.1.1 - ligase, EC:2.7.2.3 - kinase, EC:1.8.1.4 - dehydrogenase, EC:1.2.1.5 - dehydrogenase [NAD(P)+], EC:1.2.7.1 - synthase, EC:2.7.1.2 - glucokinase (phosphorylating). This indicates all the above biomolecules are highly produced and indirectly participate in plant metabolism response (Supplementary Fig. 4). The pathway enrichment analysis of DEGs that involved in upregulation of spike stage of high iron containing genotype showed that Zinc-dependent metalloprotease, Ferroxidase complex, ABC transporter family G domain, Flavonoid metabolic process and Iron ion binding were the most enriched pathways. There are 24 genes with 0.03 gene ratio for Active transmembrane transporter activity, whereas 8 genes and 3 genes for iron transport and ferroxidase complex with 0.3 gene ration, respectively. Other pathways those are responsible for iron and metal transport are depicted in Fig. 4.

3.8 Gene co-expression analysis

The co-expression and physical network of total unregulated genes during spike formation stage of High Fe barnyard millet genotype (Fig. 5A-B). Genes were detected by all methods and the connections among them were shown in coloured lines. All black lines represent the connections among these genes. Proteins such as TSC10, SDH2-3, BTS, ECA3, YSL8, AT5G48290, and NdHS are strongly expressed together and

play a critical role in Metal ion transfer, Iron-sulphur, and mineral ion transport. while in the milking stage proteins such as PNsL4, FH, AT3412100 and YSL8 are involved in metal ion transport. Particularly, FH (Frataxin protein) in mitochondria promotes the biosynthesis of heme as well as the assembly and repair of iron-sulfur clusters by delivering Fe (2+) to proteins involved in these pathways and it may play a role in the protection against iron-catalyzed oxidative stress.

3.9 Validation through qRT-PCR

Total 6 contigs representing their probable function for metal transporter, iron sulfur, metal ion binding, auxin-responsive GH3-like protein 2 and cytochrome P450 71B16 were selected for validating them in both high and low Fe containing genotypes (Table 5) during both the stages. The primers from the said contigs were designed and used in qRT-PCR for their validation. The elongation factor (EF1) transcript was considered as an endogenous control and used for normalizing the relative gene expression of transcripts. Differential expression of transcripts was observed in L/H genotypes along with their respective controls. The genes*cbf5, gh32*, and *c71bg* showed up-regulated pattern during both the stages of spike development, while *gtl1* was down-regulated during the spike emergence stage. The gene *bon1* and *tauE* were up-regulated during the spike emergence stage and down-regulated during the milking stage (Fig. 6). The transcripts related to iron and metal uptake were up-regulated during both the stages but expression level was higher during the spike emergence stage compared to milking. Comparison of transcript expression levels between transcriptome data and qRT-PCR depicted a positive correlation although the values for fold change did not exactly match but remained consistent in up and down-regulated expression.

Table 5
List of primers used for the validation through qRT PCR.

Contig No.	Probable function	name		Primer seq	Product size
Contig3088.p1	Metal transporter	cbf5	F	AGTTAGACCACTCGAAGGCA	200
			R	TTGTCATCAACGCAACACCT	
Contig16455.p2	Iron sulfur	gtl1	F	GTCGGCGCCTATAGATCTCC	201
			R	GTGGTCGTCAAGTATCCCGA	
Contig2221.p1	Metal ion binding	bon1	F	GCAAGCAGTACGTCCAGAAG	168
			R	ACTGGTGGGTGACGTAGAAC	
Contig3220.p1	Metal ion binding	tauE	F	TTGCCTGAAACCTCAACAGC	155
			R	AGGTGGACGTAGCACTTGAA	
Contig19585.p1	Auxin-responsive GH3-like protein 2	gh32	F	CCATCAACCAGTACAAGGCG	154
			R	TCCACCGTACCGAATTCCAT	
Contig21556.p2	Cytochrome P450 71B16	c71bg	F	CCCTCCACTATCACCACCAA	160
			R	GTCATACGCCGGACCAAAC	

4. Discussion

Globally spread malnutrition has been the prime focus of associated researchers across the world. The development of micronutrients rich cereal grains is one of the major approaches to overcoming this problem. The ability of genotypes to accumulate the Fe or other micronutrients varies from genotype to genotype and also depends on the stages of grain development. Transcriptome sequencing in wheat and pearl millet has been carried out to find out the genes/transcripts involved in micronutrient accumulation in the grains.

In the present investigation, we have tried to identify the genes responsible for the accumulation of Fe in the grains of barnyard millet. These genes include ABC Transporter family proteins, Calcium-dependent kinase family, Ferritin, Metal ion binding, Iron-sulfur cluster binding, Cytochrome family, Zinc finger transcription factor family, Ferredoxin–NADP reductase type 1 family, Putative laccase, Multicopper oxidase family and Terpene synthase[27,28,29]. However, the stage of spike development is also for maximum accumulation of Fe in the grains.

