



Charting oat (*Avena sativa*) embryo and endosperm transcription factor expression reveals differential expression of potential importance for seed development

Sandeep Kumar Kushwaha¹ · Åsa Grimberg¹ · Anders S. Carlsson¹ · Per Hofvander¹

Received: 29 November 2018 / Accepted: 29 April 2019 / Published online: 9 May 2019
© The Author(s) 2019

Abstract

Uniquely, oat, among cereals, accumulates an appreciable amount of oil in the endosperm together with starch. Oat is also recognized for its soluble fibers in the form of β -glucans. Despite high and increasing interest in oat yield and quality, the genetic and molecular understanding of oat grain development is still very limited. Transcription factors (TFs) are important regulatory components for plant development, product quality and yield. This study aimed to develop a workflow to determine seed tissue specificity of transcripts encoding transcription factors to reveal differential expression of potential importance for storage compound deposition and quality characters in oat. We created a workflow through the de novo assembly of sequenced seed endosperm and embryo, and publicly available oat seed RNAseq dataset, later followed by TF identification. RNAseq data were assembled into 33,878 transcripts with approximately 90% completeness. A total of 3875 putative TF encoding transcripts were identified from the oat hybrid assemblies. Members of the B3, bHLH, bZIP, C3H, ERF, NAC, MYB and WRKY families were the most abundant TF transcripts. A total of 514 transcripts which were differentially expressed between embryo and endosperm were identified with a threshold of 16-fold expression difference. Among those, 36 TF transcript homologs, belonging to 7 TF families, could be identified through similarity search in wheat embryo and endosperm EST libraries of NCBI Unigene database, and almost all the closest homologs were specifically expressed in seed when explored in WheatExp database. We verified our findings by cloning, sequencing and finally confirming differential expression of two TF encoding transcripts in oat seed embryo and endosperm. The developed workflow for identifying tissue-specific transcription factors allows further functional characterization of specific genes to increase our understanding of grain filling and quality.

Keywords *Avena sativa* · Oat transcriptome · Embryo and endosperm · Transcription factor · Seed development

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00438-019-01571-x>) contains supplementary material, which is available to authorized users.

✉ Sandeep Kumar Kushwaha
sandeep.kushwaha@slu.se

Åsa Grimberg
asa.grimberg@slu.se

Anders S. Carlsson
anders.carlsson@slu.se

Per Hofvander
per.hofvander@slu.se

Abbreviations

β -Glucan	(1 \rightarrow 3) (1 \rightarrow 4)- β -D-Glucan
bp	Base pair
DBDs	DNA-binding domains
FPKM	Fragments per Kilobase of transcript per Million mapped reads
HMM	Hidden Markov model
PE	Paired end
SE	Single end
TFs	Transcription factors
TRs	Transcriptional regulators
TPM	Transcripts per Million

¹ Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden

Introduction

Oat (*Avena sativa*) is a cereal grown for its grain and is cultivated in Europe, North America, Russia, Australia and northern China. Currently, oat is gaining in popularity due to its potential as a functional food crop and contents of health benefit promoting compounds as antioxidants, lipids and high levels of globular proteins. Oat is also well known as an excellent source of β -glucan (1–3; 1–4 mixed link β -D-glucan), a dietary fiber with documented and approved health claims regarding blood glucose stabilizing and cholesterol lowering properties (Butt et al. 2008; Rasane et al. 2015; Chen et al. 2016). These different compounds are accumulating in the oat grain in which the endosperm tissue makes out the major and the embryo tissue only a minor part. Even considering these interesting characteristics of oat, the understanding of oat seed genetics and molecular background is still not well developed. Oat is a hexaploid, consisting of three different genomes (AACCDD) with a unit size of 7 chromosomes, i.e., 21 chromosomes in a haploid oat genome ($2n = 6x = 42$) and a total size of 11,300 Mbps. Recent advances in genome and transcriptome sequencing technologies have revolutionized non-model plant research due to unprecedented sensitivity and accuracy combined with relatively low sequencing costs (Metzker 2010; Martin and Wang 2011) and it has been efficiently employed in the discovery of new genes characteristic for tissue and developmental stages, biomarkers identification, detection of new alternative splice variants, allele-specific gene expression, SNP discovery in genes and epigenetic gene regulation (Hahn et al. 2009; Marguerat and Bähler 2010; Anders et al. 2012). However, even though great progress has been made in bioinformatics data analysis, molecular data of non-model plants are still imposing new challenges to methods of bioinformatics data analysis and approaches for the genetic and molecular understanding of polyploid plants.

Spatial and temporal differential gene expression regulates different aspects of plant development, such as morphology and physiology, and controls metabolite production, protein and oil quality and quantity as well as acclimation to environmental changes. TFs are among the key regulators of gene expression (Shen et al. 2010; Van Belleghem et al. 2012; Ma et al. 2016). Broadly, gene expression regulators modulate gene transcription either positively or negatively. TFs can interact directly with DNA via DNA-binding domains or via other TFs. TFs interaction with chromatin makes genes more accessible by modifying chromatin structure and can also facilitate the recruitment of the basal transcription machinery. TFs are characterized by DNA-binding domains which can be

present in one or multiple copies in the same sequence (Charoensawan et al. 2010; Thiriet-Rupert et al. 2016). TFs have gained interest in genetic engineering of plants for directing quality and yield traits in crops as they often control specific sets of genes and thereby circumvent the individual engineering of several steps in pathways (Grimberg et al. 2015). Changes in transcription factor regulation have been quite common in crop plant development over the ages (Doebley et al. 2006). In plant breeding today genes encoding transcription factors are also candidate targets for the development of new varieties, with for example improved seed quality and flowering time (Jung and Müller 2009; Webb et al. 2016). Therefore, the identification and characterization of TFs in plants like oat are of great importance to answer basic biological questions and to identify potential targets for genetic enhancement through breeding.

In this paper, a de novo transcriptome assembly approach was used for oat embryo and endosperm transcriptome analysis since the oat genome is yet to be published and available in the public domain and the resulting oat transcriptomes were explored for the identification of transcripts encoding TFs. The objectives of this research were to establish a workflow for determining seed tissue specificity of assembled transcripts: (1) RNAseq assembly and annotation for the functional understanding of oat embryo and endosperm, (2) identification of embryo and endosperm-specific TFs, and (3) exploration and comparison of oat TF families with available wheat genetic resources.

Methods

Plant material, sample preparation, data retrieval and sequencing

Oat (*cv. Matilda*, Lantmännen, Svalöv, Sweden) was grown in controlled growth chambers (Biotron, SLU-Alnarp, Sweden) under fluorescent light ($200 \mu\text{mol m}^{-2} \text{s}^{-1}$) during a 16/8 h and 21 °C/18 °C light/dark cycle at 60% humidity. For transcriptome analyses, grains from three individual plants were harvested at a mid-developmental stage (approximately 15–18 days post anthesis) and pooled. Embryos were manually dissected from grains under a binocular using a scalpel, rinsed in water (to remove any endosperm left) and then snap frozen into liquid nitrogen. Liquid endosperms were squeezed out from grains (after embryos were removed) onto a spatula and then dipped into liquid nitrogen. For PCR validation experiments, a simplified sampling procedure was used; grains at mid-developmental stage (approximately 15–18 days post anthesis) were cut into two parts using a scalpel; embryo part and endosperm part before snap-freezing into liquid nitrogen. Samples were stored in -80 °C

until extraction. For transcriptome analyses, total RNA was extracted from frozen tissues by homogenizing material in Plant RNA Reagent according to instructions (Invitrogen, Carlsbad, USA). For PCR validation experiments, seed tissues were first ground in steel containers using metal beads in a mixer mill (MM400, Retsch, Haan, Germany), before extraction of RNA as described above. RNA integrity and concentration were determined using Experion RNA StdSens analysis kit (BioRad, Hercules, USA). Total RNA was DNase treated (TurboDNase, Ambion, Carlsbad, USA) before sequencing. Libraries were prepared at BGI (Shenzhen, China) using oligo(dT)-enriched mRNA subsequently fragmented (about 200 bp) and first-strand cDNA synthesis by random hexamer–primer according to their library preparation protocol and oat embryo and endosperm RNA libraries were sequenced through Illumina sequencing platform HiSeq 2000 as unpaired-end reads. Public hexaploid oat seed transcriptome was downloaded from NCBI SRA. Sequencing reads were trimmed by removing adapter sequences and low quality sequences and filtered reads which had a quality score of less than 20 using Nsoni clip version 0.130. Moreover, reads with a length of less than 20 bps were also removed (Chauhan et al. 2014).

