# CHAPTER 4

## ISOLATION, CHARACTERIZATION AND EXPRESSION OF GA GENES WITH PARTICULAR EMPHASIS ON *GA20ox* IN TEF (*Eragrostis tef*)

## 4.1 Abstract

To isolate *GA20ox* and *Rht* genes and further characterize and monitor the expression of *GA20ox* in *E. tef*, degenerated primers derived from rice, wheat, maize and sorghum ortholog sequences were used of full or partial sequences (422 to 1500 bp) by normal, "anchored" and nested PCR. Three putative *GA20ox* genes orthologous to the rice *sd-1* (*semi-dwarf* 1) were identified. Sequences contained the characteristic domains KLPWKET and NYYPXCXXP and the conserved H and D amino acid residues. Three more sequence orthologous to the wheat *Rht* gene, the rice *Elongated Uppermost Internode* (*Eui*) and a Cytochrome P450 monooxygenase gene involved in brassinosteroid (BR) deactivation were also isolated. The *E. tef Rht* sequence had the conserved motifs DELLA, VHYNP, VHVVD, and a C-terminus of the GRAS domain. *E. coli* expressed *GA20ox1a* catalyzed the conversion of the [$^{14}$C]-labelled gibberelline precursor $GA_{12}$/ to $GA_9$. The three *GA20ox* were further differentially expressed in different plant tissues. *EtGA20ox1a* and *EtGA20ox1b* had highest transcription in the uppermost internodes whereas *EtGA20ox2* transcription was low in most tissues. Further, in *E. tef EtGA20ox1b* was possibly the functional equivalent to the rice *sd-1* gene. All sequences showed close homology with ortholog genes from sorghum, maize, rice barley.

## 4.2 Introduction

Among the genes of GA-biosynthesis, GA 20-oxidase is shown to be an important regulator in the pathway. It catalyses several late steps converting $GA_{12}$ and $GA_{53}$ in parallel pathways to respective products $GA_9$ and $GA_{12}$ which then are converted to bioactive forms, GA4 and $GA_1$ by GA 3β-hydroxylase (see chapter 1 "Introduction"). Availability of genes encoding GA-biosynthetic enzymes has allowed understanding how these genes function. In many species the members of the dioxygenases family, GA 20-oxidases, are encoded by several genes. They show distinct spatial and temporal expression patterns with some overlapping function in plant development regulated by environmental signals and endogenous factors (Hedden and Phillips 2000; Lange, 1998; Hedden and Kamiya, 1997). Such functional redundancy in the GA 20-oxidases has been found for many crops and has been considered as the reason why null mutations in some of the genes exhibit a semi-dwarf phenotype. The existence of an overlap between expression patterns of isozymes and mobility of the GA products within a plant system prohibits severe dwarfing of the stem (Hedden and Phillip, 2000). Due to their regulatory role determining GA concentrations, members of this gene family have been the target for plant genetic manipulation and introduction of agronomically useful traits (Appleford *et al*., 2006; Carrera *et al*. 2000; Sakamoto *et al*. 2003).

Genomic DNA amplification by PCR for isolation of gene sequences is a common procedure and has also been used in this part of the study to isolate GA biosynthesis and signalling genes. A genomic survey at the NCBI database (http://www.ncbi.nlm.nih.gov), as done for this study, is also a vital tool to provide information for the design of primers for gene amplification in PCR application for reverse transcription (e.g. RACE). The Rapid Amplification of cDNA Ends (RACE) is a method for amplifying DNA sequences from a

mRNA templates between a defined or identified internal site and unknown sequences at either to the 3' or the 5' -end of the mRNA. It requires two gene sequence-specific primers that flank the region of sequence to amplify relatively few target molecules in a complex mixture (Miao *et al*., 2010). Therefore, 3' and 5' RACE methodologies offer possible solutions to the severe limitation impose on the PCR to amplify regions of unknown sequences. The 5' RACE, also known as “anchored” PCR, is a technique that facilitates the isolation of unknown 5' ends from low-copy transcriptions. Low level of concentration is expected for phytohormones and their corresponding related genes such as *GA20ox* that involve in the, GA, regulation (Hedden and Phillip, 2000).

No GA gene so far has been isolated from *E. tef* to allow studying the allelic diversity of GA genes in the genetic pool, explore their functional roles to identify a useful mutation in GA genes. Having verified the importance of GA in tef height regulation, therefore, the objectives of this part of the study was to first identify and clone full-length cDNAs from various GA genes in particular encoding the multifunctional GA20-oxidase. Secondly, the objective was to characterize these cloned GA gene sequences on the genomic level and study their expression in various plant parts during plant development.

## 4.3 Materials and Methods

### 4.3.1 Plant material and plant growth

Seed material of tef (*Eragrostis tef*) variety DZ-01-196 for this experiment was obtained from the Ethiopian Institute of Agricultural Research, Holetta Agricultural Research Centre, Ethiopia. Seeds were germinated on germination mix soil and the seedlings were grown in pots (about 225 plants per $m^2$) under a 26±2 / 18°C day/night temperature and a 14 hr day length. Plants were further supplemented with a half-strength Hoagland nutrient solution until all samples were collected.

### 4.3.2 Genomic DNA isolation

Young leaves were homogenized in liquid nitrogen and transferred to a reaction tube containing pre-heated ($60^0$C) extraction buffer according to the method of Harini *et al.* (2008) with a modification omitting the NaCl addition. DNA spooling was further replaced by precipitating the DNA. Samples not immediately used were frozen and stored at -$80^0$C. The mixture of fine-powdered plant material was immediately incubated after homogenization in extraction buffer at $60^0$C for 30 min with intermittent shaking. This was followed by adding an equal volume of chloroform: isoamyl alcohol (24:1) and subsequent centrifugation at 6000 g for 10 min. The aqueous phase was further extracted with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) and centrifuged at 8000 g for 10 min. After centrifugation, the supernatant was transferred into 2.5 volumes of chilled ethanol and genomic DNA was precipitated. Further purification was made using isopropanol which was followed by centrifugation at 8000 g for 10 min at $4^o$C to collect precipitated DNA. Washing

of precipitated DNA was carried out with 70% ethanol twice followed by centrifugation. The genomic DNA was finally dissolved either in TE buffer (10 mM Tris-HC1, 1 mM EDTA, pH 8.0) or ds$H_2O$ and further treated with RNase and precipitated again as outlined above. DNA concentration of samples was determined using the nanodrop technique (Thermo Scientific NanoDrop 3300) and then visually by running 2 µl of sample DNA on a 1% agarose gel to check for purity.

#### 4.3.3 Gene identification and isolation

Identification and cloning of GA biosynthetic genes, the rice *SD-1* homologues (*GA20-oxidases*) in *E. tef* was carried out using degenerated oligonucleotide (sense and antisense) primers (Table 3.1). Primers for specific amplification of the genes sequences were designed based on the published sequences of the *GA20ox* genes of different monocot plants accessed from the GenBank. Primers were designed based on conserved amino acid domains of rice, sorghum, maize and wheat orthologous sequences obtained from the GenBank database. The sequences of these nucleotides were designed using the primer3 program (Rozen and Skaletsky, 2000). The PCR conditions were optimized using these primers to amplify either genomic DNA or cDNA.

**Table 4.1** Primers used in PCR to amplify *GA20ox* gene fragments using *E. tef* genomic DNA

| Target gene | Primers | Degenerate/specific primers (5' – 3') |
|---|---|---|
| GA20 ox | F1 | GCTGCCGTGGAAGGAGAC |
| | F2 | CACCGATGATGATGATGATGATG |
| | F3 | CTACGCGAGCAGCTTCACG |
| | R1 | CGCCGATGTTGACGACGA |

Primers for *GA20ox* were targeted to the region encoding the amino acid sequence KLPWKET (sense) and NYYPXCXXP (antisense). PCR amplification of *E. tef* genomic DNA (gDNA) and complementary DNA (cDNA) was done using either Dream Taq (Fermentas, Canada) or FastStart Taq (Roche, UK) DNA-polymerases. A 50 µl reaction mixture consisted of the following components: 5x PCR buffer, 10 µl $MgCl_2$ (25 mM), 2 µl dNTPs (10 mM each); 1 µl each for forward and reverse primers (10 µM), 1 µl DNA (100 ng/µl); 1 µl DMSO; 0.25 µl Dream Taq DNA-polymerase (5 U/µl); 33.75 µl $H_2O$. PCR reactions were prepared on ice and run at the following standard cycling conditions: 94°C (2 min) followed by 35 cycles consisting of 94°C (30 sec); 55°C-65°C (30 sec); 72°C (1 min/kb). This was followed by DNA extension at 72°C for 7 min and then holding the reaction at 4°C.

When amplification of a high-fidelity PCR component was required, the "Phusion" polymerase proof reading enzyme was used (Finnzymes). The PCR reaction contained the following: 10 µl 5x Phusion "GC" buffer; 1 µl dNTPs (10 mM each); 1 µl each for forward

primer (10 μM); 1μl DNA (100 ng/μl); 1 μl DMSO; 0.5 μl Phusion polymerase (2 U/μl); 34 μl $H_2O$. Preparation of the PCR reactions were carried out on ice with reactions running at the following standard cycling conditions: 98°C (30 sec) followed by 35 cycles with 98°C (10 sec); 55°C-70°C (30 sec); 72°C (1 min/1kb) and DNA extension after 35 cycles at 72°C for 7 min followed by holding the reaction at 4°C. Genomic DNA amplifications were done using GC-rich buffer and addition of DMSO (molecular grade) in reactions to improve amplification efficiency. PCR products were visualized on a 1% agarose gel. For cloning, bands were excised from the gel and purified with a commercially available PCR product purification kit (Qiagen, Germany). The extracted DNA was used as a template for cloning, or nested PCR using the conditions described above. Cloning was carried out using the pGEM-T Easy cloning vector (Invitrogen, USA).

