Addis Ababa University

Institute of Biotechnology

Unit of Bioinformatics and Molecular Biology

Comparative *In silico* Study of Lodging-Resistant Genes of Tef (*Eragrostis tef* (Zucc.) Trotter) Against Wheat, Barley and Rice

By:

Tigist Shewafera

July, 2024

Addis Ababa, Ethiopia

Comparative *In silico* Study of Lodging-Resistant Genes of Tef (*Eragrostis tef* (Zucc.) Trotter) Against Wheat, Barley and Rice

By:

Tigist Shewafera

A Thesis Submitted to the Institute of Biotechnology, Addis Ababa University in Partial Fulfillment for the Requirements of the Degree of Master of Science in Bioinformatics

Advisor:

Abiy Zegeye (Ph.D)

July, 2024

Addis Ababa, Ethiopia

DECLARATION

I, the undersigned, declare that the Master thesis entitled **"Comparative *In silico* Study of Lodging-Resistant Genes of Tef (*Eragrostis tef* (Zucc.) Trotter) Against Wheat, Barley and Rice"** is my thesis work submitted to Addis Ababa University, Institute of Biotechnology. It has not been presented to any other university for any award and all the sources of materials used have been duly acknowledged.

Name: Tigist Shewafera

Signature _____ Date_____

Approval Sheet after Defense

I certify that _____'s M.Sc. thesis entitled" **Comparative *In silico* Study of Lodging-Resistant Genes of Tef (*Eragrostis tef* (Zucc.) Trotter) Against Wheat, Barley and Rice"** is the final version. Therefore; we, as examining board, approved it as the final document to be accepted as fulfilling the requirement for the degree of Master of Science in Biotechnology/Molecular Biology.

Supervisor _____Signature _____Date _____

Examining Committee

1. _____Signature _____Date _____

2. _____Signature _____Date _____

Chairman _____Signature _____Date _____

Director _____ Signature _____ Date _____

ACKNOWLEDGMENTS

First and foremost, I would like to praise almighty God, who has granted countless blessings, knowledge, and opportunities to me, allowing me to complete my thesis. I am extremely grateful to my advisor, Dr. Abiy Zegeye, for his continuous and extraordinary help, follow-up, constructive comments, and suggestions, as well as his encouragement and interest in technical discussions throughout my study. This thesis would not have been possible without his significant efforts, constructive criticisms, and discussions.

I extend my deepest gratitude to all academic and administrative staff members of the Institute of Biotechnology for their immeasurable support in providing me with a suitable working environment. Special thanks are due to Addis Ababa University and Debre Berhan University for their financial support. I also appreciate all my friends who provided advice and comments during the process of data analysis and interpretation. Finally, my heartfelt thanks go to my husband and family for their unwavering support and encouragement throughout my study.

TABLE OF CONTENTS **PAGE**

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

aas	Amino acids
Acc. No	Accession number
BAK1	Brassinosteroid associated receptor kinase1
BLAST	Basic local alignment search tool
bps	Base pairs
BR	Brassinosteroid
BRI	Brassinosteroid-insensitive1
BZR1	Brassinazole-resistant 1
CAD	Cinnayml alcohol dehydrogenase
CCR	Cinnamoyl-CoA reductase
CDS	Coding sequence
CDSf	Coding sequence first
CDSi	Coding sequence internal
CDSl	Coding sequence last
CDSO	One coding sequence
CESA1 and CESA4	Cellulose synthase A1 and A4
COMT1	Caffeic acid O-methyltransferase 1
CSA	Central Statistics Authority
CSC	Cellulose synthase complex
Cys	Cysteine pair
EBI	Ethiopian Biodiversity Institute
G	Number of gene
GA	Gibberellic acid
GAI	Gibberellic acid-insensitive
GC	Guanine and cytosine
GLIMMER	Gene locater and interpolated Markov model ER
HMM	Hidden markov model
ID	Island domain
IDD3	Indeterminate domain 3
KD	Kinase domain
LHRI	Leucine Heptad Repeat 1

LHRII	Leucine Heptad Repeat 2
LRR	Leucine-rich repeat
mRNA	Messanger ribonucleic acid
MSA	Multiple sequence alignment
NADP+	Nicotinamide adenine dinucleotide phosphate
NCBI	National center for biotechnology information
NJ	Neighbor joining
ORF	Open reading frame
Ox	Oxidase
PIF3	Phytochrome-interacting factors
PolyA	PolyA tail
Rht1	Reduced height1
SD1	Semi-dwarf 1
Seq rep	Sequence representations
SERK	Serine kinase
SIFT	Sorting in tolerant from tolerant
SLR1	Slender 1
STR	DNA strand orientation
TCP14	Teosinte Branched 1, Cycloidea, and Proliferating Cell Factor 1 (TCP14)
TFs	Transcription factors
TMD	Transmembrane domain
TSS	Transcription starting site

ABSTRACT

*Tef (*Eragrostis tef*), a cereal crop indigenous to Ethiopia, is essential in ensuring food and nutrition security. It is well-known for its gluten-free properties, rendering it suitable for individuals with celiac disease. However, tef is highly susceptible to lodging that negatively impacts its productivity. Efforts to develop varieties that are lodging-resistant have had limited success; particularly, because there have been a dearth of targeted genetic interventions. The aim of this study is to computationally predict and characterize six lodging-resistant genes (Rht1, BRI1, COMT1, CAD8C, CESA1 and CESA4) and the corresponding polypeptides of tef* vis-à-vis *three economically important cereal crops. The study began by selecting three economically important crops, namely: wheat, barley and rice and their six lodging-resistant genes. Using* in silico *probes, six homologous candidate lodging-resistant genes were retrieved from tef subgenomes. The candidate tef lodging-resistant nucleotide sequences were analyzed to computationally predict their gene structures and then comparatively analyzed against their homologues in the selected three crops. Putative polypeptide sequences of the six lodging-resistant genes, derived from their obtained predicted tef genes, were subjected to further amino acid level analysis to identify variations unique to tef, map the variations to specific functional or structural domains, predict the potential effects of the said variations and examine the possible impact of significant variations on the 3 dimensional structure of the proteins. Multiple sequence alignment of polypeptide sequences revealed distinct amino acid differences unique to tef in key functional and structural domains encoded by these lodging-resistant genes, with potential implications for tef's lodging susceptibility attributes. Variant effect predictions of counterfactual amino acid variations point to a subset of the unique tef variations having significant likelihood of causing damage in four of the six proteins: Rht1, BRI1, COMT1 and CESA1. Further examination of significant variation effects on the 3D structures of the four proteins revealed possible effects on the secondary structure of the polypeptides. Overall, the findings provide insights into several of tef's genetic variations related to lodging resistance features. Further experimental validation studies are needed to confirm these findings and elucidate the precise roles of these genes in tef's lodging susceptibility.*

Keywords/Phrases: Comparative analysis, *Eragrostis tef, In silico* probe, Lodging-resistant

1. INTRODUCTION

1.1 Background of the Study

Tef [*Eragrostis tef* (Zucc.) Trotter] is an allotetraploid ($2n = 4x = 40$) crop that is the staple food for more than 50 million people (Paff and Asseng, 2018). In Ethiopia, it is cultivated on more than three million hectares of land each year (Assefa *et al*., 2017). The cultivation of tef in Ethiopia predates historical records, existing before the introduction of wheat and barley. The grains of tef are more resistant to attack by storage pests compared to other cereals, and the seeds can be easily stored under local conditions without losing viability (Ketema, 1997).

Tef grains are rich in protein, vitamins, and minerals, including significantly higher levels of calcium and iron compared to wheat, barley, and rice. These gluten-free grains are safe for diabetics and individuals with immunological reactions to wheat and barley gluten. However, tef is quite prone to lodging, a condition where crops are permanently displaced from their natural vertical orientation (Alaunyte *et al*., 2012). As a result, up to 50% (Bennetzen *et al*., 2007) and an average of approximately 25% of tef yield is lost to the effects of lodging every year.

Scientific research on tef began in Ethiopia in the 1950s (Assefa *et al*., 2011), with early conventional breeding efforts focusing on germplasm enhancement through collection, characterization, evaluation, and conservation, as well as genetic improvement by selecting pure lines from existing germplasm (Tadele *et al*., 2018). Since the flower opening characteristics of tef were discovered in 1974, hybridization has been employed for tef improvement (Berehe, 1975). Molecular approaches, including marker development, genetic linkage maps, and genetic and molecular diversity analysis, were initiated between 1995 and 1998 (Assefa *et al*., 2011). Further advancements were made from 1998 to 2003, such as the initiation of interspecific hybridization, *in vitro* culture, and mutagenesis to improve disease- and lodging-resistance. Over the past two decades, there has been significant progress in understanding tef's genetic architecture and genomics (Belay *et al*., 2005). To increase the productivity of tef, 42 varieties improved by conventional breeding techniques and genetic transformation were released from federal and regional research centers (Merga, 2018).

Stem lodging in tef is closely linked to its height, as noted by Verma *et al*. (2005). The height of tef is determined by internode elongation, regulated by genes involved in gibberellic acid (GA) and brassinosteroid (BR) biosynthesis or signaling. GAs and BRs are crucial for tef growth and development; insufficient levels result in dwarf or semi-dwarf height, which

actually increases lodging tolerance. Similar to Rht1 in wheat and SLR1 (SD-1) in rice, which encode suppressor DELLA proteins affecting GA signaling or reducing GA synthesis, respectively, these semi-dwarf genes lead to the accumulation of DELLA proteins, resulting in the semi-dwarf phenotype (Liu and Fu, 2023).

The resistance of the stem against lodging stress depends significantly on its biochemical properties, particularly cellulose, hemicellulose, and lignin, known as structural carbohydrates (SC). The CESA genes, part of a multigene family that encodes cellulose synthases (CESs) crucial in cellulose biosynthesis, have been reported in arabidopsis, rice, maize, barley, and wheat. A functional cellulose synthase complex (CSC), essential for synthesizing cellulose in both primary and secondary cell walls, requires three distinct CESA proteins (Dong *et al.*, 2023).