Transcriptome assembly and assessment

Combined de novo transcriptome assembly strategy was adopted to explore best transcriptome assembly with trimmed reads through different RNAseq assembler like Trinity (Haas et al. 2013) and Bridger (Chang et al. 2015). To make complete transcriptome, all the generated assemblies were clustered through cd-hit clustering software (Huang et al. 2010). To assess the quality of the transcriptome assembly, sequencing reads were mapped back to the assembled transcripts using the Bowtie2 aligner (Langmead and Salzberg 2012). BUSCO analysis was performed to evaluate completeness of a de novo assembled transcriptome at default parameters (Kriventseva et al. 2015). To explore TFs families in oat, we also used publicly available oat data (SRR850241) with sequenced RNAseq and number of assemblies was generated through different Kmer size and Kmer coverage of de-Bruijn assemblers Trinity, Bridger, SOAPdenovo-trans (Xie et al. 2014) and IDBA-trans (Peng et al. 2013). This assembler generates assemblies from short-read sequences of RNAseq data using the de-Bruijn graph algorithm. Transdecoder-v.2.0 package was used to predict open-reading frames from generated transcripts.

Transcription factor exploration and functional annotation

TFs families are characterized by its conserved sequence regions, i.e., DNA-binding domains (DBDs) and used as basis

of TF families identification and characterization. HMMER V3.0 package (<http://hmmer.janelia.org/>) was used to build Hidden Markov Model (HMM) profiles to identify the oat TFs (Eddy 2011). In this study, multiple sequence alignment seeds of TF families were retrieved from the PlnTFDB. HMM profiles of the TF families from their multiple sequence alignment seeds were constructed using the hmmbuild program in HMMER V3.0. These HMM profiles were used to identify TFs. The hmmsearch program of HMMER package was used to identify TF sequences among generated non-redundant set of oat protein sequences. The protein sequences matching the HMM profiles (e value < 0.01) were considered to be TFs. TFs families' identification and family assignment rules are well described in plant TF database (PlantTFDB V3.0). Several sequences were identified for more than one TF families. We detected these sequences and removed redundant TF sequences as described by Iida (Iida et al. 2005).

Abundance estimation and differential gene expression analysis

RSEM version 1.2.7 was used for abundance estimation of the transcriptome assembly (Li and Dewey 2011). RSEM is a software package used to estimate the gene and isoform expression levels from RNA transcripts. Abundance estimation was performed using the default parameters. The relative measure of transcript abundance was TPM (Transcripts per Million) and FPKM (Fragments per Kilobase of transcript per Million mapped reads). Differential gene expression was performed through edgeR (Robinson et al. 2010). edgeR was used in identification and analysis of differentially expressed transcripts through Trinity version trinityrnaseq_r20140717 on default settings. To ensure correct assignment of differentially expressed TF transcripts, nucleotide BLAST similarity search was performed against Unigene libraries of wheat (*Triticum aestivum*) seed.

Gene ontology and pathway enrichment analysis

Gene ontology enrichment analysis of differentially expressed transcription factors (DE TFs) was performed by the AgriGO tool (Du et al. 2010) (corrected p value < 0.05) and REVIGO (Supek et al. 2011) was used for analysis of the enriched GO terms. Statistical enrichment of differential expression transcription factors in KEGG pathways (corrected p value < 0.05) was performed by KOBAS software (Mao et al. 2005).

Experimental validation of identified transcription factors

For validation experiments, total RNA was extracted (as described above) from oat grain endosperm and embryo

from three biological replicates, i.e., from seed tissue from three individual plants. Approximately, 10 µg total RNA was DNase treated (TURBO DNA-free kit, Ambion by Thermo Fisher Scientific, Carlsbad, CA, USA). Approximately, 5 µg of DNased RNA was further used for cDNA synthesis using SuperScript III First-Strand Synthesis for q-RT-PCR (Invitrogen by Thermo Fisher Scientific). Fragments from two selected target transcripts (779 bp from a bHLH transcript, and one 603 bp from a Dof transcript) were cloned by PCR amplification in 20 µl reactions in an S1000 Thermal Cycler (BioRad, Hercules, USA) using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). PCR products were separated and sizes of target fragments confirmed by visual inspection on agarose gel electrophoresis. Target bands were cut out and purified from gel using GeneJET Gel Extraction Kit, cloned into pJET1.2 vector using CloneJET PCR Cloning kit according to blunt-end protocol (kits from Thermo Fisher Scientific) and then sequenced using standard vector primers (Eurofins Genomics, Ebersberg, Germany). For semi-quantitative PCR experiments, 150 ng of cDNA was used as template in a final PCR volume of 10 µl. All PCRs were run with 0.05 µM of each primer, 0.2 mM dNTP, and 0.02 U/µl Phusion High-Fidelity DNA Polymerase. Transcripts were amplified using a program with initial 98 °C for 30 s followed by 15, 25 and 35 cycles (or 40 for cloning PCR) of 98 °C for 10 s, annealing temperature for 15 s (see below), 72 °C for 30 s and then a final extension at 72 °C for 5 min. Primers for bHLH transcript; Forward 5'ggtgtgaaggtgaaggggaa3', reverse 5'gaaacacaccagacaccgaag3' (annealing temperature 65.0 °C). Primers for Dof transcript; forward 5'catcaccatatacaagcaagtacc3', reverse 5'gacgatgacggacaaatgt3' (annealing temperature 62.5 °C). PCR products were separated by gel electrophoresis on 2% agarose gel in 1xTAE buffer containing GelRed (Thermo Fisher Scientific) at 110 V for 90 min and then visualized under UV light.

Results

Assembly of transcriptomes

Oat transcriptome was generated through de novo transcriptome assembly approach due to unavailability of oat genome sequences. Embryo and endosperm RNA samples of oat (cv. Matilda, Lantmännen, Svalöv, Sweden) were sequenced through Illumina HiSeq 2000 sequencing platform which yielded around 49.2 million unpaired-ends reads of 50 bp length. Data quality assessment was performed through the FASTQC/0.11.5 tool. High-quality reads were extracted through Neson clip version 0.130 software by removing reads containing adapter sequences, short reads (less than 20) and trimming of reads for low quality score ($Q < 20$) from raw sequence (Chauhan et al. 2014). After quality

control, 23,378,371 reads of embryo and 23,961,462 million reads of endosperm were obtained for further study. Embryo and endosperm-specific transcriptome was generated from newly sequenced endosperm and embryo reads (SE) through Trinity and Bridger de novo transcriptome assembler and assemblies generated using both assemblers were clustered through CD-HIT clustering software to generate a non-redundant set of transcripts. The used approach forms a workflow (Fig. 1). In total, 24,086,480 bases were assembled into 33,878 transcripts with a minimum length of 201 bps, N50 value of 971 bps and average transcript length of 711, with 7514 transcripts having a length of more than 1 kb. To assess the quality of assembly, reads were mapped back to the assembly and the alignment visualized with IGV v.2.3.2. The percentage of uniquely mapped reads was found 70.4% and 56.5% for embryo and endosperm, respectively. After mapping of reads, the assembly was further evaluated for transcript completeness through the BUSCO software. In the transcriptome assembly, BUSCO yielded approximately 60% complete and fragmented BUSCO genes together. It is obvious though that all the genes will not be expressed on an oat embryo and endosperm tissue level which likely is the reason for low percentage of transcriptome completeness.

Transcriptome annotation

The generated transcriptome assembly was clustered at 95% sequence similarity through CD-HIT software which produced 27,334 non-redundant transcripts. TRAPID pipeline was used for transcriptome annotation. A total of 22,026 (80.6%) transcripts with domain annotation and 18,955 (69.3%) transcripts with gene ontology were identified with at least one InterProScan annotation and gene ontology terms, whereas 26,467 transcripts were annotated together with homologous genes when searched on gene families and subfamilies using BLAST, TribeMCL and OrthoMCL. In summary, 26,467 (96.8%) transcripts were annotated together with gene families, functional gene annotation (GO) and InterProScan protein domains annotation. A Venn diagram was generated from gene family, domain and GO annotation results and shows that 17,551 transcripts were annotated through all three methods. Gene ontology and domain identification by InterProScan (IPR) were performed for the assembled transcriptome. A detailed summary of identified domain and GO annotation can be found in Supplementary File 1: Tables S2 and S3.