A total of 27 iron-sulfur transcripts were found enriched during the spike and milking stage of barnyard millet. These proteins are indirectly associated with plant-type Ferredoxin (Fd), which is a small [2Fe-2S] cluster-containing protein that plays an important role in iron-sulfur proteins interaction and also interacts

with many other plant metabolites[30]. Importantly, only ferritin-related genes were expressed in the spike stage of high Fe-containing genotype of barnyard millet which concedes with the earlier study that suggests that ferritin enables the storage of Fe within the meristematic zone of *Arabidopsis thaliana* tissue architecture[31]. Many metabolic processes including, purine-pyrimidine metabolism, phenylpropanoid biosynthesis, and terpene synthase were found putatively enriched in elite barnyard millet genotypes which were also reported in other plants in response to stress-responsive pathways[32]. In addition to this, the presence of multiple DEGs and the co-factor of Fe and ion transport-related proteins indicate their association with the regulation of several key genes.

Iron and metal ion transport assembly pathways in barnyard millet show the localization of Fe-S clusters and haem in plant cells. Iron-sulfur clusters do not exist in free form, and are only stable within a protein fold and are transferred to nucleus. The iron is accumulated by three biogenesis systems *viz*., cytosol, mitochondria and plastid. The different protein families identified in this research which include ABC transporter family proteins, ferritin, Metal ion binding, Iron-sulfur culster binding, cytochrome family, play a significant role in Fe accumulation in the grains *via* mentioned biogenesis systems. In addition to that, for the biosynthesis of Fe-S clusters, both plastids and mitochondria harbour complete assembly pathways which have been reflected in whole transcriptome of barnyard millet (Fig. 7).

The present investigation reveals that the spike emergence stage is more critical for the accumulation of micronutrients in comparison to that of the milking stage in barnyard millet. So, if the soil needs to be enriched for a higher accumulation of micronutrients in the grain, then it should be done at the time of spike emergence and not once the spike has emerged. Application of micronutrients in the soil at the time of spike emergence will accumulate maximum Fe or other micronutrients in the grain because the maximum numbers of transcripts responsible for the accumulation of micronutrients were active during the spike emergence stage as compared to the milking stage. These findings will help the breeder in developing the genotypes with a higher capability of accumulating the Fe and other micronutrients in barnyard millet and will also help the agronomist in managing the soil nutrient for the higher accumulation of micronutrients in the grains.

5. Conclusions

Barnyard millet is a climate-resilient nutritively rich cereal crop with the potential to provide sufficient Fe content required in daily diet. It has substantial genetic variability for grain Fe content among the different genotypes. Our results on transcriptome sequencing during two stages of spike development in high- and low-grain Fe-containing barnyard genotypes, revealed that the spike emergence stage is more critical for the accumulation of Fe as compared to the milking stage. The supply of Fe and other micronutrients to the plant during the spike emergence stage will accumulate more Fe and other nutrients in the grain as compared to other stages of spike development. The identified protein families include ABC Transporter family proteins, Calcium-dependent kinase family, Ferritin, Metal ion binding, Iron-sulfur cluster binding, Cytochrome family, Zinc finger transcription factor family, Ferredoxin–NADP reductase type 1 family, Putative laccase, Multicopper oxidase family and Terpene synthase family may play a

significant role in Fe accumulation in the grains. These Ferritin and Iron-sulfur cluster binding genes along with other candidate genes should be further investigated in diverse barnyard germplasm. The breeder-friendly marker systems for the improvement of grain Fe content in the barnyard millet can be developed from the transcriptome data generated in this study for the development of millet breeding programs.

Declarations

ACKNOWLEDGMENTS

The authors would like to especially thank Junagadh Agricultural University for providing laboratory facilities during the experiment. We are also thankful to the Director, ICAR-Indian Institute of Millets Research (https://www.millets.res.in/), Hyderabad for providing genotypes of barnyard millet under the Material Transfer Agreement (MTA).

AUTHOR CONTRIBUTIONS

SP and RST designed the study. SP and HD conducted the experiments. JK, SP and HD performed the field experiment, transcriptome sequencing, and other works involved in this study. JK and SP analyzed the data. RST and SP wrote the manuscript with the contribution of JK and SB.

Funding

This research was not supported by any agency.

Data availability

All the data presented in the manuscript are publicly available in NCBI with ID PRJNA748838. Some of the gene expression data generated during this study are available as a supplementary file while the rest of the data can be availed from the corresponding author(s) on request. All authors have read and accepted the MS.

References

1. Gomashe, S. S. Barnyard millet: present status and future thrust areas in millets and sorghum: biology and genetic improvement. *John Wiley & Sons: Hoboken*, NJ, USA, 184-198 (2016).
2. IIMR. Annual Report 2017-18. Hyderabad: Indian Institute of Millets Research. (2018).
3. Sood, S. et al. Barnyard millet - a potential food and feed crop of future. *Plant Breed*. **134**, 135-147 (2015).
4. Ugare, R., Chimmad, B., Naik, R., Bharati, P. & Itagi, S. Glycemic index and significance of barnyard millet (*Echinochloa frumentacae*) in type II diabetics. *J Food Sci Technol*. **51(2)**, 392-395 (2014).