Lignin's role in fortifying crop resilience against lodging primarily arises from its impact on cell wall thickness and its reinforcement of mechanical support within the stem. Various genes in the lignin synthesis pathway, such as caffeic acid O-methyltransferase (COMT) and cinnamyl alcohol dehydrogenase (CAD), have been identified as contributors to enhancing the mechanical strength of wheat stems. Low lignin or cellulose content leads to the brittleness of crop stems (Li *et al*., 2022).

Additionally, high temperatures, rainstorms and strong winds are external factors causing crop lodging. After a rainstorm, water infiltration softens the soil, weakens its ability to support root systems, and increases the weight of the plant's aboveground parts, making crops more susceptible to collapse under strong winds (Niu *et al*., 2016; Wu and Ma, 2018). Additionally, rain can dissolve organic or inorganic components within the plant, altering tissue osmotic pressure and reducing stem strength (Niu *et al*., 2016).

1.2 Statement of Problem

Tef is highly vulnerable to lodging, both in terms of shoot and root lodging. Numerous studies have been conducted to enhance lodging resistance in Ethiopian tef by investigating its biochemical and agro-morphological traits (for instance, Assefa *et al*., 2011, 2017; Bayable *et al*., 2020). Moreover, extensive research on economically significant cereal crops like wheat, barley and rice has led to the identification of genes potentially associated with lodging resistance (for instance, Nadolska-Orczyk *et al*., 2017; Shah *et al*., 2019; Niu *et al*., 2021).

Despite these efforts, the specific genes responsible for lodging resistance in tef have not been thoroughly examined. Consequently, there is a dearth of targeted point of genetic intervention to enhance lodging resistance in tef. Performing an *in silico* comparison between the polypeptide sequences of tef selected lodging-resistant genes to their homologous in wheat, barley and rice could potentially unveil important points of variation that contribute to tef lodging susceptibility. This approach holds the promise of shedding light on previously unknown avenues for improving lodging resistance in tef.

1.3 Objectives

1.3.1 General objective

The study aims to computationally find six lodging-resistant genes of *Eragrostis tef* and characterize the effects of their polypeptide level variations and predicted effects through *in silico* comparative analysis against three common cereal crops.

1.3.2 Specific objectives

- To computationally predict the homologous tef of six lodging-resistant genes and their putative proteins by *in silico* comparative analysis against wheat, barley and rice.
- Identify and catalogue unique tef amino acid variations of six lodging-resistant genes in contrast to their wheat, barley and rice homologous.
- To characterize the structural and functional aspects of uniquely tef amino acid variations of six lodging-resistant putative proteins referencing their corresponding wheat, barley and rice homologues.

1.4 Significance of the Study

Computationally identifying amino acid variations in six important tef lodging-resistant genes provides specific points of intervention for direct genetic manipulation, potentially enhancing lodging resistance in tef. As tef is an important staple crop in Ethiopia, and noting that lodging is the major factor for low productivity, improving lodging resistance will improve productivity and promote food security in the country.

2. LITERATURE REVIEW

2.1 Origin of Tef

Tef's diversity and origin are concentrated in Ethiopia and over time, the crop species has coevolved with Ethiopians. This is due to the fact that Ethiopia is home to a wide variety of crop species and is also thought to be the genesis of crop domestication, including the existence of many potential wild ancestors. In Ethiopia, where tef was domesticated prior to the birth of Christ, it is undeniable that it is a very old crop (Vavilov, 1951).

Eragrostis is one of the largest genera in the grass family, with over 350 species (Watson and Dallwitz, 1992). About 43% of these species are thought to have come from Africa, 18% from South America, 12% from Asia, 10% from Australia, 9% from Central America, 6% from North America, and 2% from Europe (Costanza *et al.*, 1979). Among the 54 species found in Ethiopia, 14 are endemic to the country (Cufodontis, 1974).

2.2 Taxonomy of Tef

All economically significant cereals, including tef, belong to the *Poaceae* family, commonly referred to as the grass family. The summary of various nomenclature names assigned to tef by different entities during different time periods (Table 2.1). However, the binomial nomenclature currently widely adopted is *Eragrostis tef* (Zucc.) Trotter. This name was proposed by Trotter in 1918 and is derived from the specific appelation "tef" previously used by Zuccagni. Tef falls under the tribe *Eragrostideae*, family *Poaceae* (formerly known as *Gramineae*), subfamily *Chloridoideae*, and genus *Eragrostis* (Assefa *et al*., 2017).

Table 2.1 Binomial nomenclatures given to tef by various establishments.

Suggested name	Year
Poa tef Zuccagni	1775
Poa abyssinica Jacquin	1781
Poa cerealis Salisb	1796
Cynodon abyssinicus (Jacq.) Rasp.	1825
Eragrostis abyssinica (Jacq.) Link	1827
Eragrostis pilosa (L.) P. Beauv. Subsp. *abyssinica* (Jacq.) Aschers and Graben	1900
Eragrostis tef (Zucc.) Trotter	1918

Source: Updated from Ebba (1975) and Ketema (1997).

2.3 Genetic Resources of Tef

The Ethiopian gene bank at the Ethiopian Biodiversity Institute (EBI) stores 5169 accessions as *ex situ* dried seed reserves for short-term storage at 4°C and long-term storage at -10°C. Among these, a total of 1854 accessions were added to the EBI gene bank through donations and repatriations, while 3315 accessions were directly collected by EBI (Tesema, 2013).

2.4 Importance of Tef in Ethiopia's Economy

Tef is the key crop in Ethiopia for both agricultural income and food security; 43% of farmers plant it. The output value of tef exceeds that of coffee, Ethiopia's primary export good. Tef is therefore, an important source of revenue, particularly for smallholder growers. However, it is frequently referred to as an "orphan crop" due to the little global scientific and funding attention it receives. This lack of attention results in it being under-researched compared to other important common cereals. Additionally, the existing seed system in Ethiopia is inadequate, providing insufficient quality and quantity of improved tef seeds, and the robust public agricultural extension system is ineffective at disseminating new technologies to farmers (Assefa *et al*., 2017). According to Central Statistical Authority (CSA) (2022), tef covers the largest cultivated area among cereals, but its yield per hectare is the lowest (Table 2.2).

Table 2.2 Cultivated area, gross grain production and average grain yield of cereals cultivated in Ethiopia (Averages of two years data of the 2021/2022 seasons).

Crop	**Cultivated area**		**Total grain production**		**Grain yield t/ha**
	Million ha	% of all cereals	Million t	% of the total of cereals	
Tef	3.094	26.699	57,34.	18.391	18.535
Maize	3.397	29.322	118.535	38.0124	34.885
Wheat	2.091	18.050	62.314	19.983	29.791
Sorghum	1.517	13.092	38.774	12.435	18.538
Barley	0.983	8.483	21.836	7.003	14.392
Millets	0.369	3.187	9.456	3.0327	25.602
Oats	0.038	0.331	0.658	0.211	17.154
Rice	0.961	0.832	2.907	0.933	30.140
Total	11.588		311.833		

Source: CSA, 2022 [t = tonne; ha = hectare].

While tef*'s* overall calorie, carbohydrate, and crude protein contents are comparable to those of other cereals, it is gluten-free, and has a higher concentration of fibre, protein, and minerals than other cereals. The high iron content of tef, which helps to prevent iron deficiency, is one of its health advantages. It is hence "nutrient-rich." Additionally, as tef is gluten-free, it can aid in the management of celiac disease (Vandercasteelen *et al*., 2018).

2.5 Genomic Relationship of Tef with Other Cereal Crops

Tef is the first whole genome sequenced member of the subfamily *Chloridoideae* under the grass family *Poaceae*. The waxy gene is crucial for the accumulation of amylose in cereal starch granules during their growth. Waxy gene has been used for phylogenetic tree construction of cereals due to its high variability, making it a suitable marker for studying genetic relationships among different species (Ma *et al*., 2013). Based on the waxy gene sequence, tef is phylogenetically closer to wheat and barley than rice, but its closest evolutionary cereal is finger millets, as shown in Figure 2.1 (Cannarozzi *et al*., 2014).