Identification of transcription factors in oat

A hybrid strategy was adopted to explore TFs in assembled oat transcriptomes. Assemblies were generated from newly sequenced endosperm and embryo reads (SE) and publicly available oat seed RNAseq (PE) (Gutierrez-Gonzalez et al.

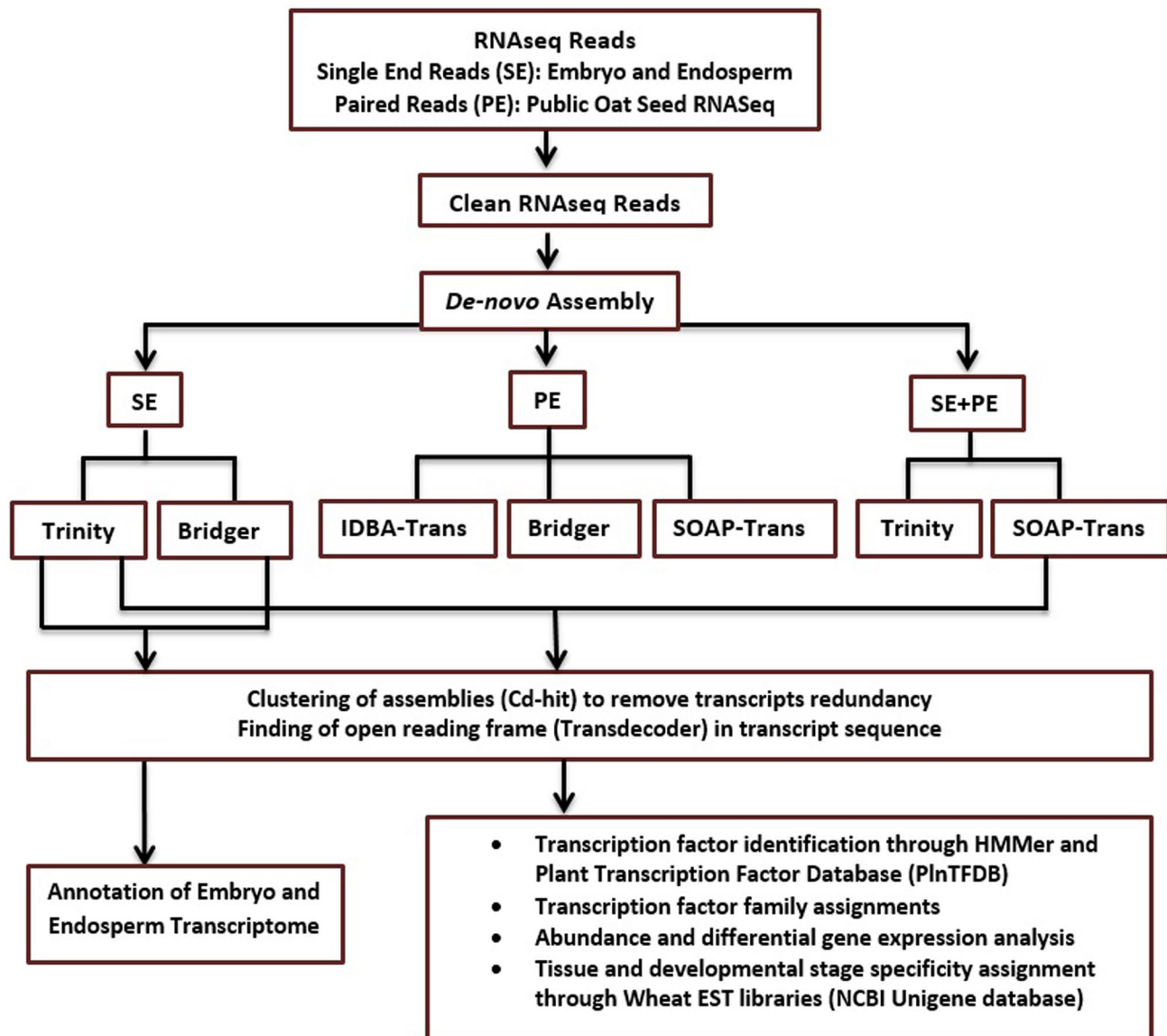


Fig. 1 Schematic representation of used approach for transcriptome assembly, annotation, and identification of differentially gene expression in oat grain tissues. *SE* single-end RNAseq reads of endosperm and embryo, *PE* paired-end reads from oat seeds (color figure online)

2013), respectively, and as a combination of both data sets (PE + SE) through Trinity, Bridger, IDBA-trans, and SOAP-denovo-trans. The approach used is integrated into the workflow (Fig. 1). In order to make a non-redundant transcriptome, all the generated assemblies were clustered through CD-HIT clustering software at 95% sequence similarity cutoff. An Hmm profile for each TF family was generated from the HMMER package through sequences downloaded from the Plant TF database PlantTFDB v3.0. Initially, all generated ORFs were searched against each TF family profile and later family assignment was performed on the basis of domains identified through a Pfam search. In total, 3875 transcripts were identified as putative TFs (Supplementary File 2: Table S4).

Analysis of differentially expressed transcription factors

To understand and decipher the potential importance and functionality of TFs in oat seed and embryo and endosperm, expression abundance analysis was performed for identified TF sequences through RSEM software. In expression abundance analysis 3838 transcripts in public seed data, 2921 in endosperm data and 3438 transcript in embryo data were found related to different TFs families with an expression value more than or equal to an FPKM value of 10. The number of expressed TF transcript sequences (FPKM \geq 10), number of common TF sequences and unique TFs of embryo, endosperm and seed were calculated (Table 1). Differential

Table 1 Number of identified transcription factor transcript sequences in each TF family and expression count (i.e., FPKM \geq 10) for the RNAseq libraries Public Oat (seed), embryo and endosperm

TF families	TF sequences	Embryo	Endosperm	Seed	Com- mon TF sequence	Unique TF sequences		
						Embryo	Endosperm	Seed
AP2	57	56	29	57	28	0	0	0
ARF	69	69	51	69	51	0	0	0
B3	146	141	86	146	83	0	0	2
BBR-BPC	12	12	12	12	12	0	0	0
BES1	16	16	16	16	16	0	0	0
bHLH	242	216	163	238	143	2	0	6
bZIP	287	277	246	287	241	0	0	5
C2H2	40	34	37	40	31	0	0	0
C3H	249	227	200	245	195	1	0	16
CAMTA	43	41	37	43	36	0	0	1
CO-like	8	6	4	8	3	0	0	1
CPP	42	41	42	42	41	0	0	0
DBB	16	16	16	16	16	0	0	0
Dof	49	32	38	49	24	0	0	3
E2F-DP	34	33	28	34	27	0	0	0
EIL	39	39	34	39	34	0	0	0
ERF	215	202	176	210	164	4	0	2
FAR1	105	95	49	102	49	1	0	8
GATA	110	105	110	110	105	0	0	0
GeBP	45	45	44	45	44	0	0	0
GRAS	187	168	126	183	108	3	0	0
GRF	26	26	8	26	8	0	0	0
HB-other	105	100	66	100	64	3	0	5
HD-ZIP	37	36	17	37	16	0	0	0
HSF	119	110	104	118	100	0	0	6
LBD	16	16	12	16	12	0	0	0
LSD	30	30	25	30	25	0	0	0
MIKC	111	57	41	110	26	0	0	38
M-type	28	18	16	28	12	0	0	6
MYB	163	104	118	162	80	0	0	22
MYB-related	384	337	298	383	275	0	0	25
NAC	234	142	173	232	125	0	1	45
NF-YA	46	46	39	45	39	1	0	0
NF-YB	69	69	58	69	58	0	0	0
NF-YC	48	48	48	48	48	0	0	0
Nin-like	49	49	41	47	41	2	0	0
RAV	5	5	5	5	5	0	0	0
S1Fa-like	1	1	1	1	1	0	0	0
SAP	37	37	37	37	37	0	0	0
SBP	37	37	30	37	30	0	0	0
SRS	1	1	0	1	0	0	0	0
TCP	18	18	15	18	15	0	0	0
TriHelix	136	126	128	134	119	1	1	1
WRKY	120	111	83	119	81	0	0	6
YABBY	23	22	3	23	3	0	0	1
ZF-HD	21	21	11	21	11	0	0	0

gene expression analysis was performed through the edgeR statistical package. In total 681, TF assigned transcripts were differentially expressed (DE) with a p value cutoff for FDR 0.001 and fourfold change when embryo and endosperm were compared. In further identification of the most differentially expressed transcripts, 514 highly differentially expressed TF transcripts were found at p value: 0.001 and 16-fold differential expression (Supplementary File 2: Table S5). Differentially expressed TFs between embryo and endosperm and number of differentially expressed TFs in family are shown in Fig. 2a, b, respectively. To narrow down, ensure occurrence and a correct assignment of these TF transcripts, we performed a nucleotide BLAST similarity search against Unigene libraries of wheat (*T. aestivum*) seed,

downloaded from NCBI Unigene libraries. 36 transcripts among the 514 differentially expressed genes were found in a similarity search of seed Unigene wheat libraries which belong to endosperm and embryo sub-tissues at dormant and ripening developmental stages using the applied e value ($1e-10$) filter (Table 2). Altogether, these 36 differentially expressed oat transcripts fell into seven families of TFs.