#### 4.3.4 Isolation of complete *E. tef GA20ox* coding regions

For extension of *GA20ox* homologous (GA20 ox1-ox3) to their unknown 5' and 3', a total of 20 primers either as specific primers SPR1, SPR2 or SPR3 (nested primers for upstream extension in 5' direction) and SPF1 and SPF2 (nested primers for downstream extension in the 3' direction) were designed and used for nested PCR amplification (Table 5.2). cDNA synthesis was done using anchor primers (supplied with cDNA synthesis kit; Invitrogen, USA) and manually designed specific primers (SP primers) (Table 5.2). Primers were allowed to anneal identified polymorphic regions to also enable selective amplifications of homologous upstream and downstream regions of alleles using reverse PCR (RACE-PCR). cDNA synthesis. PCR amplification procedures were done using the $2^{nd}$ Generation 5'/3' RACE Kit (Roche, Switzerland) following the manufacturers recommendations for full-length cDNA synthesis of genes of interest.

Primers for isolation of a full-length sequence of GA20 ox1 (1420 bp), but containing also the start and stop codon and restriction sites BamHI and HindIII (underlined), were designed as follows, sense: 5'-AGG GAT CCA GCC AGC TGC CCG TGA TG-3' and antisense: 5'-TGA AGC TT**A** ACA GAA CAG GCG GTC ATG GAT GAC-3'. A nucleotide ("A") was added to this primer to ensure that the start codon (ATG) in the GA20 ox1 is "in-frame" during after cloning into the plasmid pET-32a(+). In order to avoid error incorporated during the PCR-amplification, high fidelity "Phusion" polymerase proof reading enzyme (Finnzymes) was used and DNA sequences were determined from at least 3 -5 independent clones of the amplified product.

**Table 4.2** Primers used for GA20ox sequence RACE- PCR amplification

| Target gene | Primers | Degenerate/specific primers (5' – 3') |
|---|---|---|
| GA20 ox1 | GA20 ox1SPR3 | GCGGCAGCGTGAAGAAGGCGTCCAT |
| | GA20 ox1SPR2 | CACATCATCATCATCATCATCGGT |
| | GA20 ox1SPR1 | CTCGGCGGGTAGTAGTTGAGGCGCAT |
| | GA20 ox1SPF1 | CCTTCGTCGTCAACATCGGCG |
| | GA20 ox1SPF2 | CGGAGACAACCAAAGGAGGCG |
| GA20 ox3 | GA20 ox3SPR3 | GTCTCCTTCCACGGCAGCAATCG |
| | GA20 ox3SPR2 | CGATTGGCAGTAGTCTCGGAACA |
| | GA20 ox3SPR1 | CGGGCACGGCGGGTAGTAGTTG |
| | GA20 ox3SPF1 | GCAGGGACTTCTTCGCCGACG |
| | GA20 ox3SPF2 | CGGGAGCCATCGTCGTCAACATC |

### 4.3.5 Cloning and sequencing of PCR products

Cloning of the purified PCR products was done using the pGEM-T Easy vector system (Promega, UK). Insert and plasmid DNA (pGEM-T Easy) was mixed at a 3:1 ratio in a 10 µl ligation mixture with 1 µl T4 DNA ligase (5 U/µl) and 1x T4 DNA ligase buffer (Invitrogen, USA). The mixture was incubated at room temperature overnight. Competent Top10 or DH5α *E. coli* cells were transformed with DNA fragments legated into the pGEM T-easy plasmid by heat shock treatment at 42°C for 45 sec followed by 2 min incubation of transformed cells on ice. *E. coli* cells were then incubated for 1-1.5 hr at 37°C with rotation

(150 rpm) on a shaker after adding 250 µl SOC media (20 g/l bacto-tryptone, 5 g/l bacto-yeast extract, 0.5 g/l NaCl, 20 mM glucose, 2.5 mM KCl, 10 mM $MgCl_2$, pH7.0). Transformed cells (200 µl) were then plated onto 2YT agar plates (16 g/l bacto-tryptone, 10 g/l bacto-yeast extract, 5 g/l NaCl, 1.5% (w/v) agar, pH 7.0) supplemented with 0.5 mM IPTG, 0.01 mg/ml X-Gal and 0.1 mg/ml ampicillin allowing blue-white selection of transformed colonies.

Five white colonies were cultured overnight in 5 ml 2YT media supplemented with 0.1 mg/ml ampicillin. Plasmid DNA was purified from cells using the "QIAprep plasmid mini kit" (Qiagen, Germany) according to the manufacturer's instructions. Isolated DNA was eluted in 50 µl buffer EB (Qiagen, Germany). Positive cloned inserts were confirmed by *Eco*RI digestion followed by agarose gel electrophoresis. Confirmed DNA samples were further run for sequencing in a PCR reaction containing 2 µl BigDye Terminator 3.1, 2 µl Sequencing Buffer (5X), 1 µl M13 or custom-designed sequencing forward or reverse primers, 2 µl $sdH_2O$ and 3 µl template DNA. The PCR reaction product was then purified using Sephadex G-50 Fine Grade Slurry (Sigma, UK) and Centri-Sep columns before sequencing.

4.3.6 DNA sequence analysis and phylogenetic analysis

DNA sequences were analyzed using the BLAST program (http://www.ncbi.nlm.nih.gov /blast/) by comparing with sequences in the GenBank database. Candidate nucleotide sequences were translated using the ExPASY Translation Tool (http://au.expasy. org/ tools/dna. html) before BLAST search for determining homology of deduced amino acid sequences of closely related plant species. Phylogenetic analysis was computed using all the

amino acid sequences of related genes from the different species using CLC Bio Mainbench (Version 5.5)/ MEGA 4.1 (beta).

### 4.3.7 RNA isolation and cDNA synthesis

The frozen immature leaf sample of DZ-01-196 was homogenized in liquid nitrogen followed by total RNA extraction using the RNeasy kit (Qiagen, Germany) following the instructions by the manufacturer. Up to 2 µg of total RNA and oligo dT (18 - mer) was used for cDNA synthesis following first-strand cDNA which was reverse transcribed by using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, UK) according to the manufacturer's instruction.

### 4.3.8 Isolation of the *Rht* and other genes

Using an identical PCR approaches stated above (Section 4.2.3 to 4.2.7) were employed for designing primers, running PCR, cloning and sequencing of PCR products for the isolation of from *E. tef* homologous sequences to the wheat Rht gene and other genes (the rice Elongated Uppermost Internode (EUI) and brassinosteroid deactivation (BR)). Primers used are shown in Table A1 and A2 (See Appendix).

### 4.3.9 GA 20-oxidase expression in *E. tef*

The expression of the three *E. tef* GA20 oxidases in different *E. tef* tissues at different developmental stages was determined by a quantitative real time PCR (qRT-PCR) using purified total RNA from different tissues. The qRT-PCR analysis was done using the

LightCycler technique for quantitative reverse transcription (RT)-PCR of the mRNA levels of the gene of interest using SYBR Green (fluorophore that binds double-stranded DNA) to produce fluorescence for detection. Various internal control primers were designed (Table 4.1) based on constitutively expressed house-keeping tef Actin and 25S rRNA gene to optimize qRT-PCR measurements. The *E. tef* 25S rRNA gene was selected and amplification was optimized based on the LightCycler 480 (Roche Diagnostic, UK) result. The crossing point value (CP) was generated representing the fractional cycle number at which the amount of amplified target reaches a fixed threshold. The following protocol was used to determine the expression ratio: the $C_t$ value difference between each triplicate was kept below 1.5 and outliers with SD > 2 were removed from the analysis and the mean $C_t$ and standard deviation (SD) for each sample were determined. In the first step of analysis ($\Delta C_t$), all samples values were normalized with respect to the least expressed sample. The highest $C_t$ value was subtracted from the $C_t$ value of each sample where the least expressed sample will have a $\Delta C_t$ value of 0. The difference of samples from the reference value was computed using the function ($\Delta C_{t\ sample}$ - $\Delta C_{t\ reference}$) to obtain a $\Delta\Delta C_t$ value used to calculate $2^{-\Delta\Delta Ct}$ along with calculating the means for each group, the standard deviation and standard error of the mean (SEM). This provides a relative expression ratio (in arbitrary units) and variation between ratios.

Expression of GA20ox sequences from *E. tef* was further analyzed using total RNA extracted from different *E. tef* tissues at the same time i.e. at rapid stem elongation stage (10-15 days before panicle emergence), except the young panicle (inflorescence), which was sampled three weeks later. The RNase Plant Mini Prep kit (Qiagen, Germany) was used for RNA isolation according to the manufacturer's guide followed by DNAase treatment. RNA (2.0 µg) was converted into cDNA using the Superscript III First-Strand Synthesis System for RT-

PCR (Invitrogen, UK). The reaction was stopped by heat inactivation at 94 °C for 5 min. PCR amplification of tef GA20ox cDNA using 0.4 mM of the specific primers (Table 5.1) that have been designed on the basis of polymorphic regions of the three cloned homologous sequences of the EtGA 20-oxidase gene. PCR DNA amplification was carried out at 94 °C for 1 min followed by a 32 cycles for GA20ox1 and GA20ox1b and a 40 cycle for GA20ox2 of 94 °C for 1 min, 67 °C for 30 sec and 72 °C for 30 sec. The length of the expected amplification fragment ranged between 80-190 bp. Tef SDH gene was used as a control.