5. Kumari, S. K. & Thayumanavan, B. Comparative study of resistant starch from minor millets on intestinal responses, blood glucose, serum cholesterol and triglycerides in rats. *J Food Sci Technol*. **75**, 296-302 (1997).
6. World Health Organization. Micronutrients. https://www.who.int/health-topics/micronutrients (accessed June 2022).
7. Longvah, T., Ananthan, R., Bhaskarachary, K. & Venkaiah, K. Indian Food Composition Table. Hyderabad: National Institute of Nutrition. 1-578 (2017).
8. Tomar, R. S. Molecular markers and plant biotechnology. *New India Publishing*, New Delhi (2010).
9. Annadurai, R. S., Jayakumar, V. M. & Mugasimangalam, R. C. Next generation sequencing and de novo transcriptome analysis of *Costus pictus* D. Don, a non-model plant with potent anti-diabetic properties. *BMC Genom*. **13**, 663-672 (2012).
10. Lu, J. et al. Transcriptome analysis of Nicotiana tabacum infected by Cucumber mosaic virus during systemic symptom development. *PloS one*, **7(8)**, e43447 (2012).
11. Jaiswal, S. et al. Transcriptomic signature reveals mechanism of flower bud distortion in witches'-broom disease of soybean (*Glycine max*). *BMC Plant Biol*. **19(26)**, 1-12 (2019).
12. Strickler, J. H. et al. Phase I study of bevacizumab, everolimus, and panobinostat (LBH-589) in advanced solid tumors. *Cancer Chemother Pharmacol.* **70(2)**, 251-258 (2012).
13. Arun-Chinnappa, K. S. & McCurdy, D. W. De novo assembly of a genome-wide transcriptome map of *Vicia faba* (L.) for transfer cell research. *Front. Plant Sci*. **6(217)**, 1-9 (2015).
14. Andrews, S. "FastQC," https://qubeshub.org/resources/fastqc (2015).
15. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. **29**, 644-52 (2011).
16. Haas, B. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. **8**, 1494-1512 (2013).
17. Huang, X. and Madan, A. CAP3: A DNA sequence assembly program. *Genom Res*. **9(9)**, 868-877 (1999).
18. Limin, F., Beifang, N., Zhengwei, Z., Sitao, W. & Weizhong, L. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinform*. **28(23)**, 3150-3152 (2012).
19. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359 (2012).
20. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinform*. **26(1)**, 139-40 (2010).
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol.* **215,** 403-10 (1990).
22. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. BLAST+: architecture and applications. *BMC Bioinform*. **10(421)**, 1-9 (2009).

23. Usadel, B. et al. A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant Cell Environ*. **9**, 1211-1229 (2009).
24. Ge, S. X., Jung, D. & Yao, R. A graphical gene-set enrichment tool for animals and plants. *Bioinform*, **36(8)**, 2628-2629 (2020).
25. Thiel, T. MISA-Microsatellite identification tool. (2003). http://pgrc.ipk-gatersleben.de/misa/
26. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods*, **25(4)**, 402-408 (2001).
27. Satyavathi, C. T. et al. Stage specific comparative transcriptomic analysis to reveal gene networks regulating iron and zinc content in pearl millet [*Pennisetum glaucum* (L.) R. Br.]. *Sci Rep*. **12(276)**, 1-13 (2022).
28. Shi, S., et al. The Arabidopsis Calcium-Dependent Protein Kinases (CDPKs) and their roles in plant growth regulation and abiotic stress responses. *Int J Mol Sci*. **19 (7)**, 1900 (2018).
29. Wimalasekera, R. & Scherer, G. F. Involvement of mitogen-activated protein kinases in abiotic stress responses in plants. Plant Metabolites and Regulation Under Environmental Stress. 389-395 (2018).
30. Hanke, G. & Mulo, P. Plant type ferredoxins and ferredoxin-dependent metabolism. *Plant Cell Environ*. **36(6),** 1071-84 (2013).
31. Reyt, G., Boudouf, S., Boucherez, J., Gaymard, F. & Briat, F. Iron- and Ferritin-Dependent Reactive Oxygen Species Distribution: Impact on Arabidopsis Root System Architecture. *Mol Plant*. **8(3)**, 439-53 (2014).
32. Vogt, T. Phenylpropanoid biosynthesis. *Molecular plant*, **3(1)**, 2-20 (2010).