Gene ontology enrichment analysis

Functional classification of differentially expressed TFs was performed using a gene ontology (GO) analysis and was assigned to biological processes, molecular functions, and cellular compartments classes. GO enrichment of over- or

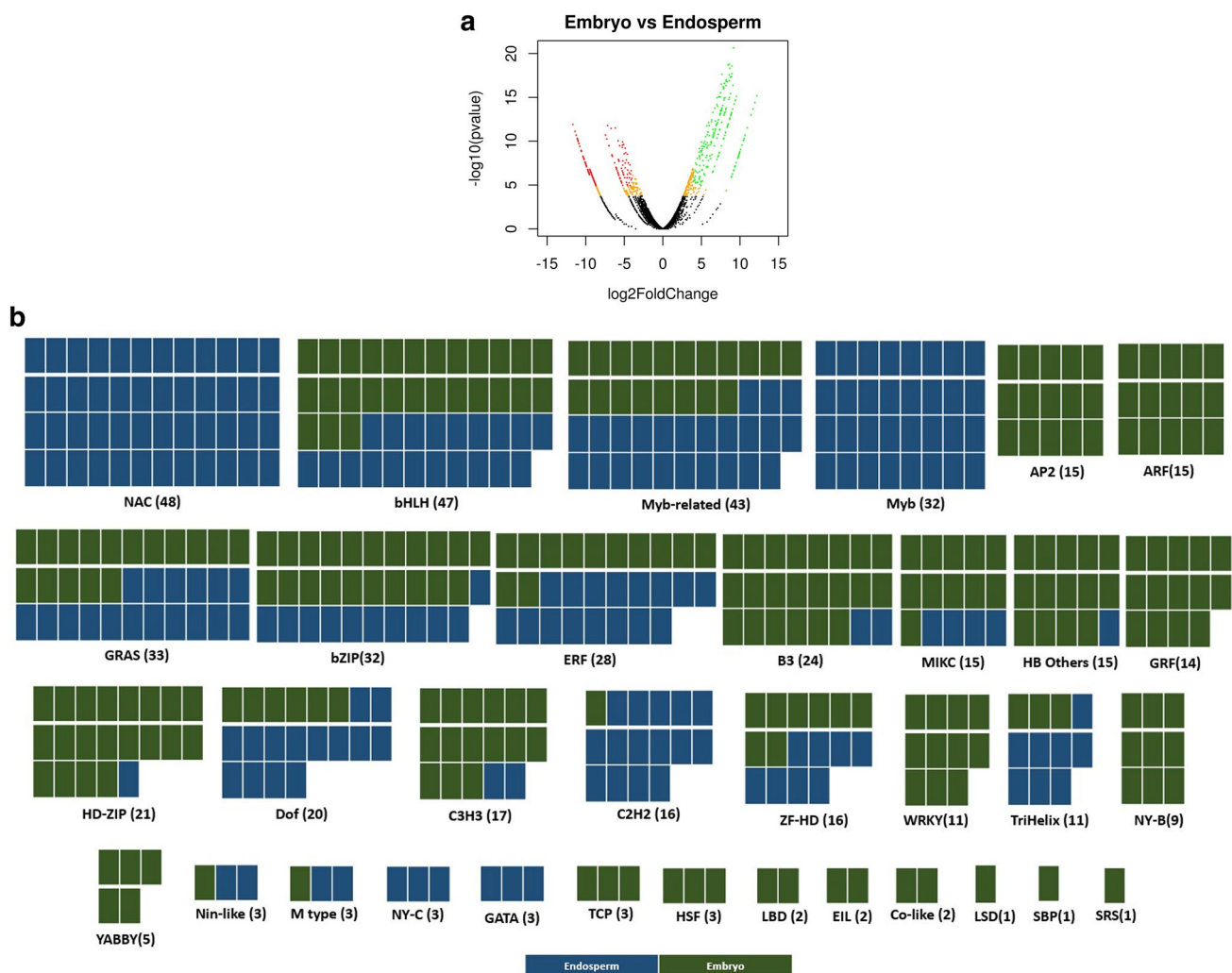


Fig. 2 **a** Volcano plots shows the fold change difference in the expression of transcripts on x -axis, and y -axis indicate the adjusted p values for the differences in expression. The up-regulated genes are represented by green dots and the down-regulated genes are represented by red dots at p value < 0.05 and log fold change 2. All insignificant genes (p value > 0.05) were represented through black dots. All genes

with significance (p value) < 0.05 and without any log fold change criteria were represented by orange dots. **b** Tissue specificity among 514 highly (16-fold) differentially expressed oat transcripts between endosperm and embryo. Individual endosperm and embryo transcripts are represented as blue and green colored boxes, respectively (color figure online)

Table 2 Oat embryo and endosperm-specific transcription factor and their expression values in FPKM

TF family	Transcripts	Embryo	Endosperm	Annotation (NCBI Blast)
B3	AsTF0237	2366.9	85.1	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0309	242.1	0	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0293	540.4	5.6	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0418	351.9	9.5	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0388	270.1	0	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0375	222.7	0	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0346	308.2	7.7	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0269	304.4	0	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0374	279.1	0	VIVIPAROUS 1 homolog transcription factor
B3	AsTF0381	217.6	0	VIVIPAROUS 1 homolog transcription factor
bHLH	AsTF0883	345.9	6.6	Protein spatula containing bHLH
bHLH	AsTF0641	457.7	0	Protein spatula containing bHLH
bHLH	AsTF0818	245.7	0	Protein spatula containing bHLH
bHLH	AsTF0820	258.9	0	Protein spatula containing bHLH
Dof	AsTF2458	1.4	711.9	Prolamin box binding factor 1
Dof	AsTF2419	2.6	803.0	Prolamin box binding factor 2
Dof	AsTF2402	1.1	757.2	Prolamin box binding factor 3
MYB	AsTF4794	14.7	1480.1	Transcription factor myb44-like
MYB	AsTF4997	26.9	914.9	Transcription factor myb44-like
MYB	AsTF4783	2.8	797.2	Transcription factor myb44-like
MYB	AsTF4983	24.1	1097.5	Transcription factor myb44-like
MYB	AsTF4756	4.5	872.1	Transcription factor myb44-like
MYB	AsTF4929	3.9	495.8	Transcription factor myb44-like
MYB	AsTF4988	17.6	629.2	Transcription factor myb44-like
MYB	AsTF4927	1.3	772.2	Transcription factor myb44-like
MYB	AsTF4928	1.4	611.2	Transcription factor myb44-like
MYB_related	AsTF5448	10.0	762.6	Transcription factor myb44-like
MYB_related	AsTF5595	21.6	406.2	Transcription factor myb44-like
MYB_related	AsTF5574	10.8	670.5	Transcription factor myb44-like
MYB_related	AsTF5509	5.8	566.2	Transcription factor
NAC	AsTF5891	6.0	3309.7	Nac domain-containing protein 100-like
NAC	AsTF5975	1.3	236.1	Protein cup-shaped cotyledon 2
NAC	AsTF5963	15.5	4832.4	Nac domain-containing protein 100-like
NAC	AsTF5888	28.9	5883.2	Protein cup-shaped cotyledon 2
Trihelix	AsTF6856	3.4	217.6	Trihelix transcription factor asr3-like isoform × 2
Trihelix	AsTF6861	3.6	115.7	Trihelix transcription factor asr3-like isoform × 2

under-represented GO terms was performed through Fisher's exact test at the p value 0.05 in the AgriGO, and gene ontologies were visualized through REVIGO tool. The GO terms "embryo development", "response to abiotic stimulus", "response to heat", "response to stress", "response to stimulus", and "developmental and multicellular organismal process" are the most highly enriched among DE TFs (Fig. 3a, b), which is in agreement with previous findings (Wickramasuriya and Dunwell 2015). GO terms also represented diverse functional activities corresponding to the mentioned biological processes. The highest numbers of DE TFs were categorized in "binding and transcriptional regulation activities", which have also been reported as

over-represented terms in embryo development (Palovaara et al. 2017). In contrast, structural molecular activities-related GO terms were found least.