**Table 4.3** Primers used in quantitative and semi-quantitative PCR amplification of GA20ox sequences

| Target gene | Primer | Primer sequence (5' – 3') |
|---|---|---|
| GA20 ox1 | GA20 ox1 02F | ATGTGGTGGGCTACTACGTCAGCAAG |
| | GA20 ox1 01R | TCATCTCCGAGCAGTAGCGCCCGT |
| GA20 ox2 | GA20 ox2 02F | CGGCCCACACCCTCTTGCTCCA |
| | GA20 ox2 02R | TTGACGACGATGGCTCCCGGCTT |
| GA20 ox3 01F | GA20 ox3 02F | GGACTACCTGGTGGGCCGC |
| | GA20 ox3 01R | TAGTAGTTGAGCCGCATGATGGAGTCG |
| SDH | SDH 01F | CACAGCTGAGCGCTACGTTCTC |
| | SDH 01R | CCCAATGCAACACCGAAAATACG |
| 25S | 25S 01F | ATAGGGGCGAAAGACTAATCGAACC |
| | 25S 01R | GAACTCGTAATGGGCTCCAGCTATC |

### 4.3.10 Expression of *EtGA20ox1a* in *E. coli*

The activity of EtGA20 ox1 was monitored using a gel-purified PCR product (1.2 kb) that was amplified using primers sense: 5'-AGG GAT CCA GCC AGC TGC CCG TGA TG-3' and antisense: 5'-TGA AGC TTA ACA GAA CAG GCG GTC ATG GAT GAC-3' from a cDNA template. The PCR amplification was done using "Phusion" polymerase proof reading enzyme (Finnzymes) and the blunt-end PCR product was gel-purified and cloned into the plasmid pET32a vector (Novagen, UK) using the BamHI and HindIII (underlined) restriction sites. Competent *E. coli* DH5-α cells were transformed with the *E. tef* GA20ox-1 containing plasmid. From transformed cells, recombinant plasmid DNA was isolated and *E. coli* BL21 cells were transformed with the pET32-a fusion by heat shock treatment. A transformed colony was selected and cells were cultured overnight in 5ml 2YT containing 0.1 mg/ml carbenicillin for selection. A cell suspension (500 µl) was then added to 50ml 2YT containing 0.1 mg/ml carbenicillin and incubated at 37°C for 2 hr and then 50 µl of 1M IPTG was added to the cell suspension. Cells were then cultivated at 25°C for 6 hr under shaking (200 rpm). The cell suspension was then centrifuged for 5 min and the cell pellet was frozen at -20°C overnight.

### 4.3.11 HPLC analysis

Defrosted cells were re-suspended in 1.5 ml lysis buffer containing in 10 ml 1 ml 1 M Tris-HCl, pH 7.5, 50 µl of 1M DTT, 200 µl of 50 mg/µl lysozyme, 8.75 ml sd$H_2O$. The cell suspension (1.5 ml) was incubated for 15 min at room temperature, treated with DNase and cell debris were removed by centrifugation for 10 min in an Eppendorf centrifuge at 13 000 g at 4°C and the cell lysate was kept for at -80°C. The lysate was thawed and the following

added to 90 µl supernatant: 5 µl dioxygenase co-factor mix (containing in 5 ml 80 mM 2-oxoglutarate, 80 mM ascorbate, 80 mM DTT, 10 mM $FeSO_4$, 40 mg mL21 BSA, and 20 mg mL21 catalase in 100 mM Tris-HCl, pH 7.5; Williams *et al*., 1998) and 5 µl substrate (GA12-$^{14}C$). The mixture was incubated for 2 hr at $30^0C$ under shaking at 200rpm. Glacial acetic acid (10 µl) was added to the mixture and further diluted with 140 µl $sdH_2O$ and centrifuged for 10 min at maximal speed in an Eppendorf centrifuge before HPLC analysis to measure the GA intermediate products catalyzed by this protein (GA20 ox1).

### 4.3.12 Southern blot analysis

A DNA hybridization probe of 450 bp was synthesized using specific primers for a coding region of GA20 ox and a DIG labelling probe synthesis kit (Roche, Switzerland). Purified genomic DNA (15 µg) was digested overnight with the restriction enzymes BamHI, BglII, EcoRI and HindIII. Digested genomic DNA was fractionated on a 0.7% agarose gel and transferred by the alkaline transfer technique (Roche, UK) to a nylon membrane (Hybond-N+, Amersham Pharmacia Biotech, UK). Transferred DNA was hybridized with a DIG-labelled probe in an oven at $42^0C$ overnight. Membranes with hybridized DNA were incubated in a color substrate solution in the dark without shaking for at least 30 min for detection of hybridized DNA bands.

## 4.4 Results

### 4.4.1 Isolation of GA genes from *E. tef*

Identification and cloning of a *E. tef* ortholog of the rice *SD-1* (*SEMI-DWARF* 1) gene was carried out using a total of 5 degenerate oligonucleotides that were designed based on orthologous sequences of the conserved amino acid regions from rice, sorghum, maize, wheat and barley. Sequences were retrieved from the GenBank database using accession numbers of known gene sequences from rice and wheat using BLAST search (NCBI). A partial sequence closely related to the 2-oxoglutarate dependent oxygenase gene family was first obtained by PCR amplification from *E. tef* DNA, which shares conserved domains of GA biosynthetic pathway genes. However, analysis revealed that this sequence is lacking the important domains that are characteristic of the 2-oxoglutarate dependent dioxygenase (2ODD) genes (GA20ox, GA3ox and GA2ox) involved in the GA biosynthetic pathway (Hedden and Kamiya, 1997). When primers F1, F2, F3 and R1 were used, three PCR fragments (*E. tef*1. *E. tef*2. *E.tef*3) were obtained with sizes between 422-523 (Figs. A.1A and B; Fig. A.2C, D, E and F see Appendix). After cloning of fragments into the plasmid pGEM-T Easy and sequencing the nucleotide sequence of these fragments showed a >60% homology with known *GA20ox* orthologous sequences from closely related cereal species such as sorghum, maize, rice, wheat and barley.

Using an identical PCR approaches GA signaling gene from *E. tef* homologous to the wheat REDUCED HEIGHT (RHT) gene was isolated. Additional gene sequences isolated were the rice Elongated Uppermost Internode (EUI) and the Cytochrome P450 monooxygenase gene involved in brassinosteroid (BR) deactivation both involved in plant height development. Primer pairs RhtF1/ F2, RhtR1/R2 and Rht SPF and RhtSPR for Rht gene; EUIF1 and EUIR1 for Eui gene

and BRF1 and BRR1 for BR gene (Tables A.1 and A.2, see Appendix) allowed PCR amplification of the following fragments: (1) 1395 bp (near full coding region of *Rht* gene) that shares over 85% homology with the above species (Figs. 5.9, 5.10 and A.6 (see Appendix)), (2) 344 bp (*Eui* gene) which shares over 60% homology with *EUI* orthologous sequences from sorghum, wheat, maize, brachypodium and rice (Figs. A.3 and A.4 and A.5, see Appendix); and (3) 749 bp (*BR* gene), another Cytochrome P450 monooxygenase gene but with brassinosteroid deactivation activity was partially cloned using semi-degenerate oligonucleotides (Figs. A.7K and L; A.8 and A.9, see Appendix). The peptide sequence shared 49 - 52% homology with orthologous peptide sequences of genes involved in BRs catabolism.

4.4.2 Putative *E. tef GA20 ox* isolation and cloning

Based on the known sequence information obtained previously for the three putative sequences of *E. tef*1-3 sequence specific primers (SP primers) (Table 4.2) were designed and used for nested PCR amplification. Generation of full-length sequences of coding regions was done using 5´/3´ RACE (Random Amplified cDNA ends). Application of RACE for *E. tef GA20ox* resulted in two 5´ and two 3´ short sequences (contigs) with sizes of 238, 263 and 121 and 305bp. No further sequence extension could be obtained for *E. tef*2 *and E. tef*3 and *GA20ox2*. The four fragments of *E. tef*1 were aligned using "contig assembly" of the vector NTI program (Vector *NTI Advance*™*9.0*). The consensus sequence of 1448bp was translated using the ExPASY protein translation tool. The sequence consisted of an open reading frame encoding a putative polypeptide of 482 amino acids. Comparison of the deduced amino acid sequence with other plant species (sorghum, rice, maize, wheat and barley) showed the presence of all characteristic conserved (consensus) amino acids sequences (domains):

LPWKET, (144-149 aa) and NYYPXCXXP (228-236 aa of the putative sequence) and three His residues for binding Fe2+ (245-247 aa counting from the start codon). BLAST results of the full coding region of putative *E. tef GA2ox1* showed high identity scores with orthologous genes from *S. bicolor* (68.3%), *Z. mays* (68.7%), *O. sativa* (59.3%), *L. perenne* (64.6), *T. aestivum* (66.6%), *H. vulgare* (67.5%) and *Z. japonica* (54.5%) (Figs. 4.1 - 4.6).