Figures

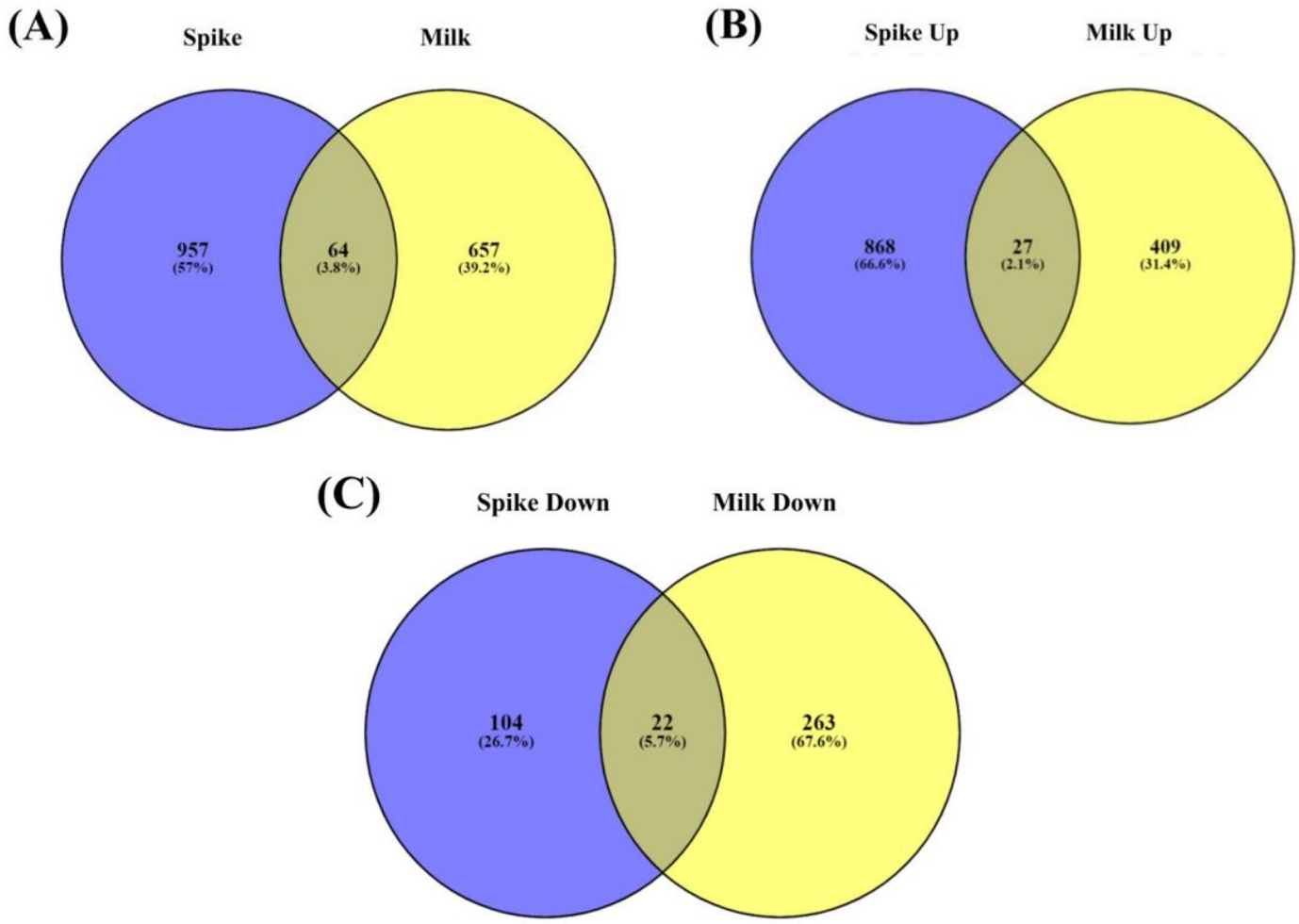

Figure 1

Over-all distribution of Differentially Expressed Genes (DEGs) in Barnyard millet genotype (A)- Venn diagram showing the distribution of unique and common DEGs in spike and milking stages of barnyard millet; (B) - Venn diagram showing the distribution of up regulated genes in spike and milking stages of Barnyard millet; (C)- Venn diagram showing the distribution of down regulated genes in spike and milking stages of Barnyard millet