Pathway enrichment analysis

To explore further molecular and biological functionality of differentially expressed TFs, the KEGG database was mapped through the KOBAS 3.0 server. In KEGG pathway analysis, 22 pathways were identified, in which two were significantly enriched with p value < 0.05. KEGG enrichment analysis showed that DE TFs are mainly involved in plant hormone signal transduction and circadian rhythm-plant

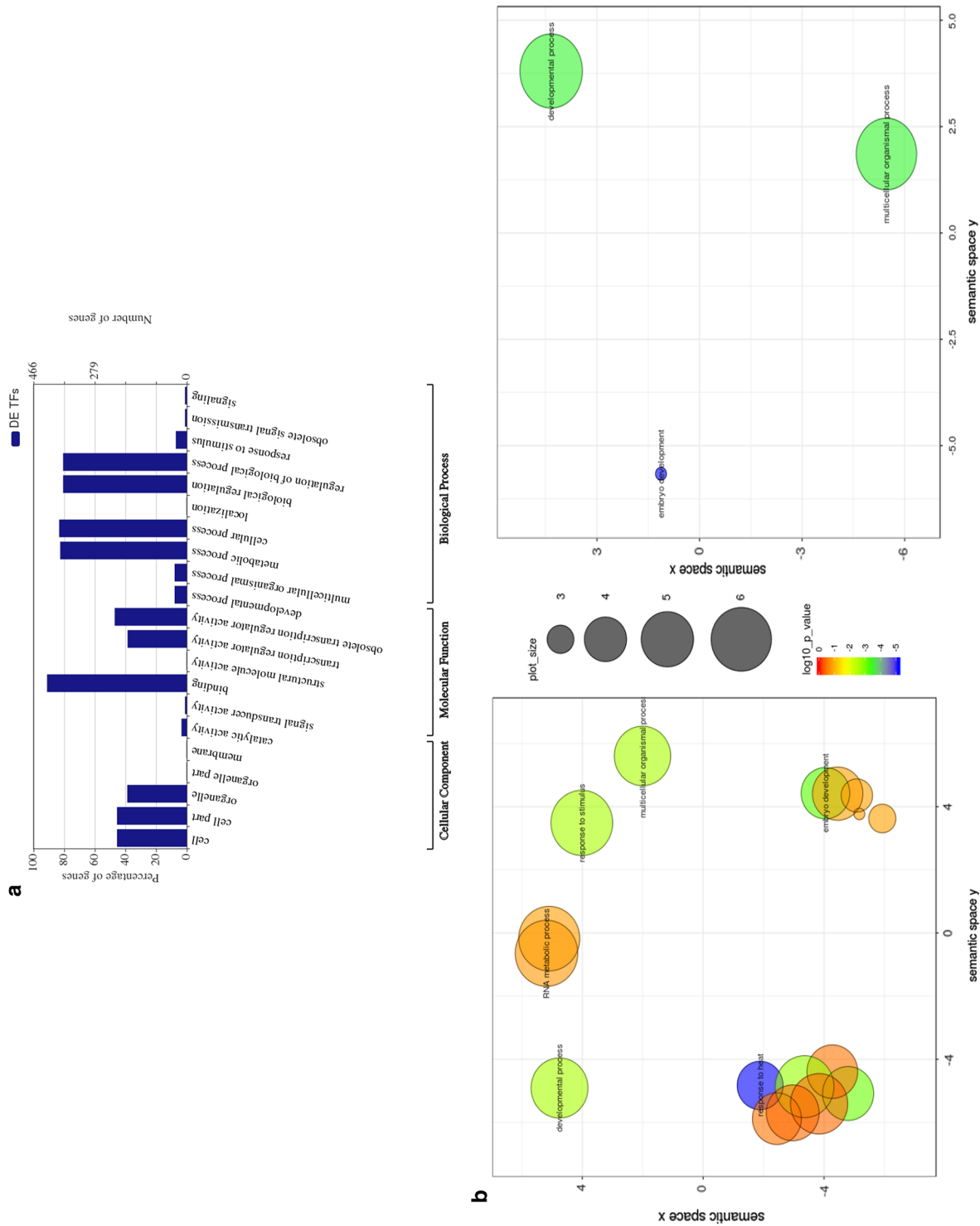


Fig. 3 **a** Gene ontology functional classifications of differentially expressed transcription factors generated by the WEGO tool. **b** Revigo scatter plot of enriched GO terms associated with differentially expressed transcripts between embryo and endosperm for biological process. Bubble color indicates *p* value ($-\log_{10} p$ value) and size indicates the frequency of the GO terms (color figure online)

pathways. Plant hormone signal transduction was the major pathway which contains the largest number of DE TFs. The number of DE TFs in circadian rhythm-plant pathway was relatively less but the richness factor was high in comparison to plant–pathogen interaction pathways (Fig. 4).

Wheat comparative genomics

To develop a more functional understanding of this dataset of 36 oat TFs that were differentially expressed between endosperm and embryo, the closest gene homologs of these oat sequences in the WheatExp database were identified through nucleotide BLAST similarity search (Choulet et al. 2014). Developmental time courses of expression of these wheat genes (Zadoks et al. 1974) in five tissues of wheat (grain, leaf, root, spike, and stem) were then extracted from the WheatExp database. The different assembled oat transcripts belonging to the same

TF family (Table 2) gave the same top gene hits in wheat (except for the group of assembled NAC transcripts that gave diverging hits but with two transcripts, AsTF5891 and AsTF5963, giving the same wheat hits and therefore selected), and usually included at least one hit in each wheat genome. The gene expression was noted for the top three genes of those (i.e., one from each genome), except for the oat transcripts encoding B3 factors which only gave one sequence hit in the wheat database (Fig. 5). TFs belonging to two families, B3 and bHLH, showed higher transcript expression in oat embryo as compared to the endosperm. The closest wheat homologs to these oat transcripts were only expressed in grain tissues, increased during development and showed no expression in other tissues (Fig. 5a, b). The oat transcripts encoding TFs of the five other families instead showed higher expression in oat endosperm as compared to embryo and belonged to the families Dof, MYB, MYB-related, NAC and Trihelix.