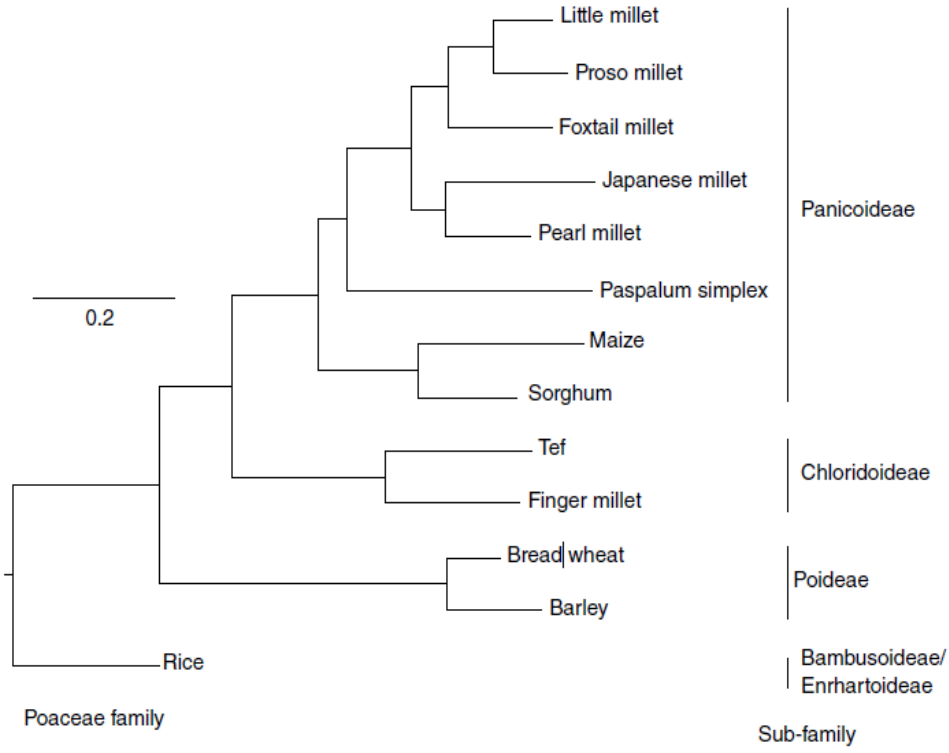


Figure 2.1. Phylogenetic tree displaying the relationships among various millets and major cereal crops worldwide based on the waxy gene. The binomial names and NCBI accession numbers of the plants used in this Figure are: little millet (*Panicum sumatrense* Roth ex Roem. & Schult., KC477404.1); proso millet (*Panicum miliaceum L.*, GU199266); foxtail millet (*Setaria italica (L.)* P. Beauvois, AB089143); Japanese millet (*Echinochloa esculenta*, AB668987.1); pearl millet (*Pennisetum glaucum (L.)* R.Br., AF488414); Paspalum simplex (AF318770); maize (*Zea mays L.*, EU041692); sorghum (*Sorghum bicolor (L.)* Moench, EF089839); tef (*Eragrostis tef* (Zucc.) Trotter, AY136939); finger millet (*Eleusine coracana Gaertn.*, AY508652); bread wheat (*Triticum aestivum L.*, KF861808); barley (*Hordeum vulgare L.*, X07931) and rice (*Oryza sativa L.*, FJ235770.1). The maximum likelihood tree was inferred using PhyML and the default model of HKY85 + G. The scale bar reflects evolutionary distance, measured in units of substitution per nucleotide site. Branch lengths reflect the estimated number of amino acid substitutions per site. ML bootstrap values were all 100% (Source: Cannarozzi *et al*., 2014).

2.6 The Problem of Lodging

Lodging, a mechanical stress experienced by plants, is a complex phenomenon shaped by a range of plant-related and environmental factors. Plant-related factors are related to particular variety or genotype and encompass morphological and biochemical traits. Specific morphological traits are often linked to lodging, resulting in the disruption of the upright position of shoots in small-grained cereals. Varieties with taller stature and weaker stems are more prone to lodging compared to semi-dwarf counterparts with stronger stems. On the other hand, environmental factors encompass elements such as light, wind, temperature,

rainfall, topography, soil type, nutrition, plant density, and diseases (Berry *et al*., 2004). Weather-related challenges compromise the ability of stems or roots to provide support, leading to susceptibility to lodging. The lack of adequate strength to withstand mechanical stress results in weak stems leading to the lodging of plants, as illustrated in Figure 2.2 (Kedisso, 2012).

Lodging has both direct and indirect consequences on crop productivity. The direct loss occurs when plants fall over, while the indirect loss stems from the underutilization of nitrogen fertilizer. Reduced stem strength leads to lodging, particularly after the application of high nitrogen fertilizers, which is counterproductive (Crook and Ennos, 1994).

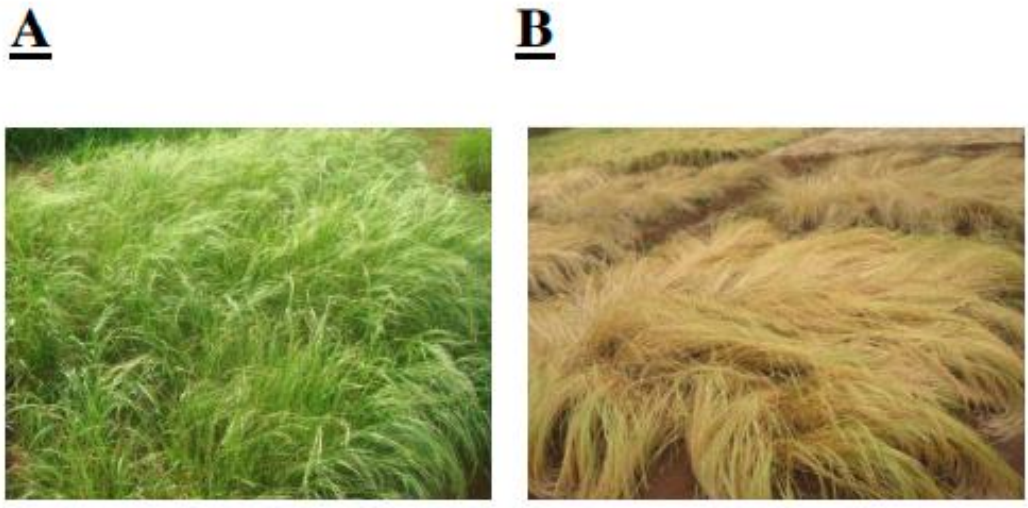


Figure 2.2. Illustrates the impact of a weak stem base, which lacks the necessary strength to support the shoot against mechanical stress, on *Eragrostis tef* plant stands in the field. The images show the plant stand at two stages: (A) during grain filling and (B) at maturity, where almost all the plants have lodged. This research was made at Holetta Agricultural Research Center in Ethiopia (Source: Gugsa *et al*., 2006).

2.7 The Role of Semi-Dwarfing Genes in Cereal to Enhance Lodging Resistance

The semi-dwarfing traits in wheat are attributed to mutations in the reduced height (Rht1) gene which codes for DELLA protein, leading to culm shortening (Gale and Marshall, 1976). Approximately 70% of modern wheat varieties contain at least one dwarfing gene, and these genes have been incorporated into various cultivars worldwide (Silverstone *et al*., 2001; Hedden, 2003).

In rice, the recessive semi-dwarf (SD-1) gene has been utilized to increase lodging resistance while reducing plant height. Native rice cultivars with the SD-1 mutation exhibited a semi-

dwarf phenotype, and this trait was introduced into improved rice lines in the 1960s (Ashikari *et al*., 2002).

Similarly, in barley, the value of shorter varieties was recognized early on. Varieties like 'Valticky,' derived from local landraces, replaced taller barley varieties in Moravia (Czech Republic) during the early 20th century (Bouma and Ohnoutka, 1991).

Since the Green Revolution, semi-dwarfing genes have been widely used in wheat and rice breeding due to their ability to increase output. Semi-dwarfs' slight (10–20%) drop in height is caused by a lack of endogenous GA's ability to promote growth; in wheat, this is caused by mutations at the Rht1 locus, which codes for DELLA, leading to "insensitive" growth. The semi-dwarfing mutations were originally spontaneous, and their continued widespread use in current types, some 50 years after their first introduction, is evidence of their agronomic significance (Chandler and Harding, 2013).

Mutations in the Rht, and SD-1 genes have played pivotal roles in achieving higher yields during the "green revolution" in wheat and rice (Hedden, 2003; Kashiwagi and Ishimaru, 2004; Yamaguchi, 2008). Consequently, average wheat yields have increased from 2.2 t ha^{-1} to 6.0 t ha^{-1}, while rice yields have risen from 1.5 t ha^{-1} to 4.2 t ha^{-1} due to the introduction of semi-dwarf varieties (Berry *et al*., 2004). Identifying, and incorporating stem height-controlling genes has been a significant factor in reducing lodging, and enabling increased fertilizer usage in wheat and rice since the 1970s (Hedden, 2003; Berry *et al*., 2004; Kashiwagi and Ishimaru, 2004; Tong, 2007).

2.8 Gibberellic Acid (GA)

2.8.1 Genetic control of biosynthetic or signaling pathways and lodging resistance

Gibberellic acids (GAs) play a critical role in regulating various aspects of plant growth, and development, particularly in stem elongation. The lengthening of internodes has significant agronomic implications, influencing crop biomass, panicle formation, and culm length (Ji *et al*., 2019). Research on mutants in rice, barley, and arabidopsis has revealed a shared mechanism controlling internode elongation involving genes linked to gibberellic acid (GA), and brassinosteroid (BR) biosynthesis or signaling (Niu *et al*., 2021). This is evident in mutants with either dwarf or semi-dwarf traits due to reduced levels of bioactive GAs, or taller crop heights due to increased bioactive GAs. Under challenging conditions, semi-dwarf rice mutants with GA deficiency or insensitivity exhibit improved resilience to lodging stress,

suggesting that reducing plant height through GA deficiency can enhance lodging tolerance (Achard and Genschik, 2009).

GA-deficient plants are more susceptible to breaking type lodging due to their shorter stature because they often have thicker, more rigid stems that are less flexible and more prone to breaking under stress, such as high winds or heavy seed heads. In shorter plants, mechanical stress is concentrated at specific points along the stem, making them more likely to snap rather than bend. Additionally, the reduced height decreases the plant's overall flexibility, preventing the stems from bending and swaying to absorb and dissipate energy, thus increasing the likelihood of breaking (Okuno *et al*., 2014).

The Rht1 protein, also known as DELLA, includes a specific DELLA domain. It is found widely across crop plants such as wheat (*Triticum aestivum*), rice (*Oryza sativa*), maize (*Zea mays*), grape (*Vitis vinifera*), barley (*Hordeum vulgare*), and soybean (*Glycine max*). Rht1 proteins are crucial negative regulators of GA signaling and belong to the GRAS [GAI (GA Insensitive), RGA (Repressor of GA1-3), and SCR (Scarecrow)] family of plant-specific nuclear proteins (Xue *et al*., 2022). The gene encoding Rht1, also called DELLA, contains essential domains: DELLA and GRAS.

The N-terminal DELLA domain of the Rht1 protein acts as a repressor of GA signaling. Additionally, the C-terminal GRAS domain, found in a family of proteins involved in transcriptional regulation, interacts with various regulatory proteins to suppress GA activity and signaling (Xue *et al*., 2022). Rht1 possesses a conserved C-terminal GRAS domain that provides its transcriptional regulatory function, classifying it as a member of the plant-specific GRAS subfamily of proteins. Its degradation in response to GA induction depends on the unique N-terminal DELLA domain (Dai and Xue, 2010; Huang *et al*., 2023).

The arabidopsis gibberellic acid insensitive (GAI) gene, maize dwarf-8 (d8) gene, rice SLR1 gene, and barley slender1 (SLN1) gene are considered orthologs of Rht1. These genes evolved from a common ancestral gene, with speciation giving rise to the wheat Rht-B1b (Rht1) and Rht-D1b (Rht2) Green Revolution genes (Figure 2.3) (Xue *et al*., 2022).