Similarity among the tef *GA20ox* homologs and comparison with orthologs of other monocot species was determined based on amino acid sequences using partial sequences trimmed to core areas present in all the three sequences. *E. tef*1 showed 89.6% and 59.8% homology to *E. tef*3 and *E. tef*3 sequences, respectively. *GA20ox* isolated from different species exhibits a conserved domain of amino acids with identity ranging from 50 % - 75% (Hedden and Kamiya, 1997). *E. tef1* and *E. tef*3 sequences were found to be closely related to each other and *E. tef GA20ox2* sequence was found to be closely related to *GA20ox2* genes sequences from various other monocot species (Figs. 4.4 – 4.6). The three homologous sequences were tentatively named as *E. tef GA20ox1a, E. tef GA20ox1b and E. tef GA20ox2.*

The *E. tef GA20ox2* partial coding sequence has high polymorphism at the N-terminus when aligned and compared to orthologs from other closely related monocot cereal species. Generally, alignment and phylogenetic relationship further showed identity with *GA20ox* sequences from other plant species such *S. bicolor (*81%), *O. sativa (sd1) (*68%), *O. rufipogon (*72%), *T. aestivum (*57%), *H. vulgare (*58%), *Z. mays (*56%), *A. thaliana (*57%) and *L. perenne (*58%) when 112-122 aa sequences present in all sequences from those species were aligned (Figs. 4.3 and 4.4). *EtGA20ox2* also has a 120 bp intron between the two conserved functional domains (Fig. 4.3). Further, *EtGA20ox1a, EtGA20ox1b* groups closely

with GA20ox1 sequences from sorghum, maize, zyocia *spp.* and rice. A similar result was found for *EtGA20ox2* closely grouping with GA20ox2 orthologs from these plant species.

```
E. tef 1   TGCGGCAGCGTGAAGAAGGCGTCCATTCGAGTGGAGTCTCGGAATCCTCAAACTCACTGCAGCAGCAGCAGCAGTAGTAGTAGTAGTCCTCCTCCTAGTCAATCAGAGAGAGAGAGAGAG 120
           AGAGAGAGAGAGAGAGAGAGAGCCAAGCCAGCTGCCCGTGATGGTGGTGGTAGCGCAGGCGCAGGCGCAGGCGCAGATAGAAGTAGAAGCAGATGAGCCGCCGCAGCAGCTTGAGGTGGTGG 240
           TGTTCGACGCGGCCCGTCTGAGCGGGCTGAGTGACATCCCGGCCCAGTTTTTGTGGCCGGAGGAGGAGAGCCCGACGCCTGACGCGGCGGAGGAGGAGCTGGACGTTCCTCTGATCGACC 360
           TCTCCGGCGACGCGTCGGAGGTGGTCCGTCAGGTGCGTGAGGCCTGCGAGGCGCACGGCTTCTTCCAGGTGGTGAACCACGGCATCGACGCCGGCCTCGTGGCGGAGGCGCACCGCTGCA 480
           TGGACGCCTTCTTCACGCTGCCGCTGCCGGAGAAGCAGCGCGCCCAGCGCCAGCCCGGCGACTGCTGCGGCTACGCCAGCAGCTTCACGGGCCGCTTCGCCAGCAAGCTGCCATGGAAGG 600
E. tef 2   -----------------------------------------------------------------------------------------------------CGATTGCTGCCGTGGAAGG 19
E. tef 3   -----------------------------------------------------------------------------------------------------CGATTGCTGCCGTGGAAGG 19

E. tef 1   AGACGCTCTCCTTCCGGGCCAAGGCCAAGGCCACCGATGATGATGATGATGATGTGGTGGCTACTACGTCAGCAAGCTGGGCGAGGCCTACAGGCGGCACGGCGAGGTGTACGGGCGCT 720
E. tef 2   AGACGCTCTCCTTC--------GGCCACCGCGAC-------------------GTCGTGGAGTACTTCACATCCACCCTCGGCAGCGACTTCAAACC-CCTAGGGAGGTGTTCCGAGACT 111
E. tef 3   AGACGCTCTCCTTCCGGGCCA-GCCCAACGTCGCCGGCCTTG-----------GTGGAGGACTACCTGGTGGGCCGCCTTGGCGACGAGTACAGGCGGCACGGCGAGGTGTACGGGCGCT 127

E. tef 1   ACTGCTCGGAGATGAGCCGGCTGTCGCTGGAGATCATGGAGGTGCTGGGCGAGAGCCTGGGCGTGGGCCGCCGCTGCTTCCGCGACTTCTTCCAGGACAACGACTCCATCATGCGCCTCA 840
E. tef 2   ACTGCCAATCGATGAAGGAGGTGTCGCTGGCGATCATGGAGGTGCTGGGCGCGAGCCTGGGCGTGGGGAGGCGCTACTGCAGGGACTTCTTCGCCGACGGCTGCTCCATCATGAGGTGCA 231
E. tef 3   ACTGCTCGGAGATGAGCCGGCTGTCGCTGGAGATCATGGAGGTGCTGGGCGAGAGCCTGGGCGTGGGCCGGCGCTGCTTCCGCGACTTCTTCCAGGACAACGACTCCATCATGCGGCTCA 247
                                                      ***   ***
E. tef 1   ACTACTACCCGCCGTGCCAGCGGCCCATGGAGACGCTGGGCACGGGCCCGCATTGCGACCCCACCTCCCTCACCATCCTGCACCAGGACCACGTCGC---CGGCCTCCAGGTCTTCGCCG 957
E. tef 2   ACTACTACCCGCCGTGCCCGGAGCCGGACCGGACGCTGGGCACGGGGCCCCACTGCGACCCGGCGGCCCTCACCCTCTTGCTCCAGGACGACGACGTGGACGGGCTCCAGGTGCTCGTCG 351
E. tef 3   ACTACTACCCGCCGTGCCAGCGGCCCATGGAGACGCTGGGCACGGGCCCGCATTGCGACCCCACCTCCCTCACCATCCTGCACCAGGACCACGTCGC---CGGCCTCCAGGTCTTCGCCG 364

E. tef 1   CCGGCCGGTGGCTCTCCATCCGCCCGCACGCCCAGGCCTTCGTCGTCAACATCGGCGACACCTTCATGGCGCTCTCCAACGGCCGGTACAAAAGTTGCCTGCACAGGGCGGTGGTCAACA 1077
E. tef 2   ACGGCGAGTGGCGGCCCGTGCGGCCCAAGCCGGGAGCCATCGTCGTCAACATCGGCGA-------------------------------------------------------------- 409
E. tef 3   GCGGCCGGTGGCTCTCCATCCGCCCGCACGCCGCCGCCTTCGTCGTCAACATCGGCGA-------------------------------------------------------------- 422

E. tef 1   GCAGCGTCCCGCGGAGGTCCCTGGCCTTCTTCCTCTGCCCGGAGATGGACAAGCTCGTCACGCTCCCGCCGCAGCTTCTTCCTGATCTGCCCGGAGACAACCAAAGGAGGCGCCCCTACC 1197
           CGGACTTCACGTGGCGCACCCTGCTCGAATTCACGCAGAAGCACACAGGGCCGACATGAAGACGCTCGAGGTCTTCTCCAACTGGCTCCGCCATGGCCAGGACAAGGTAGCGCTACCTCC 1317
           CCTATGCTAGTTATTCTCTCCATACCAATATGTAATGTAATGTAATGAATGTAGTCATCCATGACCGCCTGTTCTGTTCAAAAAAAAAAGTCGACATCGATACGCGTGGTCCGCCTCCTT 1437
           TGGTTGTCTCC 1448
```

**Figure 4.1** Nucleotide sequence alignment of three putative *E. tef GA20ox* sequences. Putative $Fe^{2+}$- binding consensus regions are indicated with highlighted asterisks (*), the 2-oxoglutarate-binding motif is indicated over-lined, and the nucleotide sequence for the conserved LPWKET region, considered to be involved in the binding of GA substrate, is shown with double over-line. In *GA20ox2,* the 120 bp intron between LPWKET and NYYPPCPEP functional domains is not included in the alignment. Dashes (-) have been inserted to maximize sequence homology. Identical and similar regions are shown by light and dark shaded areas and dots whereas number indicates the position of the conserved regions within the predicted nucleotide sequence.

```
MVVVAQAQAQAQIEVEADEPPQQLEVVVFDAARLSGLSDIPAQFLWPEEESPTPDAAEEELDVPLIDLSG-------DASEVVRQVREACEAHGFFQVVNHGIDAGLVAEAHRCMDAFFT 113
------------------------------------------------------------------------------------------------------------------------ 1
MV-------..SLG.PLLLQPPPPSL.....................Q.....AD........-...A........--------..A........R..DL...............DA.LQ.............. 105
MV-------L.AHD---------PPPL....................Q..I..AD........S.-...A.......--------.A........R..DL........G.....A.T............... 96
MVVQQ------------------EQE......V...QTE..S..I..A.....GSV.V-...E.A...VGAGA-----ER.S......G....R.....L.......E.A.LE................ 98
MVQP----------------------......L..AQ....S..I..AD........T-...H......IG.LLSGDREA.A...T.L.GD...R...................E.LGH.R-.V...... 95
MVRP----------------------......V...R....S..I..G.........-...H.....NIG.MLSGDAAA.A...T.L.G....R...................E.L.D.....V.N.... 96
MVQP----------------------......L...RA...S..I..G.........-...H......IG.MLSGDADA.A...T.L.GQ...R..L................E.L.D.....V...... 96
MVQP----------------------......V...RT...S..I..G........T-..MH......IG.MLSGDPRA.A...T.L.G....R...................Q.L.D.....V...... 96
MAAASTPNG---------------------.PPVR.LATTV.V.AVLFDIDGTLC.SD-PLHH.AFQE.LL----------.IGYNNGVPIG-..............QEPL........N.... 88

LPLPEKQRAQRQPGDCCGYASSFTGRFASKLPWKETLSFRAKAKATDD-DDDDVVGYYVSKLGEAY-RRHGEVYGRYCSEMSRLSLEIMEVLGESLGVG---RRCFRDFFQDNDSIMRLN 228
----------------------------............SPTSPAL-VE.-----.L.GR..DE.-.......................................---.................. 83
..MSD......RQ..S..........................YSDDQ-G.G--.V..D.F.D...D..-.H..........................L...........---..H..R...G........ 218
....D......RQ..S..........................YTDDDDG.KSK.V.AS.F.D....G.-.H..........................L...........---..H..R...G........ 212
...G.....RSGAR.-RTA.......................YSSAGDEEG-EEG.GE.L.R...AEHG..L.....S...H..............L.......IVGDR.HY..R...R........ 216
MS.QD.....L.R..ES.........................SCPSE------P.L..D.I.AT...DH-..L.....A..............................---.AHY.R..EG.E....... 206
M.........L.R..ES.........................SCPSD------PAL..D.I.AT...DH-..L.....A..............................---.AHY.R..EG......... 207
M.........L.R..ES.........................SCPSD------PAL..D.I.AT...GH-..L.....A..............................---.AHY.R..EG......... 207
M.........L.R..ES.........................SCPSD------PAL..D.I.AT...DH-..L.....A..............................---.AHY.R..EG.E....... 207
..........RQ.ES...........C...............YSSNPSS---P.L..D.F.E....D..-.H..A...A.................M.GI..........---.DH..R...P.E........ 201
* *
YYPPCQRPMETLGTGPHCDPTSLTILHQDHVAGLQVFAAGR-------WLSIRPHAQAFVVNIGDTFMALSNGRYKSCLHRAVVNSSVPRRSLAFFLCPEMDKLVTLPPQLLPDLPGDNQR 342
........-...........................................G..-------....-...A.......-------------------------------------------------------- 138
........YD.........................D.G......D.ATGPGTGR.R.....PG.......................R...................R.....................V.RP.AE.V------.DAN 333
........YD.........................D.G......D.AT----LA.R....RPG.......................R...................R.A.....................V.RP.KE.V------.DAN 323
...A....LD..........................G..E..W.E..-------.RA...RPG.L....V..........A..R...................TA.....................TV.RP.EE.V------.DHH 325
........N..........................D.G.....H.D..-------......R.D........................................R...K...................V.AP.GT.V------.EAN 315
........L..........................N.G.....HTE..-------.R....R.D........................................R...K...................V.AP.GT.V------.ASN 316
........Y..........................D.G.....HTE..-------.R....R.D........................................R...K...................V.AP.GT.V------.AAN 316
........L..........................D.G.....HTD..-------.R....R.D........................................R...K...................V.AP.GT.V------.AAN 316
........L............................G......TD..-------.R......G..............TGALQRPVPKL..........L...K...................V.CP.EG.V------EAGM 310

RRPYPDFTWRTLLEFTQKHTGPTRRSRSSPTGSAMARTRRYLPYASYSL 391
-------------------------------------------------- 138
P.A.........D..MR.YRSDM.TLEAFSNWLN---HGG-HLS.PPP 378
P.A.........D..MR.YRSDM.TLEAFSNWLSTSSNGGQHLLEKKK 372
P.V.......A..D...R.YRADM.TFQAFSDWLN----HHRHLQPTIY 370
P.A.......A..D.....YRADMKTLEVFSDWIQQGHQP-----AATT 359
P.A.......S..D.....YRADMKTLEVFSSWIVQQQQGQ.ALQPAMT 365
P.A.......S..D.....YRADMKTLEVFSSWVVQQQQ------AAMT 359
P.A.......S..D.....YRADMKTLEVFSSWVVQQQP-----..ART 360
P.A.........D...RRYRADM.TLEVFSNWLRHGQDKTP.TSLIHR 359
```