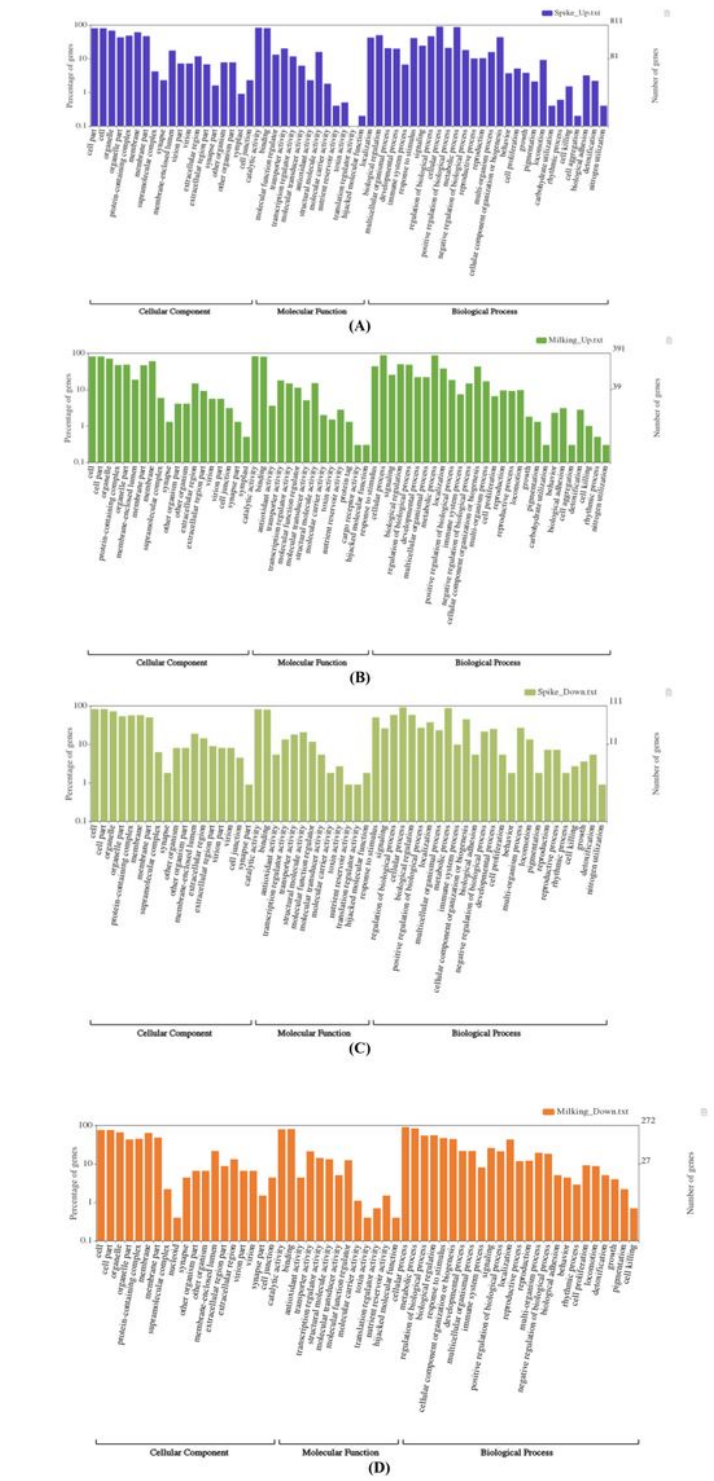

Figure 2

(A)- GO enrichment analysis of DEGs were classified into up-regulated regulated in spike stage (cellular component, molecular function and biological process sub categories); (B)- GO enrichment analysis of DEGs were classified into up-regulated regulated in milking stage (cellular component, molecular function and biological process sub categories); (C)- GO enrichment analysis of DEGs were classified into down-regulated regulated in spike stage (cellular component, molecular function and biological process sub

categories); (D)- GO enrichment analysis of DEGs were classified into down-regulated regulated in milking stage (cellular component, molecular function and biological process sub categories)

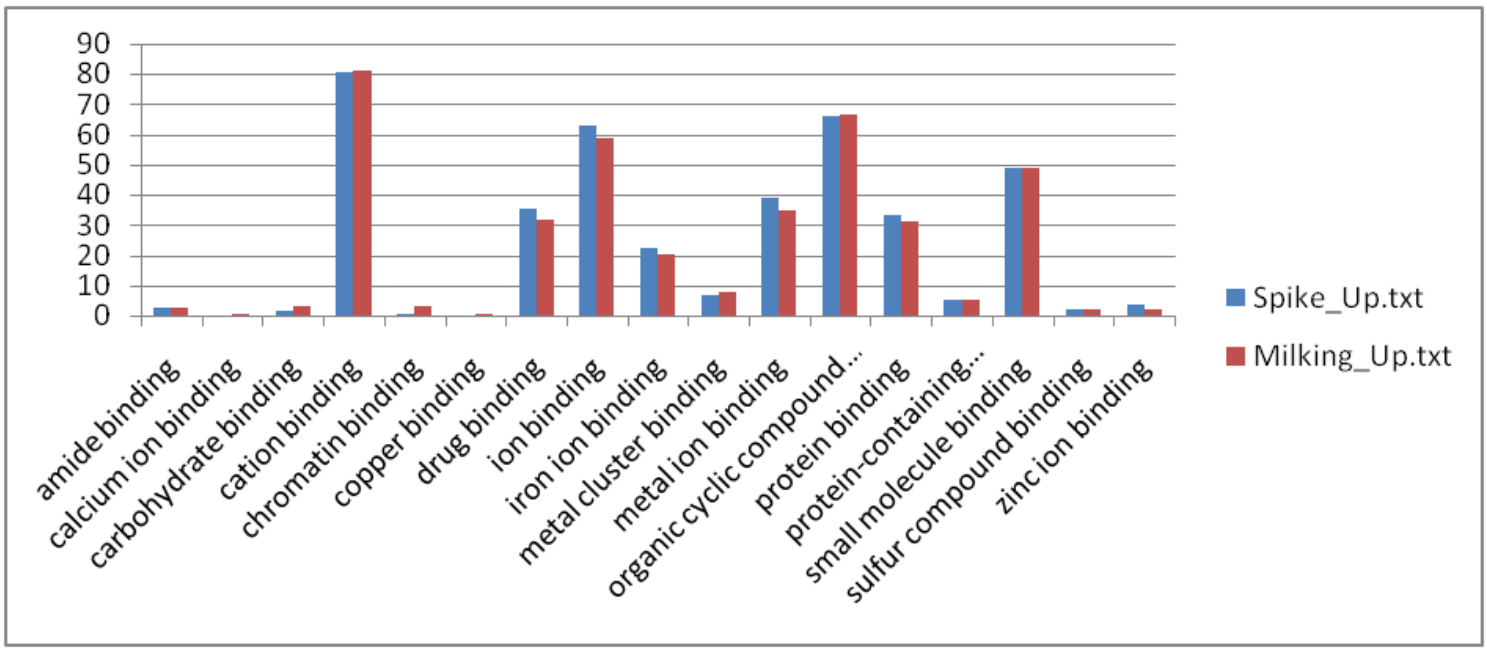

Figure 3

Classification of ion binding component of molecular function which were up-regulated during the Spike and Milking stages of Barnyard Millet