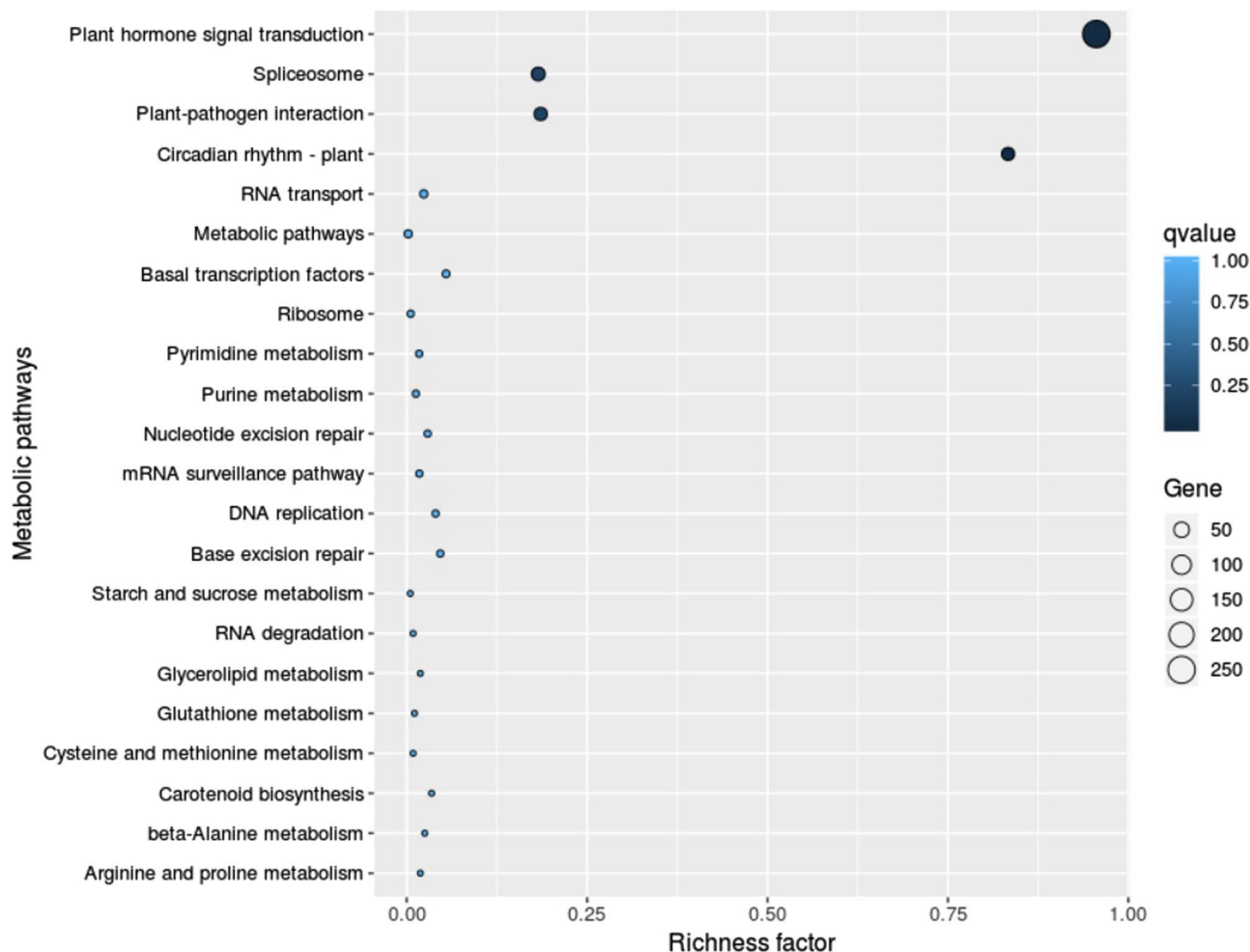


Fig. 4 List of enriched KEGG pathways in differentially expressed transcripts. The richness factor reflects the degree of enriched DGEs in a given pathway. The number of enriched differentially expressed

transcripts in the pathway is indicated by the circle area, and the circle color represents the ranges of the corrected p value (color figure online)

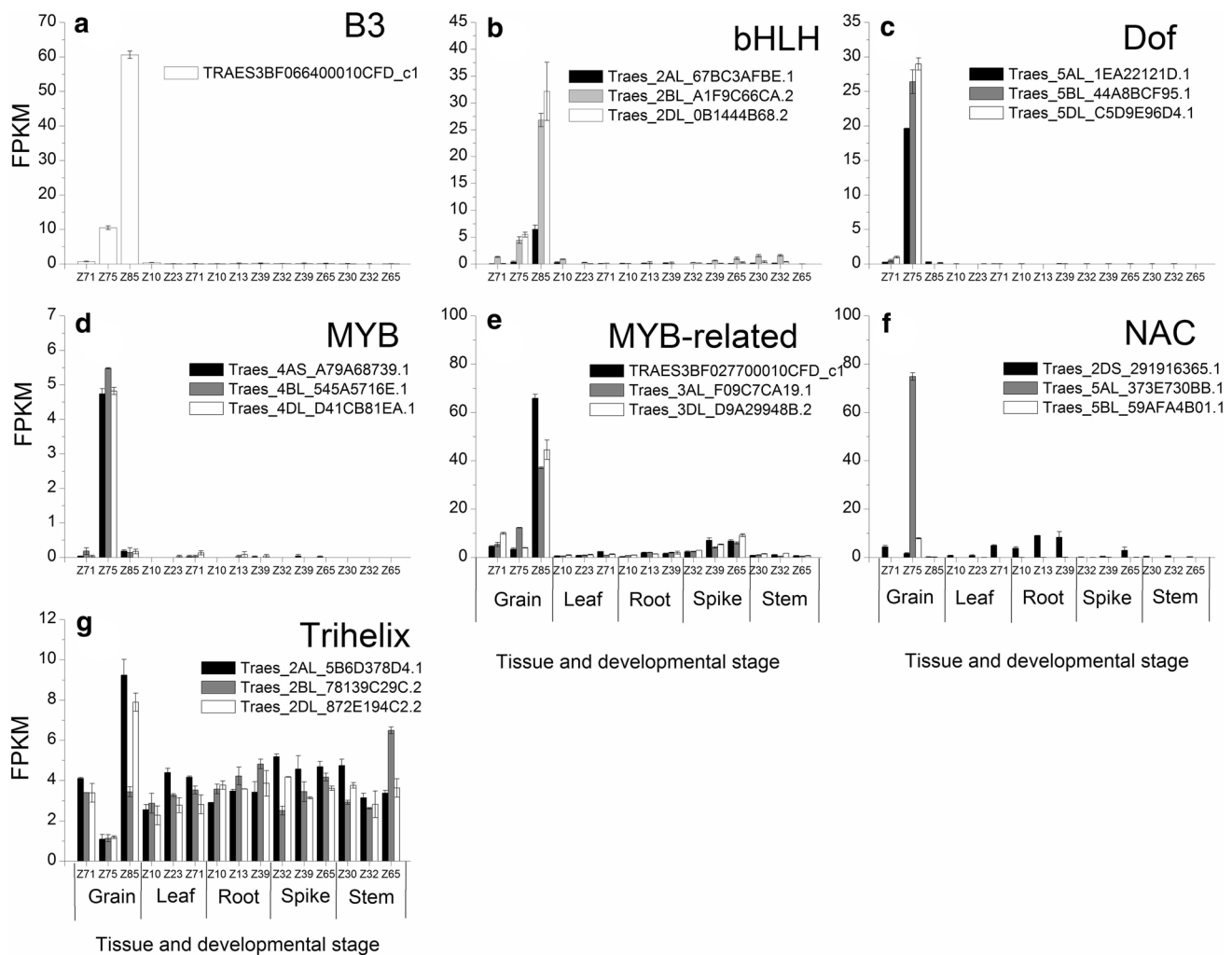


Fig. 5 Gene expression values from WheatExp database for the closest wheat homologs of identified TF oat transcripts listed in Table 2; B3 (a), BHLH (b), Dof (c), MYB (d), MYB-related (e), NAC (f), and Trihelix (g). Expression of closest homologs (one single oat transcript gave one hit from each wheat genome, except for that of B3) is given

in different wheat tissues (grain, leaf, root spike, stem). Z refers to Zadok's developmental scale of wheat (Zadoks et al. 1974). *FPKM* Fragments per Kilobase of exon per Million fragments mapped (color figure online)

The expression of the closest wheat gene homologs to all of these oat transcripts, except for that encoding the Trihelix family, was also highly biased toward grains among different wheat tissues (Fig. 5c–f). Furthermore, the wheat gene homologs encoding Dof, MYB, and NAC factors were expressed in grains at the mid-developmental stage only, while that encoding MYB-related factor was highly expressed at the late developmental stage (i.e., more similar to B3 and bHLH). In contrast to all other TF families in Table 2, the gene expression of wheat homologs encoding Trihelix factor was found in all

tissues of wheat even though the highest level was found in grain at the late stage of development (Fig. 5g).

Verification of differential expressed TFs in oat endosperm and embryo

To experimentally confirm our findings from the designed workflow, we cloned, sequenced and then finally confirmed the differential expression of two selected target transcripts in oat seed endosperm and embryo tissues of *cv. Matilda* at 18 days post anthesis (approximately at mid-stage of seed development). One cloned fragment (779 bp) encoded part

of a bHLH factor (AsTF0641) expected to show higher expression in embryo, and the other fragment (603 bp) encoded part of a Dof factor (AsTF2402) expected to show higher expression in endosperm. Sequences of the cloned fragments (Additional File 3 and 4: Figures S1 and S2, respectively) showed a very high similarity to corresponding contigs, thus confirming proper assembly of the oat transcriptomes. Further, it was clear from semi-quantitative PCR results that the transcript encoding bHLH was indeed higher expressed in embryo as compared to endosperm, while that encoding Dof was instead higher expressed in the endosperm compared to embryo (Fig. 6). It should be noted that the embryo sample contained some endosperm tissue due to the simplified sampling procedure (see “Methods”), contrary to the endosperm which contained no embryo tissue at all. Therefore, the quantification of transcript levels in the different tissues was more contrasting for Dof (which was higher expressed in endosperm) than for bHLH.