| Identity (%) | |
|---|---|
| 100 | *E. tef 1* |
| 80 | *E. tef 3* |
| 68 | *S. bicolor GA20ox* |
| 69 | *Z. mays GA20ox* |
| 59 | *O. sativa GA20ox* |
| 67 | *L. perenne GA20ox* |
| 67 | *T. aestivum GA20ox1A* |
| 67 | *D. villosum GA20ox* |
| 68 | *H. vulgare GA20ox1* |
| 54 | *Z. japonica GA20ox* |

**Figure 4.2** Derived amino acid sequence alignment of two putative *E. tef GA20ox* (*E. tef1* and *E. tef3*) sequences with orthologous *GA20ox* gene sequences from sorghum (*S. bicolor*; Acc No. XP_002463483.1), maize (*Z. mays;* Acc No. ACF83905.1), rice (*O. sativa*; Acc No. P93771.2), wheat (*T. aestivum;* Acc No. 004707.1 ), barley (*H. vulgare*; Acc No. AAT49058.1), lolium (*L. perenne*; Acc No. AAG43043.1), Zyocia (*Z. japonica*: Acc No. ABG33927.1) and Dasypyrum (*D. villosum*; Acc No. ACU40946.1). Putative $Fe^{2+}$- binding consensus regions are indicated with asterisks (*), the conserved motifs, NYYPXCXXP and LPWKET are shown with deep dark shades. Identical and similar regions are shown by light and dark shaded areas and dots whereas number indicates the position of the amino acid within the predicted peptide.

```
E. tef 2                LPWKETLSFGHRD----------VVEYFTSTLGSDFKPLGEVFRDYCQSMKEVSLAIMEVLGASLGVGRRYCRDFFADGC 70
                                                TTVEIASFATFFFVRTRDRSSESFYNDCTDRAMHGRR

S. bicolor GA20 ox2      .............RRTSG--SHV..D..............V..YQN..NA............I.V......S.Y........S 78
O. sativa sd1           ..........FH.RAAAP----V.AD..S....P..A.M.R.YQK..EE...L..T...L.EL....E.G.Y.E....SS 76
O. sativa GA20 ox2      ............HANAAGNNSST.AD..S-...D...H....YQE..EA.E..TK...A...E.....GG.Y.E..E.SS 79
O. rufipogon GA20 ox2   ..........FH.RAAAP----V.AD..S....P..A.M.R.YQK....E....L..T...L.EL....E.G.Y.E....SS 76
P. sativum GA20 ox      .........QFS.EKNS---SNI.KD.LSN...E..QQF....YQE..EA.SKL..G...L..M.....KECF.....EENK 77
A. thaliana GA20 ox2    .........QFSNDNSG---SRT.QD..SD...QE.EQF.K.YQ...EA.SSL..K...L..L....N.D.F.G..EEND 77
Z. mays GA20ox          .........RYT.DDDGDKSKDV.AS..VDK..EGYRHH....YGR..SE.SRL..EL.....E.........HF.R..QGND 80
L. perenne GA20ox       .........RSCPSE-----PDL..D.IVA...E.HRR.....YAR..SE.SRL..E.......E.......AHY.R..EGNE 75
T. aestivum GA20ox1A    .........RSCPSD-----PAL..D.IVA...E.HRR.....YAR..SE.SRL..E.......E.......AHY.R..EGND 75
H. vulgare GA20ox1      .........RSCPSD-----PAL..D.IVA...E.HRR.....YAR..SE.SRL..E.......E.......AHY.R..EGNE 75

                                                                  Identity (%)
E. tef 2                SIMRCNYYPPCPEPDRTLGTGPHCDPAAHTLLLQDDDVDGLQ- 112    100
S. bicolor GA20 ox2      ....................E.............S.L.V.....G...----  117     81
O. sativa sd1           ....................E.............T.L.I.....-..G..EV 118     68
O. sativa GA20 ox2      ....................E.............S.L.V.....G......V 122     72
O. rufipogon GA20 ox2   ....................E.............T.L.I.....-..G..EV 118     69
P. sativum GA20 ox      .....L........QK..L...............TSL.I.H...-Q.G...V 119     57
A. thaliana GA20 ox2    .....L.H......QT..L...............SSL.I.H...-H.N...V 119     57
Z. mays GA20ox          .....L........QR.YD...............TSL.I.H...-..G...V 122     56
L. perenne GA20ox       .....L........QR.NE...............TSL.I.H...-..G...V 117     58
T. aestivum GA20ox1A    .....L........QR.LE...............TSL.I.H...-N.G...V 117     57
H. vulgare GA20ox1      .....L........QR.LE...............TSL.I.H...-..G...V 117     58
```

**Figure 4.3** Derived amino acid sequence alignment of putative *E. tef GA20ox2* sequence (*E. tef2)* with orthologous GA20ox gene sequences from other grass species namely sorghum (*S. bicolour*; Acc No. XP_002441117.1), maize (*Z. mays*; Acc No. ACN25832.1), rice (*O. sativa and O. rufogen,* Acc No. AAT44252.1 and BAK39011.1), wheat (*T. aestivum;* Acc No. 004707.1) barley (*H. vulgare*; Acc No. BAK04752.1), pea (*P. sativum*; Acc No. AAF29605.1) and Arabidopsis (*A. thaliana*; Acc No. NP_199994.1). The position of the 37 aa intron region is indicated in the first row 23aa downstream LPWKETL motif between amino acids "G" and "E". Identical and similar regions are shown by light and dark shades respectively and numbers indicate the position of the amino acid within the predicted peptide. Percentage of identity is shown in the parenthesis.