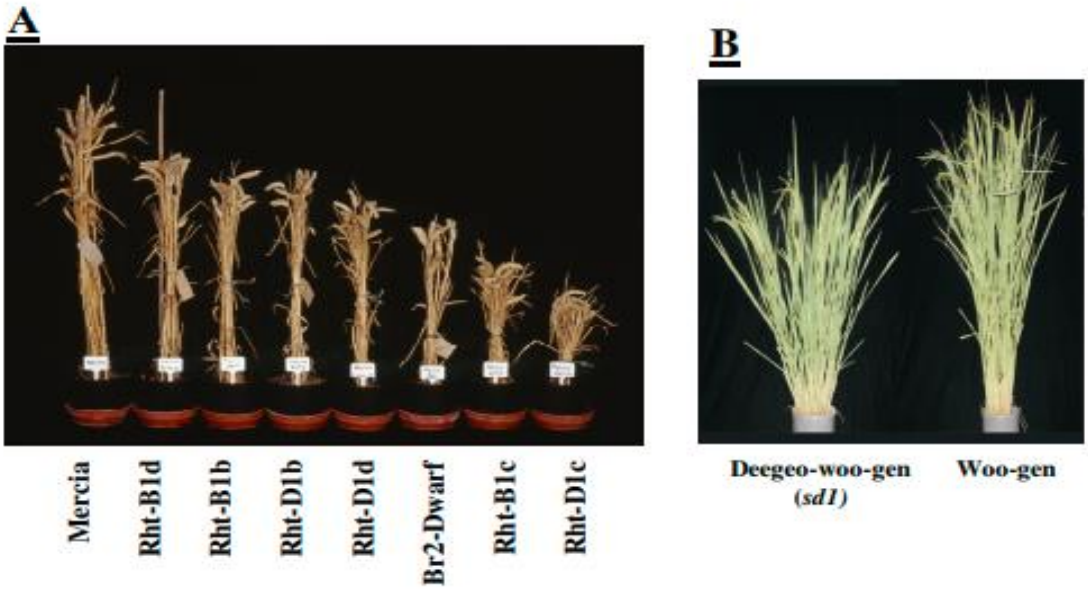


Figure 2.3. Two sets of phenotypic variations: A) showcases allelic diversity for semi-dwarfing traits associated with the wheat Rht gene, excluding the Br2-dwarf variant, which is characterized as brassinosteroid insensitive dwarf (Peng *et al*., 1999) and B) presents the phenotypic variations for the semi-dominant (SD-1) semi-dwarfing gene in rice (Source: Monna *et al*. (2002) and Spielmeyer *et al*. (2002).

2.8.2 Tef GA 20-oxidases

Tef GA 20-oxidases were studied by Beyene *et al*. (2022) through a comparative genomics analysis, which identified four tef GA 20 ox genes (EtGA 20ox1-4) and characterized them based on their homologous alleles. According to VanBuren *et al*. (2020), EtGA 20-ox tef genes were named based on their nucleotide and deduced amino acid sequence similarities to rice GA 20-ox genes. The nucleotide and amino acid sequence identity among the two species ranged from 90.8%–95.4% and 88.2%–92.3%, respectively. Comparatively, the wild-type rice SD-1 (OsGA20ox2) exhibited a close relationship with EtGA20-ox2 (EtSD-1). Beyene *et al*. (2022) targeted tef SD-1 (EtSD-1) for knockout mutation due to its similarity to the rice SD-1 gene, resulting in high-efficiency tetra-allelic mutations which led to reduced plant height.

2.9 Genetic Control of Brassinosteroid Biosynthetic or Signaling Pathways and Lodging Resistance

Brassinosteroids (BRs), a group of steroidal phytohormones, play a pivotal role in regulating various processes across the entire plant life cycle and aid plants in responding to abiotic stressors (Colebrook *et al*., 2014; Gruszka, 2020). The BR receptor BRI1 in wheat, barley, and rice consists of a putative signal peptide, two conservatively spaced cysteine pairs, a

domain with 22 leucine-rich repeats (LRR), a single transmembrane domain, and a kinase domain (Li and Chory, 1997; Yamamuro *et al*., 2000; Chono *et al*., 2003).

In rice, a 22 leucine-rich repeat (LRR) structural motif is present within the region spanning positions 90-676. These repeats fold to form the LRR domain, important in protein-protein interactions (Kajava and Kobe, 2002; Matsushima and Miyashita, 2012). The first few LRRs after the signal peptide are particularly crucial for the proper folding of BRI1 (Hou *et al*., 2019). Unique sequences in LRR22, just downstream of the island amino acids, are critical for BR-binding ability (Kinoshita *et al*., 2005).

An island between LRR21 and LRR22 contains 70 amino acids that fold back into the interior of the superhelix's core, interacting extensively with LRRs 13–25 through polar and hydrophobic interactions, forming a surface pocket that can bind brassinolide, crucial for BR signaling, as described in *Arabidopsis thaliana* (Matsushima and Miyashita, 2012). A similar characteristic is observed in OsBRI1, where the amino acid sequence between the 18th and 19th LRRs corresponds to the island domain (ID) in OsBRI1 (Li and Chory, 1997; Yamamuro *et al*., 2000). This ID is vital for brassinosteroid signal transduction; mutations in this region can disrupt BRI1 function (Li and Chory, 1997).

The kinase domain is essential for the trans-phosphorylation of BRI1 with BAK1 on multiple sites, fully activating BRI1 and initiating downstream signaling that leads to significant changes in nuclear gene expression (Hothorn *et al*., 2011).

In rice, the BR signaling pathway includes OsBRI1, its coreceptors BRI1-associated receptor kinase, Glycogen Synthase Kinases 1 and 2 (OsGSK1/2), and the Brassinazole-Resistant 1 (OsBZR1) transcription factor (Nakamura *et al*., 2006; Li *et al*., 2009). This complex controls gene expression through OsBZR1, a significant inhibitor of BR signaling (Bai *et al*., 2007). Application of exogenous BR can inhibit rice growth, as BR-deficient mutants exhibit shorter roots, reduced leaf size, and diminished plant height (Tong *et al*., 2014). Loss-of-function mutations in the OsBRI1 gene result in mutants with erect stature, semi-dwarfism, and BR insensitivity (Yamamuro *et al*., 2000; Hayat *et al*., 2011).

Overexpression of the wheat TaBRI1 gene in arabidopsis accelerates seed germination, earlier flowering, and increased seed output (Singh *et al*., 2016). In barley, several alleles of the homologous HvBRI1 gene, including the uzu1.a allele in Northeast Asian short culm cultivars and landraces, are known for their semi-dwarfing trait (Chono *et al*., 2003). In contrast, loss-of-function mutations in the OsBAK1 gene lead to upright leaves and BR

insensitivity, with minimal impact on plant height, reproduction, or grain yield. Wheat homologs of OsBAK1 belong to the SERK family proteins (Hayat *et al*., 2011). The BRI1 gene signaling pathway involves various proteins and molecules in signal transduction downstream of hormone receptors. Transcription factors activated by this pathway regulate genes involved in growth, development, and other physiological processes governed by brassinosteroids, as depicted in Figure 2.4.

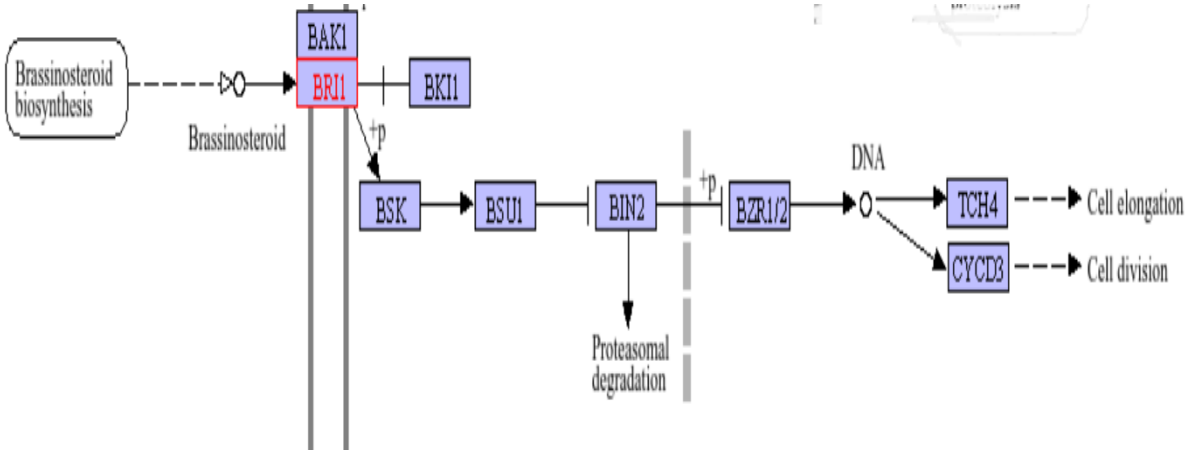


Figure 2.4. The BRI1 gene pathway (Source: KEGG PATHWAY: Plant hormone signal transduction - Oryza sativa japonica (Japanese rice) (RefSeq)).

2.10 Lignin Biosynthesis Gene's associated with Lodging Resistance

Lignin and cellulose, primary constituents of the cell wall, play a vital role in boosting plant vigor and providing defense mechanisms against both biotic and abiotic stressors, including plant lodging (Chen *et al*., 2011). Higher lignin content within the vascular bundles contributes to the increased physical strength of plant stalks. Strong correlations have been observed between total lignin levels in the basal second internodes of rice and wheat, and the brittleness and elasticity of their stems (Okuno *et al*., 2014; Zheng *et al*., 2017). Elevated levels of lignin accumulation have been linked to improved physical stability in wheat culm internodes (Peng *et al*., 2014).