Discussion

TFs are involved in the regulation of gene expression by switching on or off whole gene ‘programs’ in specific tissues or during specific developmental stages (Zhang 2003). TF research has usually been focused on the function and evolution of several members of a particular TF family, e.g., expressed under various stress conditions

and investigating molecular mechanisms of responses to specific TFs under various abiotic or biotic conditions (Yanagisawa 2004; Shen et al. 2010; Rahaie et al. 2013). No study has to our knowledge been addressing the systematic identification of seed TFs of the oat grain, i.e., endosperm and embryo. In this study, we identified 3875 TFs in oat grains, belonged to different TF families through transcriptomic characterization and comparative analysis. A TF family level analysis of 3875 identified oat seed TFs (Table 1) showed that the MYB superfamily, ERF, bHLH, bZIP, C3H, NAC families were the six most abundant TF families, whereas ARR-B, G2-Like HB-PHD, HRT-like, LFY, NF-X1, NZZ-SPL, STAT, TALE-family, VOZ, Whirly family and WOX TF families were absent in oat seed transcriptome. These six TF families accounted for 45.8% of the 3875 oat TFs. In several of the classes, it is most likely that the same genetic loci are represented several times from being assembled as individual contigs which is most likely due to the hexaploid nature of oat with three different genomes. Members of these TF families are also abundant in *Arabidopsis*, rice, wheat and maize (Shen et al. 2010; Katiyar et al. 2012; Rahaie et al. 2013).

Thus, it is of high interest to explore what TFs may govern specific features of each tissue development and characteristics. One perspective that can be taken is investigating the differential expression of genes to find factors of importance. This resulted in a set of 514 differentially expressed TFs. Of these, 284 were differentially expressed in embryo while 230 were differentially expressed in endosperm. The number of TFs of differentially expressed TFs was thus quite balanced between embryo and endosperm (Additional File 2, Table S5). However, when scrutinizing different TF families in oat, it was evident that for many there was a dominance for either tissue with B3 (Carbonero et al. 2017), AP2 (El Ouakfaoui et al. 2010), and NF-YB (Zhao et al. 2017) being differentially expressed in embryo while NAC (Borrill et al. 2017), DOF (Hernando-Amado et al. 2012), and C₂H₂ (Royo et al. 2009) were dominated by differential expression in endosperm (Fig. 2b) which is very similar to previous embryo and endosperm studies in different plants (Le et al. 2010; Abraham et al. 2016; Huang et al. 2017; Palovaara et al. 2017).

In gene ontology analysis, embryonic development (GO:0009790), multicellular organismal development (GO:0007275), developmental process (GO:0032502), multicellular organismal process (GO:0032501), and response to heat (GO:0009408) were the top five enriched GO terms among differentially expressed TFs. Further, GO enrichment of DE TFs at tissue level revealed that only embryo has significant (p value < 0.05) GO enrichment terms (Fig. 3b). It may suggest that some regulators might be shared between the embryo and endosperm to complete the development

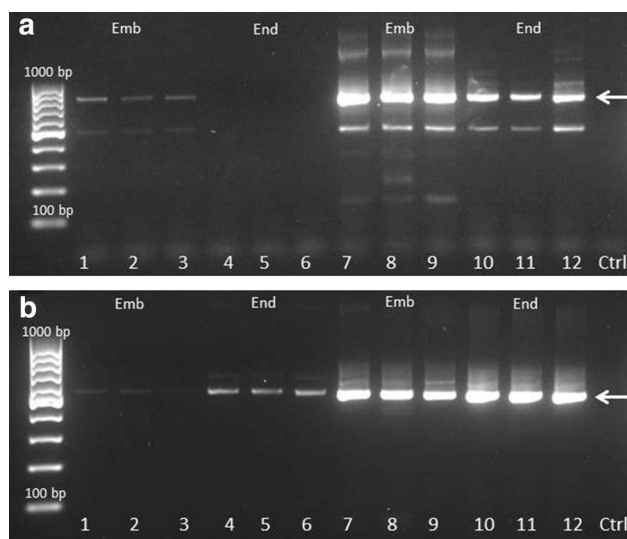


Fig. 6 Semi-quantitative PCR confirming differential expression of transcripts between oat grain endosperm and embryo 18 days post-anthesis. Photo shows results from gel electrophoresis of PCR-amplified transcripts after 25 cycles (lanes 1–6) and 35 cycles (lanes 7–12). Target transcripts (indicated with arrows) were encoding parts of a bHLH factor (**a** 779 bp), and a Dof factor (**b** 603 bp). Embryo tissue (Emb); lanes 1–3 and 7–9, endosperm tissue (End); lanes 4–6 and 10–12, all in three biological replicates each (color figure online)

process at the same time. It also shows that endosperm activity is limited to a number of metabolic processes which could be of importance to save resources such as lipids to provide nutrients to the embryos (Lafon-Placette and Köhler 2014).

In our KEGG enrichment analysis, DE TFs were mainly concerned with plant hormone signal transduction, circadian rhythm-plant, DNA excision repair and replication, plant–pathogen interaction, amino-acid, glutathione and pyrimidine and purine metabolism (Fig. 4). Plant hormone signal transduction and circadian rhythm-plant were shown to have high significance (p value < 0.05) in enrichment analysis. Enrichment of these two pathways among DE TFs could be hypothesized because circadian plant rhythm involves information as light, temperature and nutrient status to synchronize to internal biological rhythms through plant hormone signal transduction pathway with surrounding environments (Salomé et al. 2008; Oracz and Karpiński 2016). DE TFs involvement in plant hormone signal transduction pathways was found for cell enlargement and plant growth, cell division, elongation and shoot initiation, induced germination and stem growth, whereas phytochrome interacting factor 3 (PIF3) was enriched in circadian rhythm-plant metabolism.

To further narrow down the number of targets, BLASTN similarity search was performed against Unigene libraries of wheat (*T. aestivum*) to the data set of 514 differentially expressed TFs. This resulted in a total of 36 transcripts fulfilling all filter criteria applied in the complete workflow (Table 2). To be observed is that, since oat is hexaploid with A, C and D genomes, the number of differentially expressed TFs representing unique genetic loci may be significantly lower than that indicated in our figures. When looking closer at the subset of 36 transcripts it is evident as discussed earlier that they do not represent 36 unique genetic loci but are representations of the three different genomes present in hexaploid oat as well as being allelic variants. B3 and bHLH TFs differentially expressed in embryo most likely represent one gene locus. A similar situation applies to Dof, MYB, MYB_related, NAC and TriHelix, where only the MYB_related and NAC groups are likely to be represented by more than one gene locus. WheatExp database used do not yield an expression resolution on the level of grain sub tissues. Interestingly, BLAST analysis and investigating expression in wheat revealed all but the TriHelix closest homologs to be grain specific. Since the data on oat only reveal expression in grain, it is not certain that these transcripts also are grain specific in oat but rather show that the workflow used can deliver TFs candidates which most likely are grain specific and could be of great importance for seed development. The B3 TF transcript was found most related to VP1 of maize or ABI3 of Arabidopsis. VP1 and ABI3 are well known for their importance in seed and embryo development (Giraudat

et al. 1992; Suzuki et al. 1997). This is thus a good candidate for being the orthologous oat gene.

However, the other TF transcripts identified which closest homologs were grain specific in wheat have not been functionally characterized and thus represent research targets of differentially expressed TFs between endosperm and embryo which if pursued would add further information to the puzzle of factors guiding seed development. This is a finding from the output of the presented workflow (Fig. 1) that points at differences which are likely to be of biological significance and could serve as list of TFs for further research. A final validation of the ability, through the designed workflow, to appropriately filter out TF transcripts was done by cloning, sequencing and performing semi-quantitative PCR of two selected transcripts in cDNA derived from isolated embryo and endosperm RNA. The results confirmed the differential expression between endosperm and embryo of those two transcripts encoding one bHLH and one Dof factor, thus verifying the designed workflow. Further functional characterization of highly differential expressed transcript for enriched biological process will increase our understanding for grain filling and quality.

Conclusions

In the study, we reported a workflow resulting in a list of the 36 most differentially expressed genes encoding TFs between embryo and endosperm of oat which fell into seven TF families, i.e., B3 (10), bHLH (4), Dof (3), MYB (9), MYB_related (4), NAC (4) and TriHelix (2). In our analysis, we found that the closest wheat gene homologs of oat TF transcripts have differential expression between embryo and endosperm, and majority of TF found to be expressed in grain tissues of wheat, thus supporting our findings. Dof, MYB, MYB_related, Trihelix and NAC family TFs were highly expressed in endosperm whereas B3 and bHLH TFs abundant expression were found in embryo tissue. We verified our findings of differentially expressed TFs between endosperm and embryo experimentally by semi-quantitative PCR of two selected target transcripts. This workflow approach can be used for other crops for TF identification and characterization.