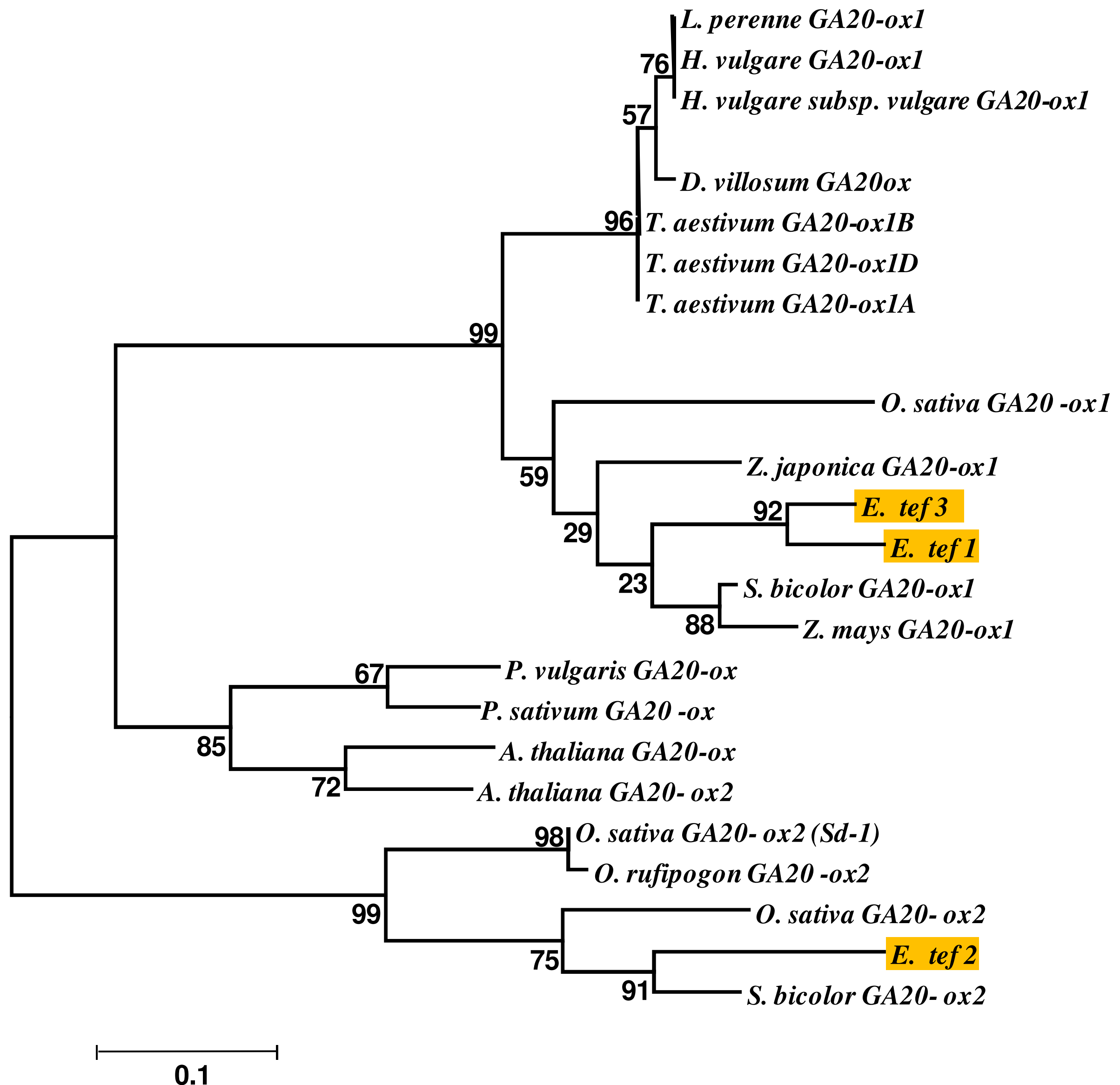


**Figure 4.4** Molecular phylogenetic analysis of the homologous of *E. tef* GA20oxsequences (*E. tef GA20ox1, GA20ox1b and GA20ox2*; underlined). The tree was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. Initial alignment was done by mafft (http://mafft.cbrc.jp/alignment) software. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically as follows: when the number of common sites was < 100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. The tree is drawn to scale with branch lengths measured in the

number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 105 positions in the final dataset.

### 4.4.3 *EtGA20ox* copy number

To determine the copy number of *EtGA20ox* gene in *E. tef*, Southern blot analysis of genomic *E. tef* DNA restricted with different enzymes (BamHI, EcoRI and HindIII) and probed with a 450-base-pair coding sequence of *EtGA20ox* was carried out. The probe hybridized with four fragments after the HindIII digest (one is very slightly visible); four fragments after BamHI digest, and two very visible and two hardly detectable fragments after EcoRI digest (Fig. 4.5). Therefore, four copies of *EtGA20ox* are possibly present in the allotetraploid *E. tef* genome.

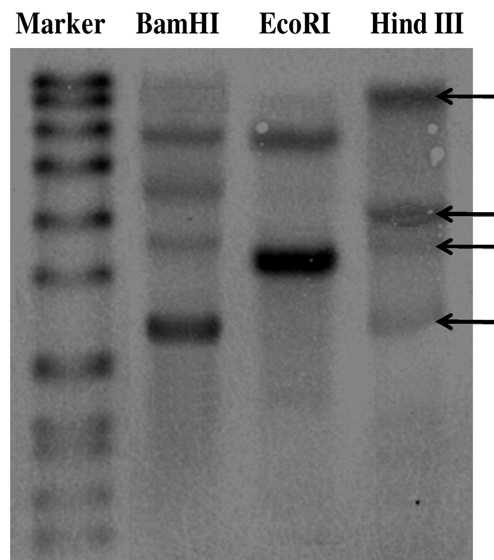


**Figure 4.5** Detection of EtGA20 ox gene copies in the *E. tef* genome after restriction enzyme digest using Southern blotting.

#### 4.4.4 *GA20ox* expression in *E. tef*

In both *E. tef* genotypes, Gea Lammie (short phenotype) and DZ-01-196 (long phenotype), *GA20ox1* was expressed in germinating hypocotyls, leaf, stem and inflorescence. Relatively higher expression was found at earlier stages in young stem and leaf (Fig. 4.17) with slight differences between the two genotypes. Further, when semi-quantitative PCR (RT-PCR) was used to study expression of the three *E. tef GA20ox* homologous sequences in leaf, leaf sheath, uppermost two internodes, nodes from these internodes, peduncle and inflorescence both *EtGA20ox1a* and *EtGA20ox1b* expression was highest in the uppermost internodes followed by nodes (Fig. 4.7). However, *EtGA20ox1b* expression was greater than *EtGA20ox1a* expression in the internodes, nodes, leaf sheath and panicle. In contrast, *EtGA20ox2* was expressed at relatively lower rates in the nodes and panicle but also in the upper most internodes when compared to the other two genes.

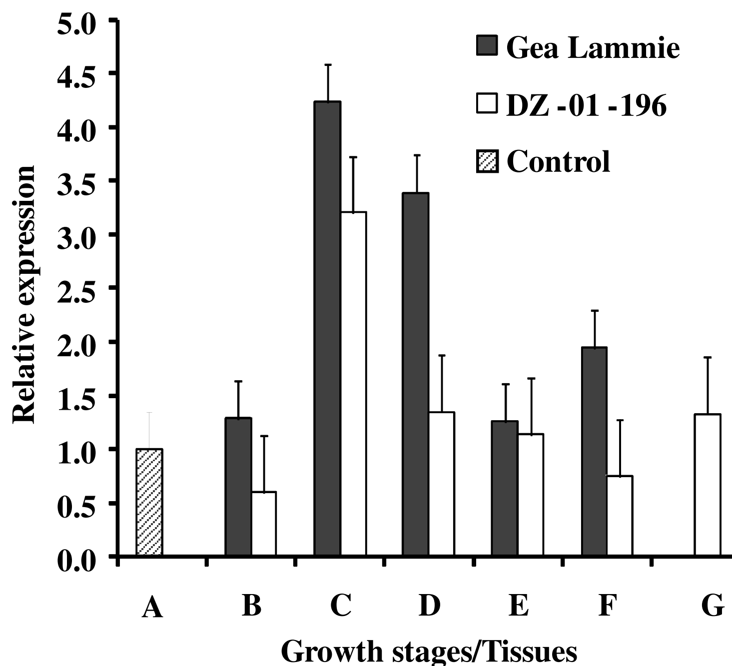

**Figure 4.6** Relative expression of GA20ox1 in two *E. tef* genotypes DZ-01-196 and Gea Lammie in different plant tissues and growing stages. (A) Control; (B) 3$^{rd}$ week stem + leaf; (C) 5$^{th}$ week stem + leaf; (D) 8$^{th}$ week stem; (E) 8$^{th}$ week leaf (F) Old (10$^{th}$ week) stem and (G) 10 days old inflorescence. The qRT-PCR was repeated three times, and gene expression level was calculated following the expression $2^{-DCt}$ (Yang *et al.*, 2005).

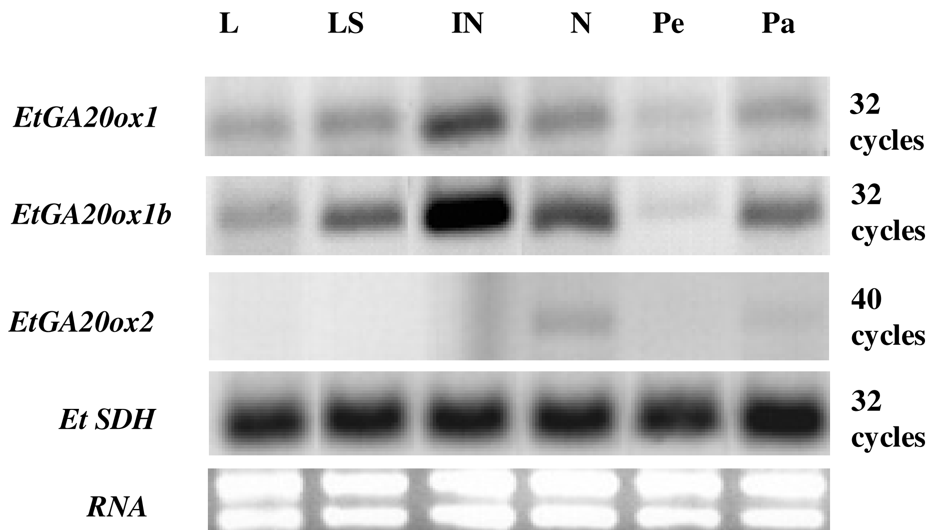


**Figure 4.7** Semi quantitative RT-PCR Expression analysis of three *EtGA20ox* genes in various plant tissues (L=Leaf; LS=Leaf sheath; IN= Internode; N= Node; Pe= Peduncle; Pa= Panicle) from cultivar DZ-01-196 sampled at the stage of stem elongation and panicle initiation (for panicle). RT-PCR was performed with 32 and 30 cycles for *EtGA20ox* and *EtSDH*, respectively. The experiment was repeated three times with similar result. *EtSDH* was used as an internal PCR amplification control.