During secondary cell wall production, lignin is deposited within the carbohydrate matrix of the cell wall, reinforcing the plant's structural integrity and facilitating upward growth (Del *et al*., 2012; Hyles *et al*., 2017). In mutant rice genotypes and wheat, the accumulation of cellulose, lignin, and hemicellulose enhances the strength of the culm and secondary cell wall. Increased lignin and cellulose content in rice cell walls contribute to enhanced lodging resistance (Shah *et al*., 2019).

Numerous studies have consistently shown a strong correlation between lodging resistance and increased levels of lignin, pectin, cellulose, and protein in plant stems. Lodging-resistant wheat varieties have higher levels of hemicellulose and lignin in their culms compared to susceptible ones (Berry *et al*., 2003; Chen *et al*., 2011). Research has revealed that the expression of wheat homologues of lignin biosynthesis genes is significantly more abundant in stem tissue than in other plant tissues such as the leaf sheath and leaf blade. Enzymes related to lignin synthesis, including phenylalanine ammonia lyase (PAL6), p-coumarate 3-hydroxylase (C4H), CoA ligase1 (4CL1), cinnamoyl-CoA reductase (CCR2), and ferulate 5-hydroxylase (F5H1 and F5H2), show a substantial correlation with lignin content (Bi *et al*., 2011).

Caffeic acid O-methyltransferase 1 (COMT1) is involved in lignin biosynthesis, catalyzing the methylation of hydroxylated monomeric lignin precursors. COMT plays a central role in lignin deposition through lignification, significantly increasing cell wall stiffness (Liang *et al*., 2022). The TaCOMT gene accumulates in the basal second internodes, leading to increased lignin and reduced lodging in wheat. Additionally, COMT displays consistent expression across stem, leaf, and root tissues (Zhou *et al*., 2009; Liang *et al*., 2022). The wheat COMT protein consists of single N-terminal dimerization domain (27–78 aas) that is involved in dimer formation and does not require metal ions for activity. The evolutionarily conserved C-terminal methyltransferase domain (121–341 aas) is the catalytic domain that consists of an S-adenosyl-L-methionine (SAM) and substrate binding sites (Naaz *et al*., 2013; Wang *et al*., 2018).

CCR1 and cinnamyl alcohol dehydrogenase 1 (CAD) transcripts are notably abundant in the stem, with higher enzyme activity compared to other tissues. During the heading stage, lodging-tolerant wheat cultivars exhibit increased transcript abundance of CCR1, COMT1, and CAD1 genes, and higher enzyme activities in the stem. CAD8C is involved in the last stage that is unique to the lignin monomer manufacturing biosynthetic process. It catalyzes the reduction of coniferaldehyde, 5hydroxyconiferaldehyde, sinapaldehyde, 4-coumaraldehyde, and caffeyl aldehyde to their respective alcohols in a NADPH-dependent mechanism. These alcohols serve as direct precursors for monolignol synthesis and their eventual polymerization into lignin (Tobias and Chow, 2005).

Enol alcohol + $NADP^+$ = Enol aldehyde + H^+ + NADPH *

* This reaction proceeds in the backward direction.

These factors correlate strongly with lignin content and stem mechanical strength (Li *et al*., 2022).

2.11 Cellulose Biosynthesis Genes

Plant cell walls require cellulose for various reasons. Cellulose, present in the basic cell wall, acts as a sturdy and flexible scaffold, supporting cell expansion while maintaining their unique three-dimensional shape. Additionally, cellulose provides mechanical strength to the secondary cell wall, formed after cell growth ceases, enabling it to withstand gravity and resist both biotic and abiotic pressures. Weak stems due to lack of cellulose are more prone to lodging, leading to decreased grain production (Houston *et al*., 2015).

The majority of plant vegetative biomass comprises the secondary cell wall, which is thicker and deposited within the primary cell wall. While the secondary cell wall forms as cell growth nears completion, the primary cell wall is deposited during cell division and expansion stages. Cellulose synthesis in plants is facilitated by multimeric protein complexes, comprising hexameric, rosette-like structures in the plasma membrane. These complexes, referred to as "cellulose synthase A" (CESA), with "A" denoting the catalytic subunit, consist of individual members responsible for cellulose synthesis (Kaur *et al*., 2016).

For instance, OsCESA1, OsCESA3, and OsCESA8 are subsequently primarily deposited in the primary cell wall, while OsCESA4, OsCESA7, and OsCESA9 are predominantly deposited in secondary cell walls in rice (Shah *et al*., 2019). Similarly, in barley, HvCesA1, HvCesA2, and HvCesA6 are involved in primary cell wall biosynthesis, whereas HvCesA4, HvCesA7, and HvCesA8 are associated with secondary biosynthesis. Mutations affecting the synthesis of primary and secondary cell walls in arabidopsis result in the collapse of xylem cells due to their inability to withstand the negative pressure generated by the transpiration stream (Taylor, 2008).

2.12 Hemicellulose Biosynthesis

Hemicellulose is a polysaccharide that is a component of the secondary cell wall structure. Its heterogeneous nature, including xylans, xylolucans, mannans and glucomannans, is important for its interactions with cellulose and linkages with lignin (Pękala *et al*., 2023). The fine structure hemicellulose of varies between different species. The major hemicellulose in the secondary cell wall is xylan, with the other sugars making up smaller proportions. In Arabidopsis, the IRX8 and PARVUS genes are important for synthesizing the tetrasaccharide end sequence putatively involved in xylan biosynthesis (Pauly *et al*., 2013).

Hemicellulose is synthesized in the Golgi by glycosyltransferases and transported for deposition in the cell wall where, along with cellulose and lignin, and provides structural support to plants (Scheller and Ulvskov, 2010).

2.13 Computational Approach

A computational approach utilizes computational methods, models, and algorithms to analyze and interpret biological data, model biological processes, and predict outcomes in biological systems (Gayathiri *et al*., 2023). The biological data required as input for such analysis are reposed in many publicly available databases. Primary among these is the website database NCBI (http://www.ncbi.nlm.nih.gov), which contains a variety of polynucleotide and polypeptide sequence data along with accompanying search and alignment tools such as Basic Local Alignment Search Tool (BLAST). In addition, UniProt (https://www.uniprot.org) serves as the primary repository for collating and linking information from diverse and extensive sources, constituting the most exhaustive collection of protein sequences, functional annotations and 3 dimensional structures of polypeptides. It leverages the robust bioinformatic infrastructure and scientific proficiency of the European Bioinformatics Institute (EBI), Protein Information Resource (PIR), and Swiss Institute of Bioinformatics (SIB), and it is freely accessible and user-friendly for researchers (UniProt Consortium, 2007).

In addition to the extensive sequence databases, NCBI also offers BLAST (Altschul *et al*., 1990) that serves to conduct local pair-wise searches of the database using a query sequence. The platforms behind BLAST algorithm allows to narrows the search space by taxa, type of sequence dataset as well as adjust parameters to optimize the search for a match to a query sequence. There are also several algorithms hosted by various entities for performing multiple sequence alignments. One of the most popular, in this regard, is Clustal Omega (Madeira *et al.,* 2019), which utilizes a progressive alignment approach to bring the most characters in multiple sequences into alignment. European Bioinformatics Institute's (EBI), hosts Clustal Omega among many other algorithms that perform multiple sequence alignments.

Linux (Torvalds, 1969) is platform that provides a powerful and flexible command line-based string analysis of sequence data. The functionalities of Linux are character counting, pattern searching and editing. Furthermore, many open-source bioinformatics applications are developed and maintained for the Linux environment. The advent of genome sequencing has ushered in the development of gene prediction tools by computational methods.

Generally, there are two classes of computational gene prediction methods: similarity based and *ab initio.* Gene prediction in eukaryotes is particularly challenging because of low density of genes and the presence of introns in protein coding genes. The most recent generation of *Ab initio* gene prediction programs rely on recent advances in Support Vector Machine (SVM), Hidden Markov Model (HMM) and Neural Network (NN) to differentiate coding from non coding regions. *Ab initio* gene prediction programs currently widely used are Genscan (Burge and Karlin, 1997), Augustus (Stanke and waack, 2003), MAKER (Cantarel *et al*., 2008), GLIMMER (gene locater and interpolated Markov model ER) (Delcher *et al*., 2007) and FGENESH (Solovyev *et al*., 2006). Of those genes prediction programs listed GLIMMER is a tool suited for microbial genome, while FGENESH utilizes HMM and is the most efficient and precise gene finder available (Sashankar *et al*., 2021). FGENESH is freely available online at softberry.com.

Biological macromolecules such as DNA and protein contain information about the evolutionary patterns and relationship that can be constructed in to trees that depict the evolutionary history. Molecular phylogeny serves as a fundamental tool for understanding the evolutionary history of species. Phylogenies are typically represented as trees, where nodes represent taxa and edges represent genetic or evolutionary distances between taxa (Chikkagoudar, 2010). There are two broad approaches to undertaking phylogenetic analysis: character based and distance based. However, there are many various methods that followed different principles that output differing tree topologies with varying degree of confidence. Among those that distance based the three major methods of phylogenetic tree construction are Neighbor Joining (NJ) (Saitou and Nei, 1987), Maximum Likelihood (ML) (Felsenstein, 1981) and Bayesian (Ronquist and Kuelsenbeck, 2003). The ML and Bayesian methods outperform the NJ method in generating more accurate phylogenetic trees because they use rigorous statistical models and probabilistic frameworks that account for various factors in sequence evolution. In contrast, the NJ method relies on pairwise distances without utilizing these advanced models, leading to less accurate representations of evolutionary relationships (Guindon and Gascuel, 2003). However, ML and Bayesian methods require a large amount of computational time compared to the simpler NJ method. An efficient implementation of the NJ method is a component of MEGA (Tamura *et al*., 2011).

The Sorting Intolerant from Tolerant (SIFT) (Ng and Henikoff, 2003) algorithm was among the first web server tools to predict amino acid substitution on protein function. It has gained renown as a technique for characterizing missense variation because of their significant

impact on protein function and their relevance in medical, agricultural, and evolutionary research. SIFT computes a score that it uses to build a receiver operating characteristic curve upon which it determines whether an amino acid substitution to categorize them as tolerated them or deleterious. It has been used to assess the effects of mutations in agricultural plants (for instance, Till *et al.*, 2007), arabidopsis (for instance, Günther and Schmid, 2010), and animals model organisms (for instance, Guryev *et al*., 2004).