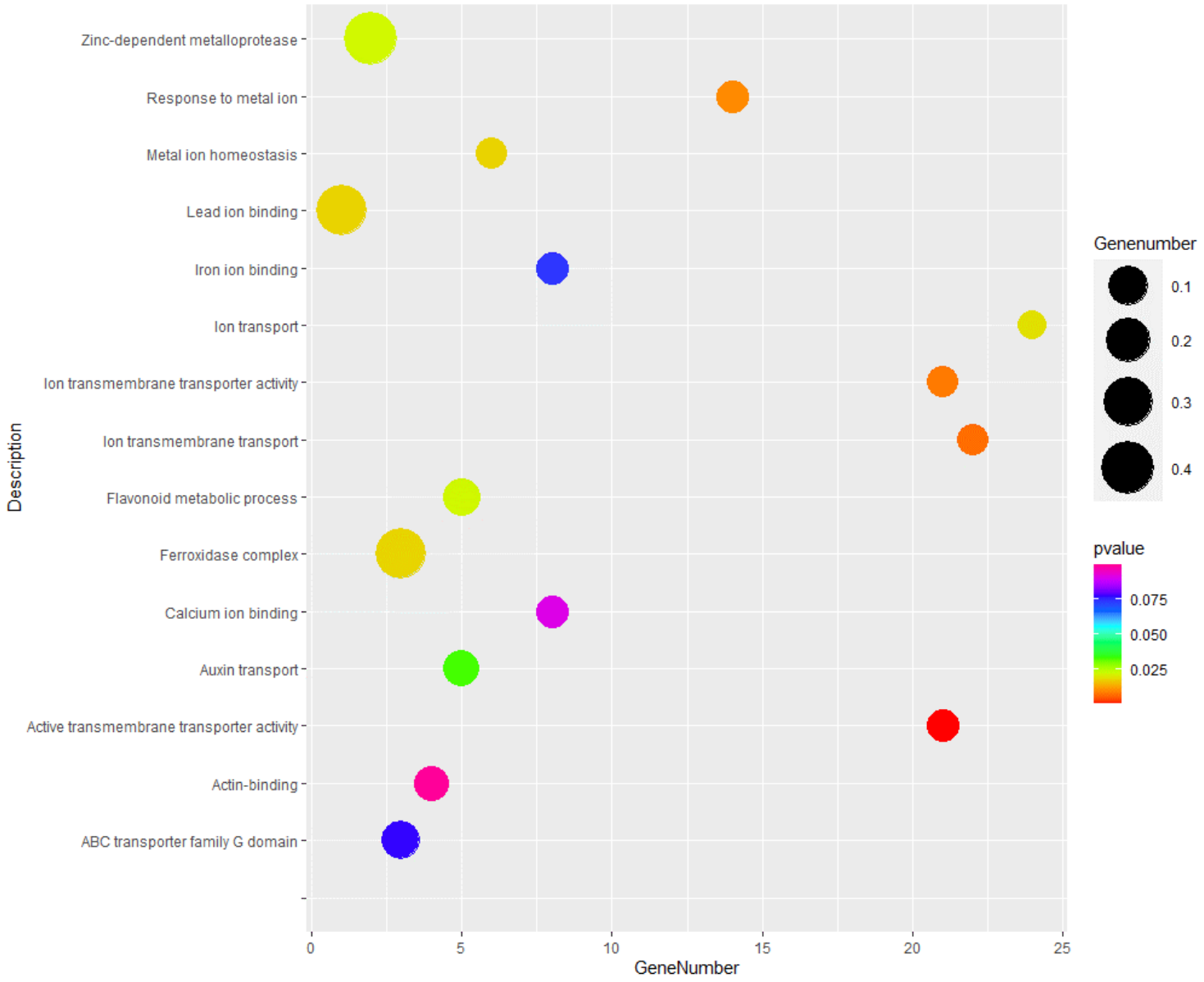

Figure 4

Metal and iron uptake Pathway enrichment analysis of spike stage of high iron containing genotype. Each individual in the figure represents a pathway, the ordinate represents the name of the pathway, and the abscissa is the Gene Number, indicating the ratio of the gene proportion annotated to a pathway in the differential gene to that annotated to the pathway in all genes. The colour of the block represents p-value corrected by multiple hypothesis tests.