### 4.4.5 *In vitro* enzymatic activity of *GA20ox* in a heterologous system

In *E. coli* expressed *GA20ox1* protein catalyzed the conversion of the [$^{14}$C]-labelled gibberellin precursor $GA_{12}$/ to $GA_9$ which could be detected by full-scan GC-/MS. In some clones only a partial conversion of the substrate was found, possibly due to sub optimal reaction conditions, and intermediates of the reaction pathway, $GA_{24}$ and $GA_{15}$, were also detected by HPLC analysis (Fig. 5.8). Clone 1 showed optimum activity by complete conversion of $GA_{12}$ into $GA_9$>$GA_{15}$>$GA_{24}$. Clones 2 and 3 showed only partial activity.

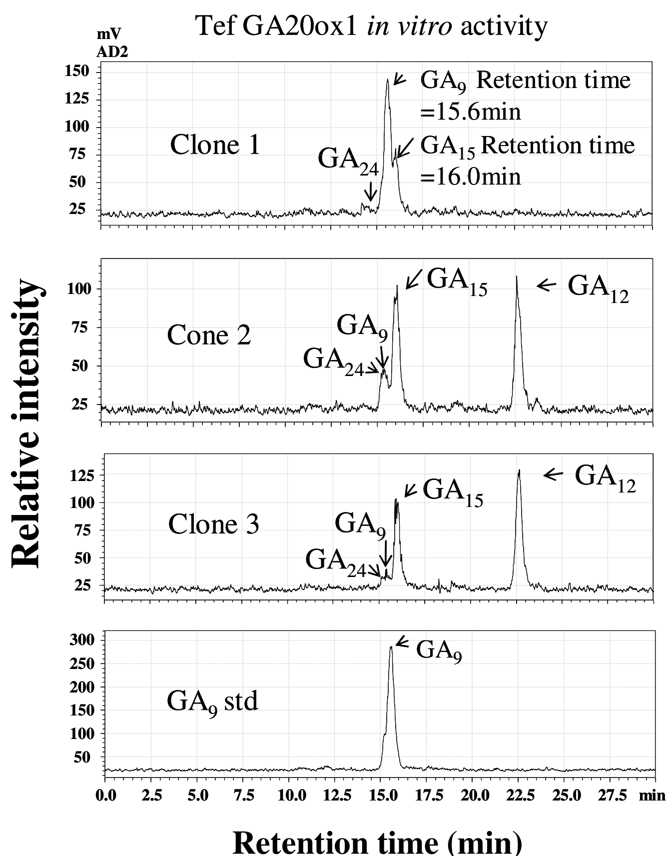


**Figure 4.8** Radiochromatograms after HPLC of *E. tef GA20ox1* activity products after incubation with a $^{14}$C-labeled $GA_{12}$ as a substrate (standard shown at the bottom).

4.4.6 Isolation of the *Rht* and other genes

For isolation and cloning of *E. tef Rht,* specific primers (SP) (Table A.1) were designed based on a 344 -base-pair known sequence for the *Rht* gene. Primers were used for nested PCR to amplify a putative *Rht* orthologous gene from *E. tef* similar to the procedure used for *GA20ox* to obtain a full-length sequence. PCR amplification resulted in two 5´ and two 3´ contigs providing a total fragment size of 1395 bp after assembling the contigs. The sequence analysis of the amplified cDNA fragment (1395bp) consisted of an open reading frame encoding a putative polypeptide of 465 amino acids. Comparison of the *E. tef Rht* amino acid sequence with sequences from other plant species, such as sorghum, rice, maize, wheat and barley, showed the characteristic conserved amino acids domains such as the DELLA and GRAS domains for the *Rht* gene (Figs. 5.9 and 5.10). When the aligned sequences were trimmed to core areas present in all sequences (438-465 aa) a high identity to sequences of *Z. mays* (90%), *O. sativa* (86%), *S. bicolor* (85%), *T. aestivum* (84%) and *A. thaliana* (55%) were found. The above gene sequence in *E. tef* was named *EtRht* according to the nomenclature used by Cloes *et al*. (1999).

Partial sequence for brassinosteroid inactivating gene that encodes the Cytochrome P450 monooxygenase family gene, was also cloned using semi-degenerate oligonucleotides. After optimization of PCR conditions degenerate primers F1: 5´- ACA GCC GCA GCG TCT CGT T(GCT) - 3´ and primer R1: 5´ - A(GA)(CG) (CG)(TC)C (CA)AC GGC GA(CGA) (GC)(CG)(AT) GTT, were used to obtain 749 bp long fragment. Further sequencing and comparison through BLAST-search showed homology range between 49 - 52% with ortholog gene sequences of the Cytochrome P450 monooxygenase family for closely related species including the rice CYP734A1 involved in BRs catabolism (Fig. A.10 and Fig. A.11).

Domain I Domain II

```
EE-VDELLAALGYKVRSSDMADVAQKLEQLEMAMGMGGVP----AADDGFVSHLATDTVHYNPSDLSSWVESMLSELNAPPPPLPPAPAPPAPQL-VSTSSTVT-GGGSG 103
.D-.....................................GGAGAT.....................................A.....--T...R.-A........S.AAA. 106
..-M..M...V.............................GGAGAT....I..................................---------...R.-A........S.AAA. 100
.D-......................................A..SAPG-........................................L..I....--..ARH-A........G....- 104
GNMD.....V.......E..E..L.....TM.SN---------VQE..-L..............E.Y..LDN.......P...L.-------------A.SNGLDP------ 82
..-......A..........................G-AGA.P..S.AT..............T............................-----....NA........GS.--- 100

AGYFDPPPAVDSSSSTYALKPIP-SPVAA-PADP----SADSAREPKRMRTGGGSTSSSSSSSSMGGGGARSSVVEAAPP---ASAAANAPAVPVVVVDTQEAGIRLVH 204
....L........-....PS..-----T.........................D..RT........ATQ.....G..................... 211
....L........-...V.-S..S---T..T......................D..RT........ATQ.....G......M..P.......... 205
-.F.EL.A.A.......R..S-L..V.-T..S---A....DT..............L...AS.G.........AMQGA...................... 208
--VLPS.EICGFPA.D.D..V..GNAIYQF..IDS---.SS.NNQN..LKSCSSPD.MVT.T.TGTQI..VIGTT.TTTTT---TTT..AESTRS.IL..S..N.V.... 184
-....L..S.......I...R...-..AG.-T.PADL--......V.D.......S..............L...-.............--V.A..NAT..L............... 202

ALLACAEAVQQENFSAAEALVKQIPMLASSQGGAMRKVAAYFGEALARRVYRSPPPPPTAPSSTPPSPTSSTPHFYESCPYLKFAHFTANQAFLEAFAGCRRVHVVDFGI 314
..............................................FR..DSS-LLDAAFADLLHA...................I.......................... 320
..............D...........................FR.T.DSS-LLDAAVADFLHA...................I.......................... 314
............A.........T..A.................FR.-ADST-LLDAAFADLLHA...................I.......................... 316
..M.....I..N.LTL.........GC..V...A.......T...A.....I..LS.-.QNQ--IDHCLSDTLQM....T..............I....E.KK....I..SM 291
............L...........L..A.....................F.FR.Q.DSS-LLDAAFADLLHA...................I.......................... 311

EQGMQWPALLQALALRPGGPPSFRLTGVGPPQPDETDALQQVGWKLAQFAHTIRVDFQYRGLVAATLADLEPFMLQPEGE-ENDEEPEVIAVNSVFEMHRLLAQPGALEK 423
K..................................................................................D-DT.D...............L............ 429
K..L..........................H....................................................D-DK.................L............ 423
K..................................................................................ADAN..................L............ 426
N..L.....M.......E.....T......I....A..NS.H.HE..C....L.EA.H.E.E...F..NS....DAS..ELRPS-----DT.AV.......L.K...GR..GI.. 396
K..................................................................................EDPN...............................  421

VLGTVRAVRPKIVTVVEQEANHNSGSFLDRFTQSLHYYSTMF 465
..........R...............T......E........ 471
..........R...............T......E........ 465
.....H....R.......................E........ 468
...V.KQIK.V.F......S....GPV......E........L. 438
..........R...............T......E........ 463
```

| | Identity (%) | |
|---|---|---|
| 465 | 100 | *E. tef Rht* |
| 471 | 90 | *Z. mays D8* |
| 465 | 85 | *S. officinarum GAI* |
| 468 | 86 | *O. sativa Indica GAI* |
| 438 | 84 | *A. thaliana GAI1* |
| 463 | 55 | *T. aestivum rht-D1a* |

**Figure 4.9** Amino acid sequence alignment of the putative tef *Rht* sequences with orthologous amino acid sequences from maize (*Z. Mays*; Acc No. AAL10325.1), rice (*O. sativa*; Acc No. EAY91579.1), wheat (*T. aestivum;* Acc No. Q9ST59.1) and sugarcane (*S. Officinarum*; Acc No. AAZ08571.1) and Arabidopsis (*A. thaliana,* Acc No. BAC42642.1). The putative sequence shows the characteristic domains (dark shaded) for this gene: a conserved N-terminal: DELLA motif (Domain I), a VHYNP motif (Domain II) and VHVVD. The red boxes in *Z. maize* and *T. aestivum* in domain I and domain II show positions of small internal deletions and introduction of a stop codon (in wheat) in alleles that cause semi-dwarf phenotype (Benetzen and Mulu, 2000). Identical and similar regions are shown by grey and light grey shaded areas and dots whereas numbers indicate the position of the amino acid within the predicted peptide.