Proteins are versatile amino acid polymers that organize into unique three-dimensional arrangement that is essential for their cellular function. Understanding the spatial arrangement of the constituent amino acids of a protein, therefore, opens an opportunity to examine how changes in amino acid(s) can affect the 3D structure of the protein, and subsequently, its function. SWISS-MODEL workspace (Bordoli *et al*., 2009), is an integrated fully automated web-based homology protein modeling expert system accessible via Expasy (http://swissmodel.expasy.org). It draws upon a library of experimentally determined protein structures to serve as appropriate templates for a specified target protein for which a predicted 3D structure is sought. A three-dimensional model of the target protein is created based on a sequence alignment between the target protein and the template. The reliability of the final model is estimated using model quality evaluation techniques. Currently, SWISS-MODEL homology modeling is widely employed in numerous biological applications producing trustworthy structural models.

3. MATERIALS AND METHODS

3.1 Selection and Extraction of Reference mRNA and Coding Sequences

The selection of the most economically significant cereal crops, namely wheat, barley and rice, forms the basis of this study. This is because these crops have been extensively studied in terms of their lodging resistance genes and their genome has been sequenced. Among the many genes that contribute to lodging-resistant traits, it was decided to study six such genes, namely: Rht1, BRI1, COMT1, CAD8C, CESA1 and CESA4, because of numerous research works demonstrating their role in lodging resistance. These genes are known to be involved in various morphological and biochemical aspects of plant architecture, cell wall composition, or hormone signaling pathways, all of which are known to influence lodging resistance. In the second step, six lodging-resistant genes were selected for their morphological and biochemical manifestations on lodging charactersitics, as described in the research articles and review papers (Nadolska-Orczyk *et al*., 2017; Shah *et al*., 2019). The six reference mRNA sequences corresponding to the selected genes in wheat (*Triticum aestivum L*) (Ta) (taxaid: 4565) were acquired from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) for further downstream analysis.

In order to acquire the homologous of the six wheat lodging-resistant genes using the ref mRNA in barley (*Hordeum vulgare)* (Hv) and rice (*Oryza sativa*) (Os) from NCBI ref mRNA dataset, default parameters of the nucleotide BLAST algorithm (Altschul *et al*., 1990) were used; including a maximum of 100 aligned sequences, an expectation value of 0.05 for matches in a random model, a seeding alignment length of 28 nucleotides, and the restriction of matches to a query range. Additionally, a reward of 1 for match and a penalty of -2 (in a linear gap penalty scheme) for mismatching bases were applied. The linear gap model applied gap costs were used to establish and extend gaps in the alignment. Megablast was used to calculate linear costs based on match/mismatch scores. The sequences were masked and filtered for low complexity regions. The searches were limited to specific datasets of barley (taxaid: 112509) and rice (taxaid: 4530). At the end of this exercise, a total of six for three taxaid (a total of 18) lodging-resistant genes ref mRNA sequences were retrieved from the NCBI.

The CDSs of the 18 mRNA sequences described above, corresponding to lodging resistance genes in wheat, barley and rice, were extracted using the exon-intron boundary information provided by NCBI GenBank annotation. These CDSs provided complete sequences that

spanned from the start to the end of translation including all exons but excluding introns, along with their corresponding polypeptide sequences. To support subsequent downstream investigations, the sequences were archived in FASTA format.

In addition, the accession numbers of the ref mRNA sequence of wheat, barley and rice of each lodging-resistant gene and its associated relevant metadata were retrieved. This metadata includes details such as coding sequence length, exon-intron structure (representative of wheat), and the categorized role (morphological and biochemical) of the identified proteins in relation to phenotypic traits (Table 3.1).

Table 3.1 The taxaid of wheat, barley and rice, and accession number of their lodging-resistant gene (and their aliases), along with the length (in bps) of their corresponding mRNA and CDS and traits and characteristics they influence. The representative exon-intron structure of the corresponding wheat gene is shown.

Affected traits	Wheat (taxaid: 4565)	Barley (taxaid: 112509)	Rice (taxaid:4530)	Characteristics
Morphological features	Name: Rht1 Acc.No: XM_044512168.1	Name: SLN1 (SD-1) Acc.No: XM_045123987.1	Name: SLR1 Acc.No: NM_001418508.1	Reduced plant height
	mRNA = 2545 bp CDS: 192...2057 (1866 bp)	mRNA = 2574 bp CDS: 228...2084 (1857 bp)	mRNA = 2545 bp CDS: 268...2145 (1878 bp)	
	Gene structure* (*Number of exons=1*) 			
	Name: BRI1 Acc.No: XM_044493483	Name: BRI1 Acc.No: XM_045120961.1	Name: BRI1 Acc.No: XM_015765544	
	mRNA = 3890 bp CDS: 151...3513 (3363 bp)	mRNA = 3593 bp CDS: 207...3563 (3357 bp)	mRNA = 2545 bp CDS: 153...3518 (3363 bp)	
	Gene structure* (*Number of exons=1*) 			
Biochemical features	Name: COMT1 Acc.No: XM_044574894	Name: COMT1 Acc.No: XM_045103585.1	OsCOMT1 Acc.No: NM_1403482	Lignin and stalk lodging mechanical strength
	mRNA = 1476 bp CDS: 104...1186 (1083 bp)	mRNA = 1454 bp CDS: 89...1171 (1083 bp)	mRNA = 1514 bp CDS: 123...1229 (1107 bp)	
	Gene structure* (*Number of exons=2*) 			
	Name: CAD8C Acc.No: XM_044541461.1	Name: CAD8C Acc.No: XM_045092878.1	Name: CAD8C Acc.No: NM_001419168.1	
	mRNA = 1649 bp CDS: `124..1389` (1266 bp)	mRNA = 1650 bp CDS: `163..1428` (1266 bp)	mRNA = 1686 bp CDS: `216..1310` (1095 bp)	
	Gene structure* (*Number of exons=3*) 			
	Name: CESA1 Acc.No: XM_044592776.1	Name: CESA1 Acc.No: XM_045110835.1	Name: CESA1 Acc.No: NM_001402501.1	Primary cell wall formation

	mRNA = 3853 bp CDS: 284...3526 (3231 bp)	mRNA = 3605 bp CDS: 61...3303 (3237 bp)	mRNA = 4082 bp CDS: 292...3522 (3231 bp)	
	Gene structure* (*Number of exons=14*) 			
	Name: CESA4 Acc.No: XM_044485391.1	Name: CESA4 Acc.No: XM_045121113.1	Name: CESA4 Acc.No: NM_001405442	Secondary cell wall formation
	mRNA = 3425 bp CDS: 99...3239 (2925 bp)	mRNA = 3473 bp CDS: 152...3286 (2949 bp)	mRNA = 3479 bp CDS: 132...3101 (3231 bp)	
	Gene structure* (*Number of exons=13*) 			

* Representative gene structure is that of wheat. [*Not scaled to length of mRNA*].

3.2 *In silico* Probe Design

In order to computationally fish out the six candidate homologues lodging-resistant gene in tef, it was necessary to design *in silico* probes. The design of *in silico* probes is a key step aimed at optimizing computing time and eliminating redundant processes in downstream analysis. Specifically, six *in silico* probes sequences were designed one for each set of three homologous lodging-resistant genes exclusively using CDSs of wheat, barley and rice. For this purpose, Clustal Omega, a multiple sequence alignment program, was employed with default parameters (Madeira *et al.,* 2019**).** Of the output of the multiple sequence alignment (MSA), only those perfectly conserved matching sequences across all species with minimum length of 23 bps or longer were selected as potential *in silico* probes. Additionally, the reverse complements of these *in silico* probes were developed using Python codes in the Jupyter Notebook, specifically in the Anaconda3 environment (Wang and Oliphant, 2012) (see Annex I**).**

Subsequently, 15-base pair sliding window probe sub-sequences was generated from both the forward and reverse complement of the candidate *in silico* probes. In the Linux environment (Torvalds, 1969), seqkit tool 'split -W 15 -O *directory_name filename.extension*' command were used to split the *in silico* probes into smaller sub-sequence probes with size of 15 bps (Shen *et al*., 2016) (see Annex II**)**. The formula used for determining the new window probe sub-sequences was: New window = old window - new window + 1.

Developing 15 bps sliding window *in silico* probes sub-sequences in the forward and their reverse complements was done to minimize false positives and enhance the likelihood of identifying matching sequences in tef. This approach makes the analysis more robust and enables a more thorough exploration of the potential homologous candidate lodging-resistant genes in tef.

3.3 *In silico* Probe Mapping to Exons

Mapping of *in silico* probes to exons were determined by aligning wheat's six lodging-resistant gene sequences with their corresponding CDS, facilitated by DotLet's dot plot analysis (Junier and Pagni, 2000). By virtue of identifying the exon location of the *in silico* probes on their respective CDSs, a follow-up multiple sequence alignment, with default parameter (Madeira *et al*., 2019), was undertaken of each set of identified exon, CDS and *in silico* probe. This alignment allowed the approximation of the length of the regions spanning 5' and 3' of the *in silico* probes that covers the entire breadth of the genes. Since the lengths of the six lodging-resistant genes, as well as the CDS sequences, of wheat, barley and rice are approximately equal length, and have a very similar in exon-intron structure, it is a reasonable assumption that the estimated lengths of the genes 5' and 3' of the *in silico* probes will also hold true for tef.

3.4 Retrieval of Tef Genome Sequence

The ref genome sequence of *Eragrostis tef* of subgenomes 1A-10A and subgenomes 1B-10B dabbi cultivar (VanBuren *et al.,* 2020) were downloaded using wget command line on the Linux terminal (Torvalds, 1969) from NCBI (Annex III).