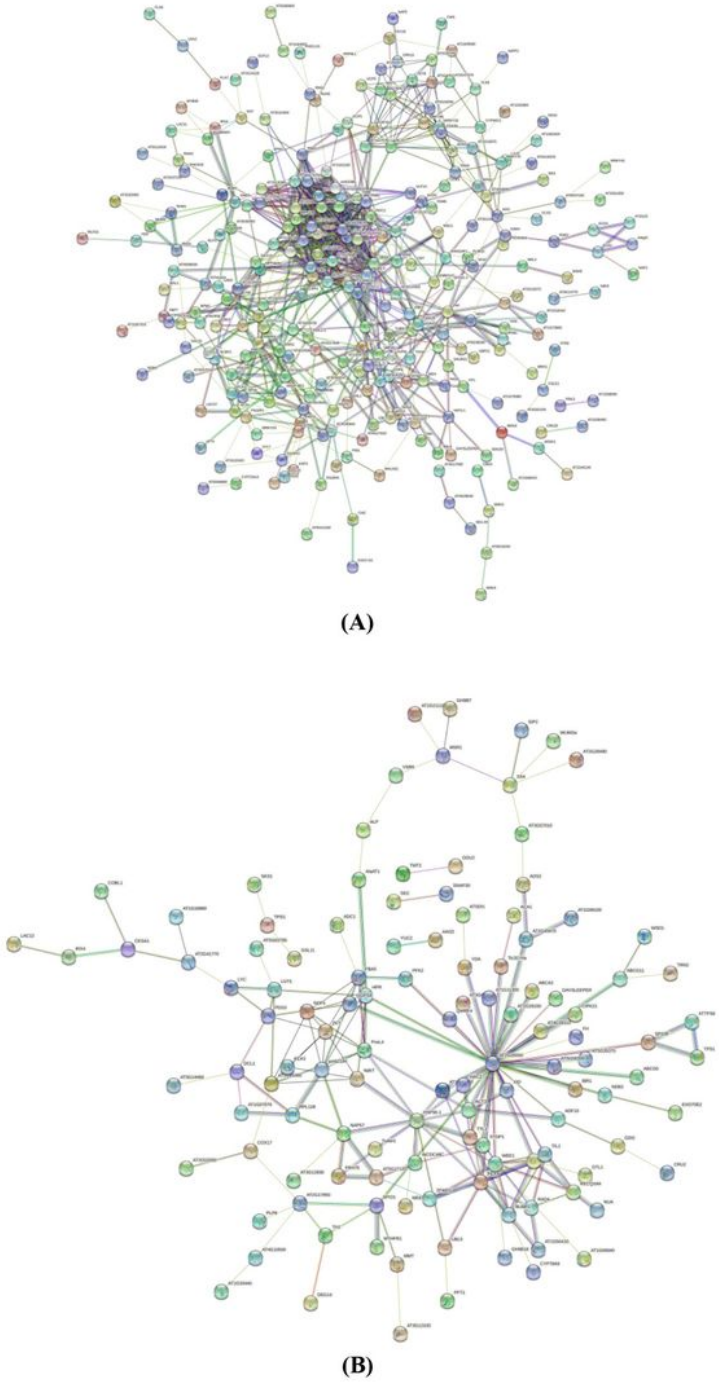

Figure 5

Gene co-expression analyses of up-regulated genes in barnyard millet. (A) - Gene co-expression analyses of up-regulated genes in spike stage of barnyard millet; (B) - Gene co-expression analyses of up-regulated genes in milking stage of barnyard millet