Acknowledgements The authors acknowledge support from Uppsala Multidisciplinary Center for Advanced Computational Science for access to the UPPMAX computational infrastructure. We also would like to acknowledge Svetlana Leonova for preparing oat tissues for transcriptome sequencing.

Author contributions SKK conceived the study, performed bioinformatics analysis and lead role in manuscript writing; ÅG performed the wet lab experiments and analysis; PH and AC were involved in study design discussion and provided the overall guidance for this project. All authors wrote and approved the final manuscript.

Funding This work was supported by research funds from Plant Breeding Platform, Swedish University of Agricultural Sciences, SSF (Swedish Foundation for Strategic Research) and the strategic research program Trees and Crops for the Future (TC4F) supported by the Swedish Government.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Availability of data and material Raw RNAseq dataset involved in this study is available in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) database under accession number SRP132328.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abraham Z, Iglesias-Fernández R et al (2016) A developmental switch of gene expression in the barley seed mediated by HvVP1 (Viviparous-1) and HvGAMYB interactions. *Plant Physiol* 170:2146–2158
- Anders S, Reyes A et al (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* 22:2008–2017
- Borrill P, Harrington SA et al (2017) Genome-wide sequence and expression analysis of the NAC transcription factor family in polyploid wheat. *G3 (Bethesda, Md)* 7:3019–3029
- Butt MS, Tahir-Nadeem M et al (2008) Oat: unique among the cereals. *Eur J Nutr* 47:68–79
- Carbonero P, Iglesias-Fernández R et al (2017) The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds. *J Exp Bot* 68:871–880
- Chang Z, Li G et al (2015) Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol* 16:30. <https://doi.org/10.1186/s13059-015-0596-2>
- Charoensawan V, Wilson D et al (2010) Lineage-specific expansion of DNA-binding transcription factor families. *Trends Genet* 26:388–393
- Chauhan P, Hansson B et al (2014) De novo transcriptome of *Ischnura elegans* provides insights into sensory biology, colour and vision genes. *BMC Genom* 15:808. <https://doi.org/10.1186/1471-2164-15-808>
- Chen H, Qiu S et al (2016) New insights into the antioxidant activity and components in crude oat oil and soybean oil. *J Food Sci Technol* 53:808–815
- Choulet F, Alberti A et al (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science* 345(6194):1249721. <https://doi.org/10.1126/science.1249721>
- Doebley JF, Gaut BS et al (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321
- Du Z, Zhou X et al (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38:W64–W70. <https://doi.org/10.1093/nar/gkq310>
- Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- El Ouakfaoui S, Schnell J et al (2010) Control of somatic embryogenesis and embryo development by AP2 transcription factors. *Plant Mol Biol* 74:313–326. <https://doi.org/10.1007/s11103-010-9674-8>
- Giraudat J, Hauge BM et al (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* 4:1251–1261
- Grimberg Å, Carlsson AS et al (2015) Transcriptional transitions in *Nicotiana benthamiana* leaves upon induction of oil synthesis by WRINKLED1 homologs from diverse species and tissues. *BMC Plant Biol* 15:192. <https://doi.org/10.1186/s12870-015-0579-1>
- Gutierrez-Gonzalez JJ, Tu ZJ et al (2013) Analysis and annotation of the hexaploid oat seed transcriptome. *BMC Genom* 14:471. <https://doi.org/10.1186/1471-2164-14-471>
- Haas BJ, Papanicolaou A et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512
- Hahn DA, Ragland GJ et al (2009) Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly *Sarcophaga crassipalpis*. *BMC Genom* 10:234. <https://doi.org/10.1186/1471-2164-10-234>
- Hernando-Amado S, González-Calle V et al (2012) The family of DOF transcription factors in *Brachypodium distachyon*: phylogenetic comparison with rice and barley DOFs and expression profiling. *BMC Plant Biol* 12:202
- Huang Y, Niu B et al (2010) CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680–682
- Huang J, Deng J et al (2017) Global transcriptome analysis and identification of genes involved in nutrients accumulation during seed development of rice tartary buckwheat (*Fagopyrum tararicum*). *Sci Rep* 7:11792
- Iida K, Seki M et al (2005) RARTF: database and tools for complete sets of arabidopsis transcription factors. *DNA Res* 12:247–256
- Jung C, Müller AE (2009) Flowering time control and applications in plant breeding. *Trends Plant Sci* 14:563–573
- Katiyar A, Smita S et al (2012) Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genom* 13:544
- Kriventseva EV, Zdobnov EM et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Lafon-Placette C, Köhler C (2014) Embryo and endosperm, partners in seed development. *Curr Opin Plant Biol* 17:64–69. <https://doi.org/10.1016/j.pbi.2013.11.008>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359
- Le BH, Cheng C et al (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci USA* 107:8063–8070
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform* 12:323. <https://doi.org/10.1186/1471-2105-12-323>
- Ma X, Ma J et al (2016) Genome-wide identification of TCP family transcription factors from *Populus euphratica* and their involvement in leaf shape regulation. *Sci Rep* 6:32795. <https://doi.org/10.1038/srep32795>
- Mao X, Cai T et al (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21:3787–3793
- Marguerat S, Bähler J (2010) RNA-seq: from technology to biology. *Cell Mol Life Sci* 67:569–579
- Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nat Rev Genet* 12:671–682
- Metzker ML (2010) Sequencing technologies—the next generation. *Nat Rev Genet* 11:31–46
- Oracz K, Karpiński S (2016) Phytohormones signaling pathways and ros involvement in seed germination. *Front Plant Sci* 7:864

- Palovaara J, Saiga S et al (2017) Transcriptome dynamics revealed by a gene expression atlas of the early *Arabidopsis* embryo. *Nat Plants* 3:894–904
- Peng Y, Leung HCM et al (2013) IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics* 29:i326–i334
- Rahaie M, Xue G-P et al (2013) The role of transcription factors in wheat under different abiotic stresses. In: Vahdati K, Leslie C (eds) *Abiotic stress—plant responses and applications in agriculture*. InTech, Rijeka
- Rasane P, Jha A et al (2015) Nutritional advantages of oats and opportunities for its processing as value added foods—a review. *J Food Sci Technol* 52:662–675
- Robinson MD, McCarthy DJ et al (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140
- Royo J, Gómez E et al (2009) Transcriptional activation of the maize endosperm transfer cell-specific gene *BETL1* by *ZmMRP-1* is enhanced by two C2H2 zinc finger-containing proteins. *Planta* 230:807–818
- Salomé PA, Xie Q et al (2008) Circadian timekeeping during early *Arabidopsis* development. *Plant Physiol* 147:1110–1125
- Shen B, Allen WB et al (2010) Expression of *ZmLEC1* and *ZmWRI1* increases seed oil production in maize. *Plant Physiol* 153:980–987
- Supek F, Bošnjak M et al (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Suzuki M, Kao CY et al (1997) The conserved B3 domain of *VIVIP-AROUS1* has a cooperative DNA binding activity. *Plant Cell* 9:799–807
- Thiriet-Rupert S, Carrier G et al (2016) Transcription factors in microalgae: genome-wide prediction and comparative analysis. *BMC Genom* 17:282. <https://doi.org/10.1186/s12864-016-2610-9>
- Van Belleghem SM, Roelofs D et al (2012) De novo transcriptome assembly and SNP discovery in the wing polymorphic salt marsh beetle *Pogonus chalceus* (Coleoptera, Carabidae). *PLoS One* 7:e42605. <https://doi.org/10.1371/journal.pone.0042605>
- Webb A, Cottage A et al (2016) A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (*Vicia faba* L.). *Plant Biotechnol J* 14:177–185
- Wickramasuriya AM, Dunwell JM (2015) Global scale transcriptome analysis of *Arabidopsis* embryogenesis in vitro. *BMC Genom* 16:301
- Xie Y, Wu G et al (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30:1660–1666
- Yanagisawa S (2004) Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant Cell Physiol* 45:386–391. <https://doi.org/10.1093/pcp/pch055>
- Zadoks JC, Chang TT et al (1974) A decimal code for the growth stages of cereals. *Weed Res* 14:415–421
- Zhang JZ (2003) Overexpression analysis of plant transcription factors. *Curr Opin Plant Biol* 6:430–440
- Zhao H, Wu D et al (2017) The *Arabidopsis thaliana* nuclear factor Y transcription factors. *Front Plant Sci* 7:2045

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.