3.5 Candidate Homologous Lodging-Resistant Genes and Their Putative Proteins in Tef

In order to find the tef homologous of the six lodging-resistant genes, all the *in silico* probes' 15 bps sub-sequences, for both the forward and reverse complement, of each gene was exhaustively queried against tef subgenomes 1A-10A and 1B-10B. To achieve this, "grep" pattern identification command was executed on a Linux platform (Torvalds, 1969) (see Annex IV).

Based on the maximum length match of the *in silico* probes sub-sequences, the location of the candidate tef homologues of the six lodging-resistant candidate genes' on the tef subgenomes were determined. Relative to the expected lengths of the six lodging-resistant genes, the *in silico* probes were relatively short. However, the approximate length the genes 5' and 3' of the *in silico* probes estimated from the wheat genes was also used to approximate the length of the tef sequence that extends 5' and 3' of the *in silico* probes location on the tef subgenomes. To extract the estimated entire tef homologous genes, the 'grep' command options, "--before-context=*length*" and "--after-context=*length*" (where 'length' provides the

length 5’ or 3’ of the *in silico* probes) that utilized the approximate lengths estimated from the wheat's homologous lodging-resistant genes (see Annex IV).

The subgenome sequence whose approximate lengths is estimated to cover the entirety of the tef lodging-resistant genes were interrogated using the gene finding program of Softberry (Solovyev *et al*., 2006) with the parameter matrix option set to "*Eragrostis tef* (tef)". The homologous candidate tef lodging-resistant genes’ exon-intron structures were predicted using Softberry’s FGENESH-HMM gene finder program. The program also generated the predicted genes’ corresponding putative peptides sequences. The outputs from FGENESH-HMM were retained for further downstream characterization.

3.6 Structure and Characteristics of Six Candidate Homologous Tef Lodging-Resistant Genes Compared to Wheat, Barley and Rice

In order to confirm the validity of the candidate tef lodging-resistant genes, it was necessary to compare them to their corresponding homologous in wheat, barley and rice. In addition to the exon-intron structure, length and number of exons and their GC composition, the phylogenetic relationship among the orthologues in the four species was determined using a ML approach; and the accompanying branch length estimated.

3.7 Comparison of Polypeptide Sequences of the Six Lodging-Resistant Genes and Identification of Variations Unique to tef

In order to determine the percent identity putative tef polypeptide sequences with that of their homologous of wheat, barley and rice, the sequences were aligned using the MSA tool Clustal Omega (Madeira *et al.*, 2019). Also, based on the MSA output amino acid variations unique to tef (*i.e.* conserved in the other three species) were identified and catalogued. These unique variations were then mapped to their functional or structural domain of the proteins.

3.8 Characterization of Functional Consequences and Variation Effect Prediction of Unique Tef Variations

The unique tef amino acid variations, which are otherwise perfectly conserved in wheat, barley and rice, were mapped to the protein sequences. Those amino acid variations located in known functional or structural domains of the respective proteins were filtered in for further in-depth analysis. Hence, the filtered unique tef variations were further subjected to qualitative and quantitative examination to elucidate their potential consequences on the structure and function of the proteins.

Qualitative analysis involved examination of the individual amino acid variations with respect to their physico-chemical properties *vis-à-vis* the amino acid conserved in wheat, barley and rice. Quantitative analysis utilized the SIFT variation effect prediction model (Ng and Henikoff, 2003), in a counterfactual approach where the identified tef amino acid variation(s) was/were altered to the amino acid conserved in the other three species and interrogated to determine its effect if it had remained as is. This alteration was examined to determine its predicted effect had it remained unchanged.

3.9 Predicted 3-Dimensional Structure of Tef Putative Lodging-Resistant Polypeptides

Aside of tef Rht1 polypeptide, whose predicted 3-dimensional (3D) structure is already found in the UniProt database (https://www.uniprot.org), the 3D structures of the computationally derived polypeptide sequences of tef BRI1, COMT1, CAD8C, CESA1, and CESA4 were generated using the online SWISS MODEL (http://swissmodel.expasy.org). The wheat and rice homologous polypeptide structures were obtained from UniProt database for comparison with tef predicted 3D structures. Only the effects of those variations with significant SIFT scores of less than or equal to 0.1 (*i.e.* could manifest damaging phenotypes) were considered to evaluate potential structural changes.

4. RESULTS

4.1 Lodging-Resistant Homologous Genes in Wheat, Barley and Rice

This study benchmarks economically significant cereal crops, specifically wheat, barley and rice, which were subjected to extensive investigation to comprehend their lodging-resistant genes. Further, the study selected six genes whose importance for lodging resistance in the benchmarked three cereal crops is well documented.

4.2 *In silico* Probes' Mapping to Exons

The six *in silico* probe obtained from perfectly conserved regions across the six lodging-resistant genes in the three cereal crops, both forward and reverse complements, measured in base pairs (bps) are presented in Table 4.1. All *in silico* probes were equal to or longer than 23 bps, with the longest, at 53 bps for Rht1.

Table 4.1 Forward and reverse complement *in silico* probes.

No	Lodging-resistant gene	Orientation*	*In silico* probes' sequences 5' to 3'	Probe length (in bps)
1	Rht1	F	GGCACGGACCAGGTCATGTCCGAGGTGTACCTCGGCCGGCAGATCTGCAACGT	53
		R	ACGTTGCAGATCTGCCGGCCGAGGTACACCTCGGACATGACCTGGTCCGTGCC	
2	BRI1	F	GAAATGGAGACCATTGGCAAGATCAAACACCG	32
		R	CGGTGTTTGATCTTGCCAATGGTCTCCATTTC	
3	COMT1	F	GACGGCGGCATCCCGTTCAACAAGGCGTACGGGATG	36
		R	CATCCCGTACGCCTTGTTGAACGGGATGCCGCCGTC	
4	CAD8C	F	CAGATCGAGGTCGTCAAGATGGACTACGTCAACCAGGC	38
		R	GCCTGGTTGACGTAGTCCATCTTGACGACCTCGATCTG	
5	CESA1	F	AATGAACAGTTCTGGGTCATTGG	23
		R	CCAATGACCCAGAACTGTTCATT	
6	CESA4	F	GTGCTCTGGTCTGTCCTGCTCGCCTCC	27
		R	GGAGGCGAGCAGGACAGACCAGAGCAC	

*F represents 'forward' and R represents 'reverse complement'.

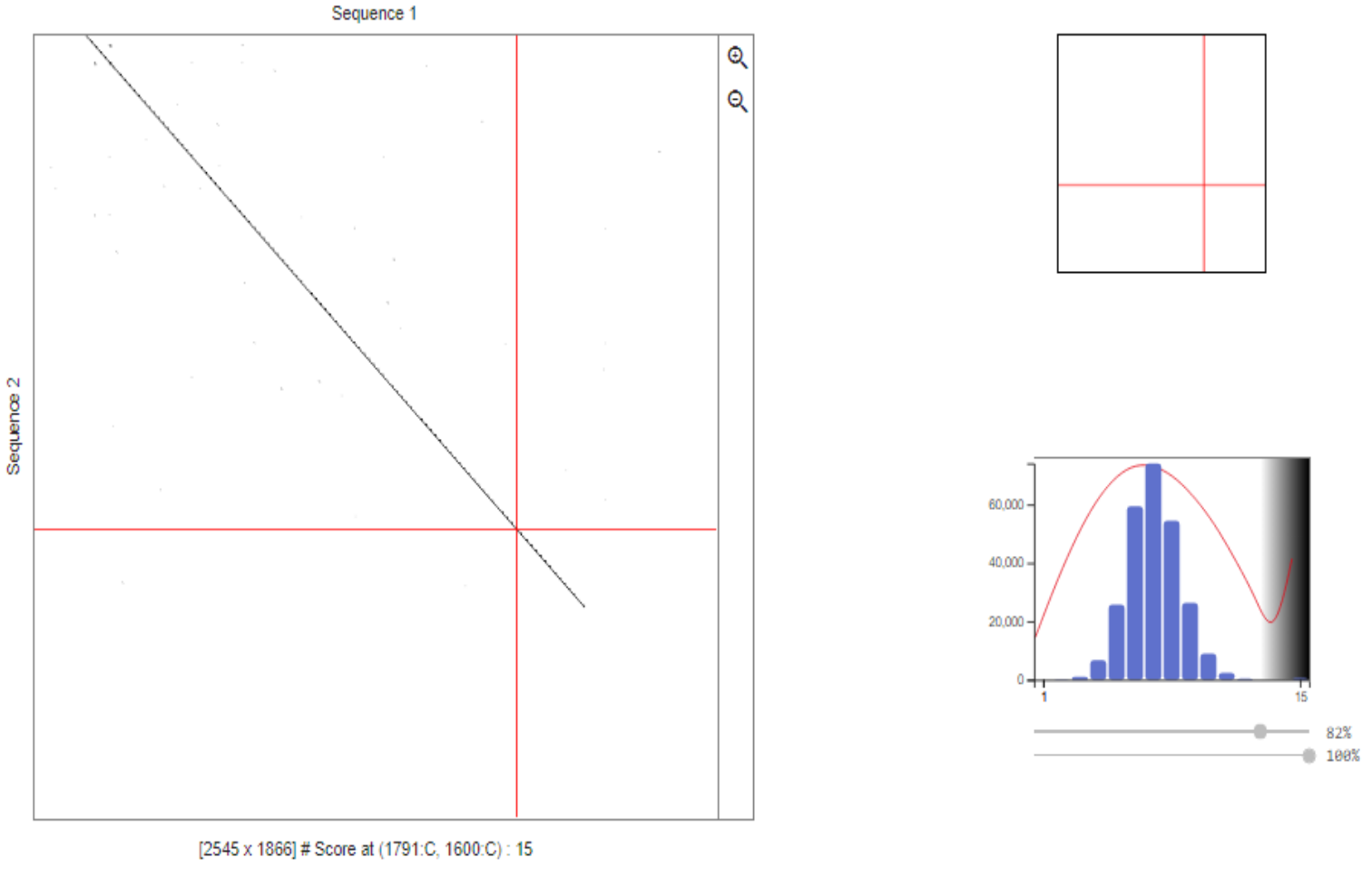
Using the CDS sequences as a guide, the six forward *in silico* probes were mapped to the exons of their respective reference wheat gene sequences, and hence, the precise exonic locations of the *in silico* probes were determined. The exact locations of these *in silico* probe sequences were mapped within the exonic regions of each lodging-resistant gene in wheat. All *in silico* probes were equal to or longer than 23 bps, with the longest, at 53 bps for Rht1, as presented in Figure 4.1.