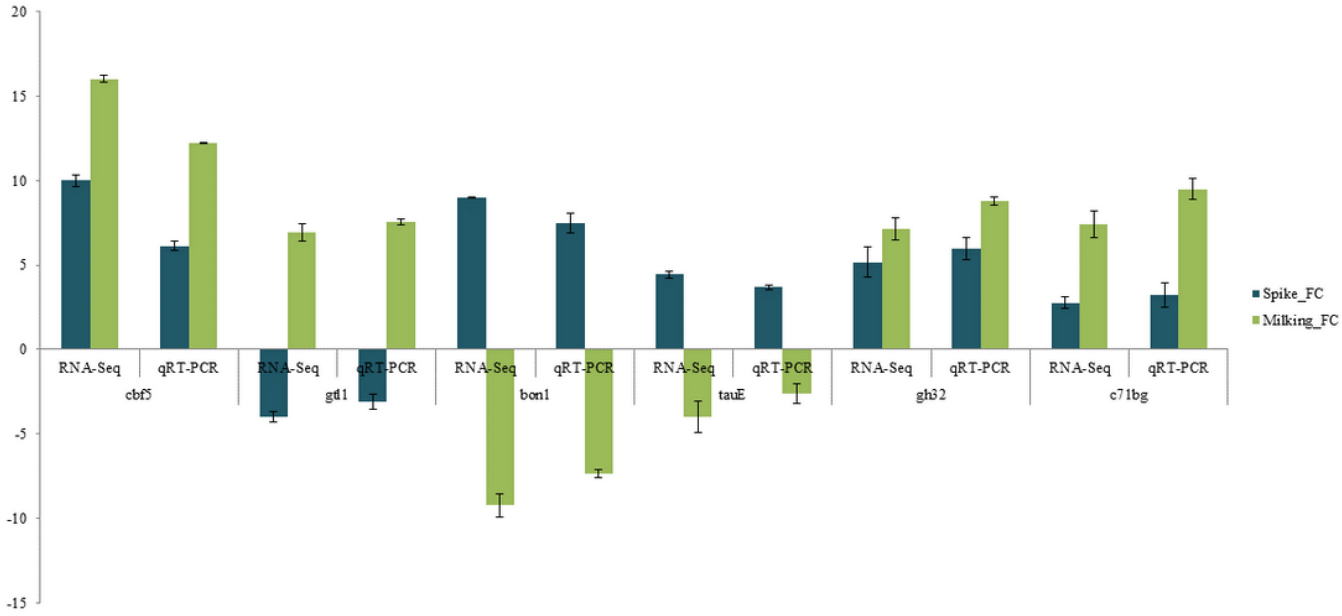

Figure 6

qRT PCR validation of transcripts

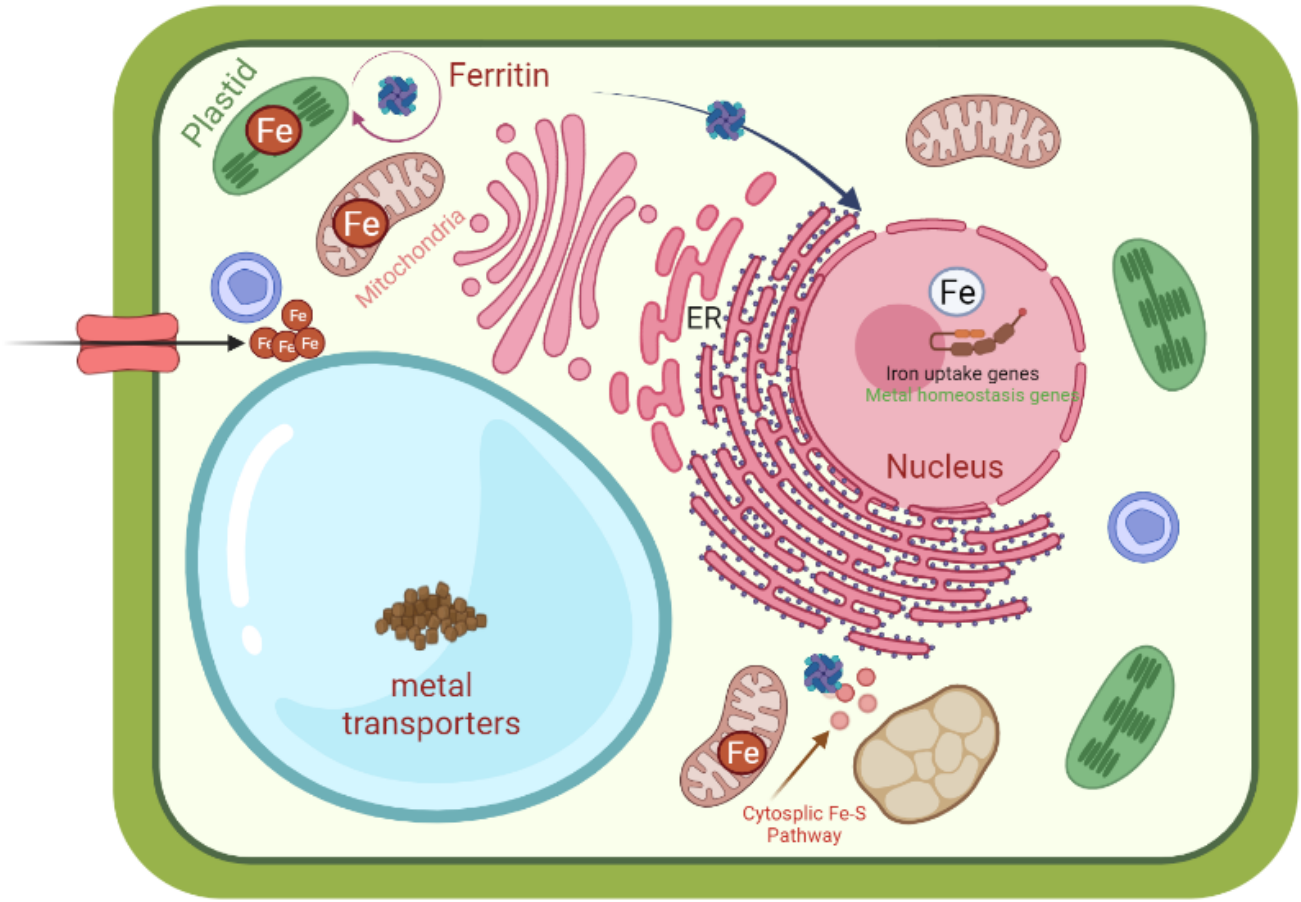

Figure 7

Graphical summary of transcriptome profiling of high and low grain-iron containing Indian barnyard millet

Supplementary Files

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

- SupplementaryFigure1.docx
- SupplementaryTableS1.xlsx
- SupplementaryTableS2.xlsx
- SupplementaryTableS3.xlsx