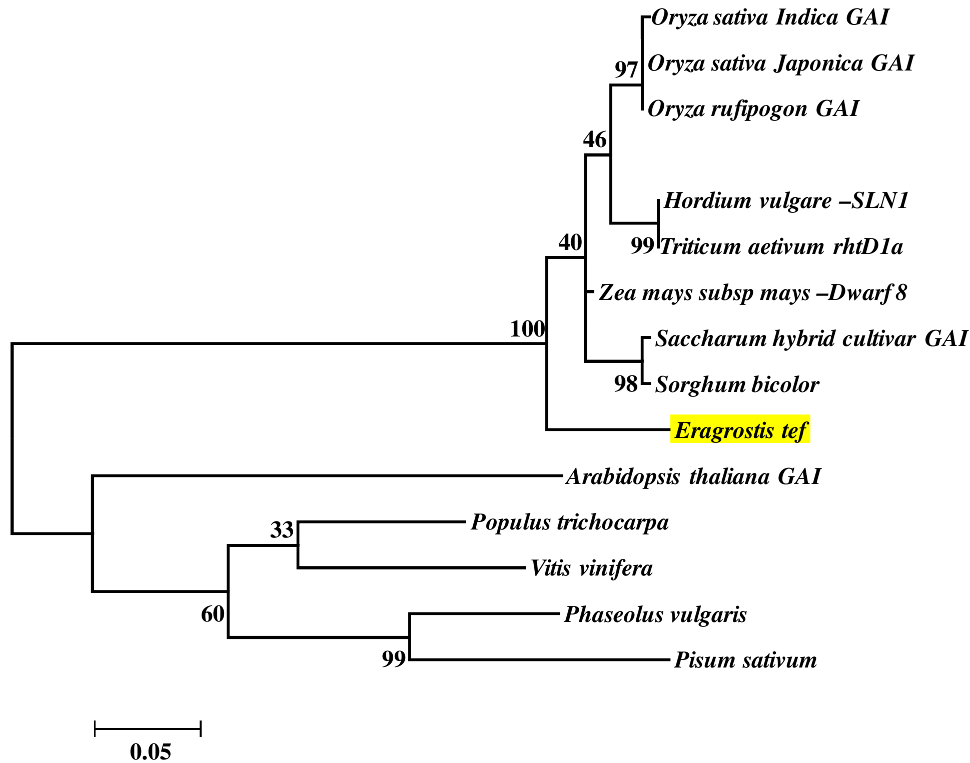


**Figure 4.10** Molecular Phylogenetic analysis of *E. tef* putative *Rht* gene by Maximum Likelihood method. The evolutionary history was inferred based on the JTT matrix (Jones *et al.*, 1992). Initial alignment was done by mafft (http://mafft.cbrc.jp/alignment) software. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was < 100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There were a total of 478 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).

## 4.4 Discussion

In this study full and partial coding sequences of three GA biosynthesis gene homologs for the GA20ox has been isolated, characterized and the expression analyzed. The putative *E. tef GA20ox* genes showed close homology with genes of other cereal species (sorghum, maize, rice, wheat, barley) and a grass species *Zyocia japonica*. The putative *GA20ox* genes belong to the GA 2-oxoglutarate-dependent dioxygenase (2-ODD) gene family as confirmed by the heterologous expression assay producing final and intermediate products of the GA20ox enzymatic actions. For convenience, gene nomenclature has been kept consistent with the naming of genes used in several recent publications about GA-biosynthesis and signalling genes in plant hormone metabolism (Thomas *et al*., 1999; Sakamoto *et al*. 2004; Spielmeyer *et al*. 2004).

The three *E. tef* GA20 oxidases also have characteristic conserved amino acid residues, namely the 2-oxoglutarate co-factor, LPWKET motif, very likely involved in the binding to the GA substrates (Xu *et al*., 2002), and the NYYPXCQKP motif for common co-substrate binding (Miao *et al*., 2010). The conserved H and D residues are involved in Fe2+ binding at the active site of isopenicillin N synthase (Hedden and Phillip, 2000 and Xu *et al*., 1995). In *E. tef GA20ox2* an intron (120 bp) has been further found *23 aa* downstream of the LPWKET domain.

The three *E. tef GA20ox* genes (hereafter named EtGA20ox) showed differential transcription in different plant tissues when semi-quantitative RT-PCR was applied. Generally, *EtGA20ox1a* and *EtGA20ox1b* were similarly expressed in tissue with highest transcription in the uppermost

internodes followed by nodes. In contrast, *EtGA20ox2* transcription was relatively lower than transcription of either *EtGA20ox1a* or *EtGA20ox1b*. These two genes might therefore act to promote internode elongation and panicle growth and might further also be involved in reproductive growth of the panicle. Such overlapping transcription has also been found in Arabidopsis where *AtGA20ox1* and *AtGA20ox2* act redundantly promoting elongation in the hypocotyls and internode, flowering time, and elongation of anther filaments including seeds number per silique (Rieu *et al*., 2008). Also, in Arabidopsis GA20ox1 greatly contributes to internode elongation (Galun, 2010). Further, in rice *OsGA20ox2* (*SD-1*) is a predominant dwarfing gene with a role in height control (Spielmeyer *et al*., 2002, Monna *et al*., 2002). Results in this study indicate, however, that *EtGA20ox1b*, and not *EtGA20ox2*, is the functional gene in *E. tef* for height control equivalent to *OsGA20ox2* (*SD-1*). However, a future study is required to investigate whether *EtGA20ox1b* regulation will affect *E. tef* plant growth as previously found in rice.

*EtGA20ox1a* also showed differential transcription when quantitative RT-PCR was used. This was dependent on growth stages and plant parts in the two genotypes, Gea Lammie and DZ-01-196. *EtGA20ox1a* transcription was found in DZ-01-196 in the immature (emerging) inflorescence, germinating hypocotyls, young stem as well as leaf. However, *EtGA20ox1a* transcript abundance was generally higher in the shorter genotype, Gea Lammie, than in the taller genotype DZ-01-196. It is still unclear if higher transcript abundance in the short genotype depends on a feedback response interfering GA biosynthesis since the level of bioactive GA measured in Gea Lammie in this study was about half the amount in DZ-01-196. Increased transcript levels (reduced sensitivity to exogenous GA application) have been reported for

response pathway mutants in different crops due to impairment in the response genes (Olszewski, *et al*, 2002)**.** Also a low bioactive GA amount despite relatively high transcription of *GA20ox* has to be further investigated.

In a heterologous expression, *EtGA20ox1a* substrates are converted through several successive oxidations at C-20 of $GA_{12}$ to the alcohol and aldehyde intermediates $GA_{15}$, $GA_{24}$ and $GA_9$ (Hedden and Kamiya, 1997 and Yamaguchi 2008). Further, it has been shown that $GA_9$ is converted by *3β*-hydroxylation to the bioactive $GA_4$ (Junttila *et al*., 1992 and Rood and Hedden, 1994). All these products could be detected when *EtGA20ox1a* was expressed in *E. coli* and reacted also with $GA_{12}$ as a *EtGA20ox1a* substrate instead of $GA_{53}$ which is the natural substrate in tef plants. This sequence was successfully expressed converting substrates to final intermediates product except in few cases where $GA_{15}$ was a more abundant product due to incomplete downstream reactions.

In this study, a *Rht* ortholog and a partial sequence of an *Eui* and a BR deactivating gene were also isolated from *E. tef*. However, none of them were further characterized regarding their transcription in *E. tef*. The *E. tef Rht* gene has a conserved N-terminal DELLA motif (Alevy and Harberd, 2005). In general, when the *Rht* gene is expressed it represses downstream genes in the absence of GA inhibiting plant growth (Hedden, 2006). Other conserved *Rht* motifs which were also found in *E. tef Rht* include the TVHYNP domain essential for the perception of an upstream GA signal, VHVVD, and a C-terminal GRAS domain which are the functional domains responsible for transcriptional regulation such as the suppressive function of DELLA proteins against GA action (Peng *et al*., 1997; Silverstone *et al*., 2001; Ueguchi-Tanaka *et al*., 2005). Two

conserved N-terminus DELLAs domains (I and II) (Fig. 12; see Appendix) found in *E. tef Rht* are highly conserved regions in angiosperms (Yasumura *et al*., 2007) presumably necessary for DELLA interaction of the peptide during GA signalling process (Alevy and Harberd, 2005).

The *Eui* gene also isolated from *E. tef* is a single recessive gene responsible for culm length modifications. *Eui* mutants have a longer culm length due to elongation of the uppermost internodes (Zhang *et al*., 2008). Ectopic expression of the *Eui* coding sequence under the control of the rice *GA3ox2* and G*A20ox2* gene promoters reduced plant height (Zahang *et al*., 2008). Further, the isolated *E. tef BR* gene has been found to be closely related to the rice CYP724As (Cytochrome P450 monooxygenases) involved in BR catabolism (Sakamoto *et al*., 2011). Several of closely related sequences also identified through a BLAST search in this study were also related to BR inactivation. Studies have shown that BR mutants develop a dwarf phenotype. In rice, CYP734As control bioactive BRs by direct inactivation of castasterone (CS), a bioactive BR, and by the suppression of CS biosynthesis thus decreasing the levels of BR precursors. In Arabidopsis, CS and brassinolide (BL) are inactivated mainly by two Cytochrome P450 monooxygenases (CYP734A1/BAS1and CYP734A1/BAS1) that inactivate CS and BL by C-26 hydroxylation (Ohnishi *et al*., 2009).

In summary, genes were isolated in this part of the study allowing further characterization for their function. Given the redundancy of the GA biosynthetic genes reported in this study for tetraploid *E. tef*, it can be assumed that finding for instance a recessive dwarfing *sd-1* homologue mutant may be challenging since it requires independent mutations in both its A and B genomes. In such a case the semi-dominant Rht mutation version (full coding region cloned) is an easier

alternative for TILLING application and easy to express phenotypically due to its semi-dwarf nature. They can now be directly used for genetic engineering approaches using the transformation protocol developed for *E. tef*, which is part of this PhD study. Further, sequences might allow mutagenesis and selection through TILLING or Eco-TILLING and marker assisted breeding in the conventional and modern *E. tef* development process targeting lodging resistance. The genomic information developed through this part of the study therefore provides useful information for further studies to understand and establish the precise roles in plant growth of these isolated genes and for direct use in lodging resistance improvement through modern and conventional techniques.