The Rht1 gene in wheat has a single exon and the forward *in silico* probe was near the 3' end (Figure 4.1). It spanned 53 bps, and begins 1762 bps and 1571 bps downstream from the transcription start site and translation start site, respectively. Similarly the BRI1 gene has a single exon and the forward *in silico* probe of 32 bps begins at 2705 bps within the gene and at 2555 bps within the CDS (Figure 4.1 and Table 4.2).

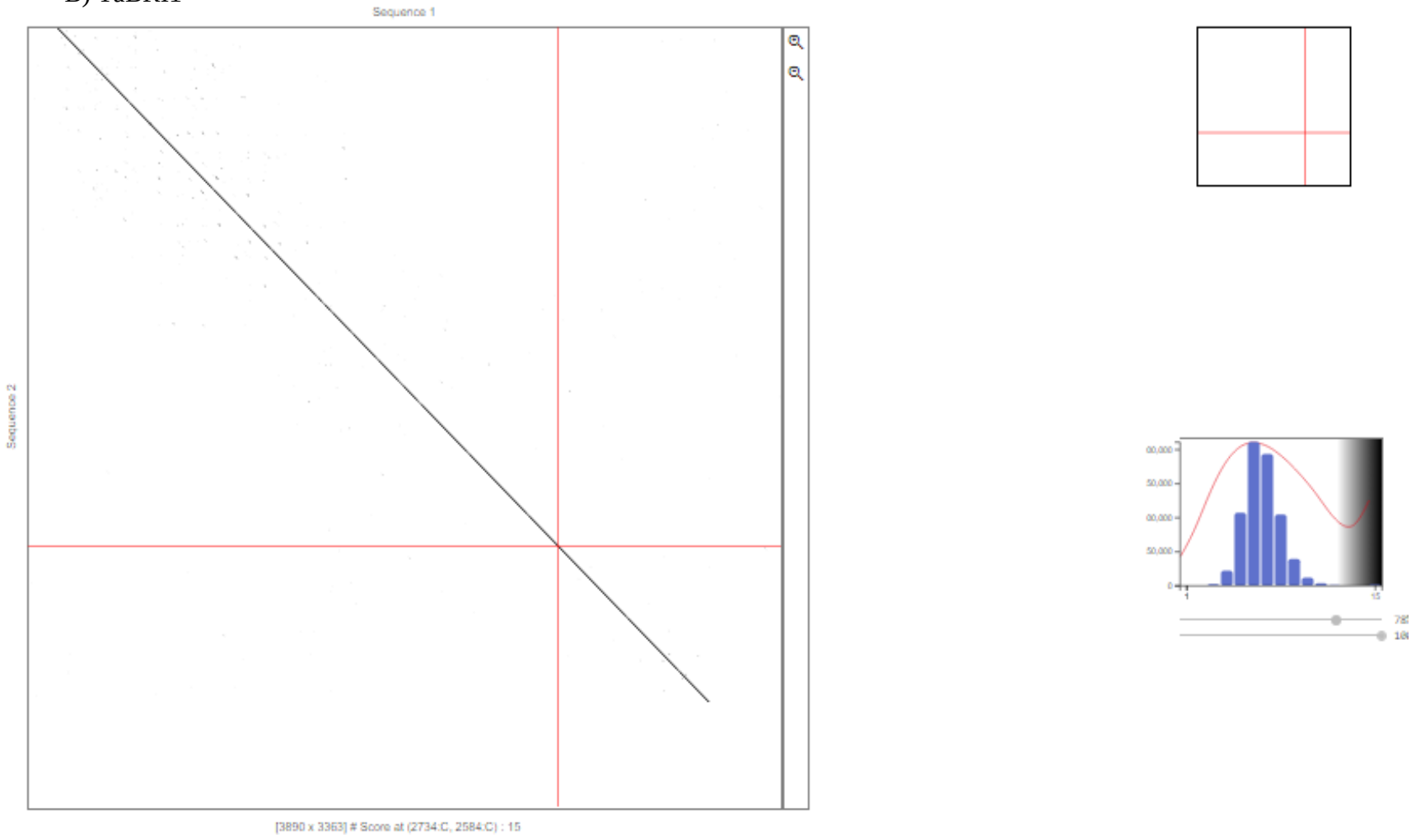
The 36 bps long *in silico* probe of the COMT1 gene is found in the last, and also the second, exon of the reference wheat gene (Figure 4.1). It begins at position 2272 bps within the gene and at 464 bps within the CDS. The 38 bps long *in silico* probe for the CAD8C gene in wheat was located near the 3' end of exon 3, beginning at 3012 bps within the gene and at 1195 bps within the CDS (Figure 4.1 and Table 4.2).

Additionally, the 23 bps long *in silico* probe of the CESA1 gene is found in exon 14 of the reference wheat gene (Figure 4.1). It begins at position 5003 bps within the gene and 2774 bps within the CDS, as presented in Table 4.2. Lastly, the 27 bps *in silico* probe for the CESA4 gene in wheat was found near the 3' end of exon 13, beginning at 5264 bps within the gene and at 2858 bps within the CDS. The information obtained from mapping the exonic location of the *in silico* probes was used to estimate the length of the genes both 5' and 3' of the *in silico* probes (Figure 4.1 and Table 4.2).

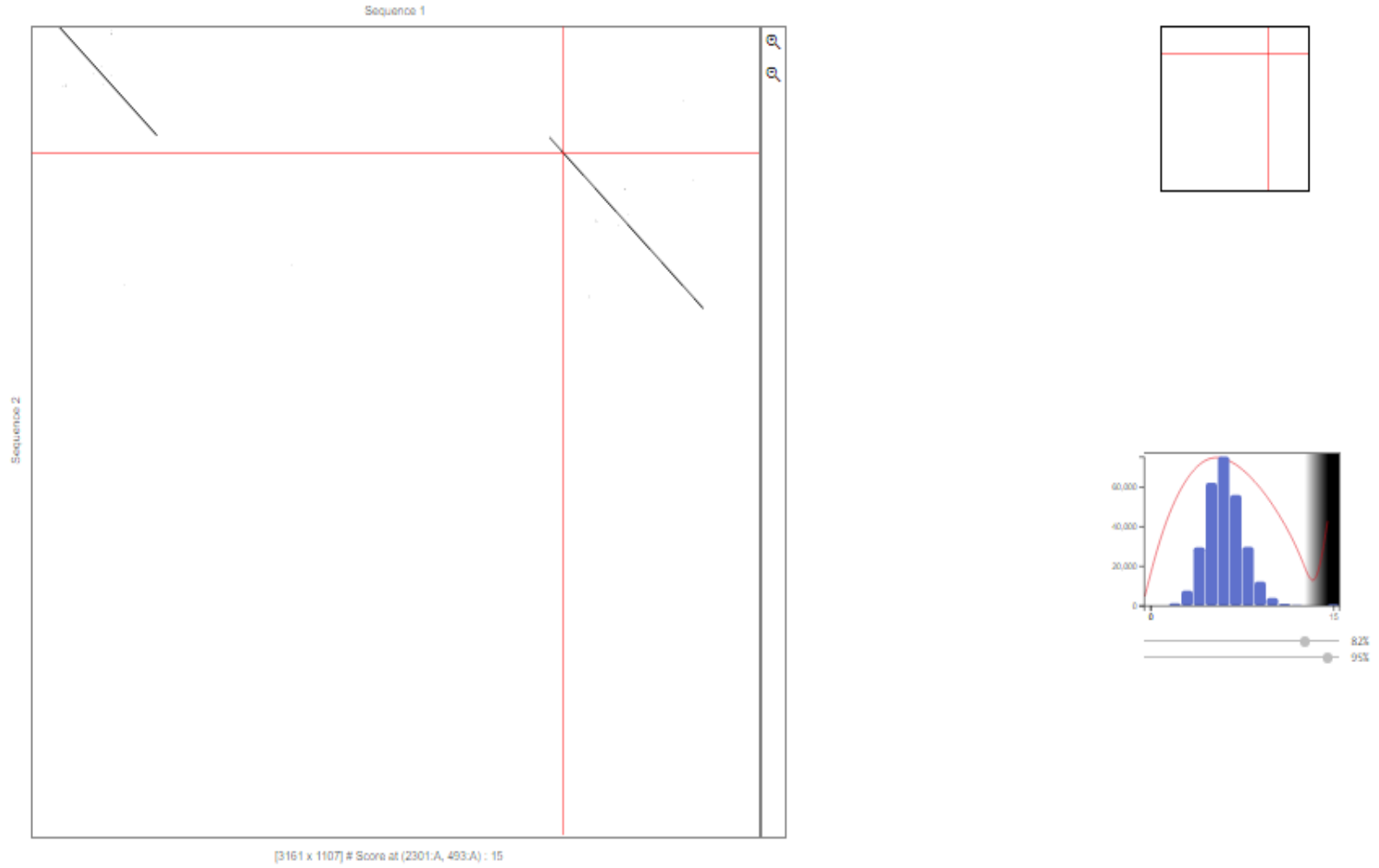
A) TaRht1


Sequence 1
Sequence 2
[2545 x 1866] # Score at (1791:C, 1600:C) : 15
60,000
40,000
20,000
0
1
15
82%
100%
Seq1:1791
GGCACGGACCAGGTCATGTCCGAGGTGTACCTCGGCCGGCAGATCTGCAACGTGGTGGCCTGCGAGGGGGCGG
GGCACGGACCAGGTCATGTCCGAGGTGTACCTCGGCCGGCAGATCTGCAACGTGGTGGCCTGCGAGGGGGCGG
Seq2:1600

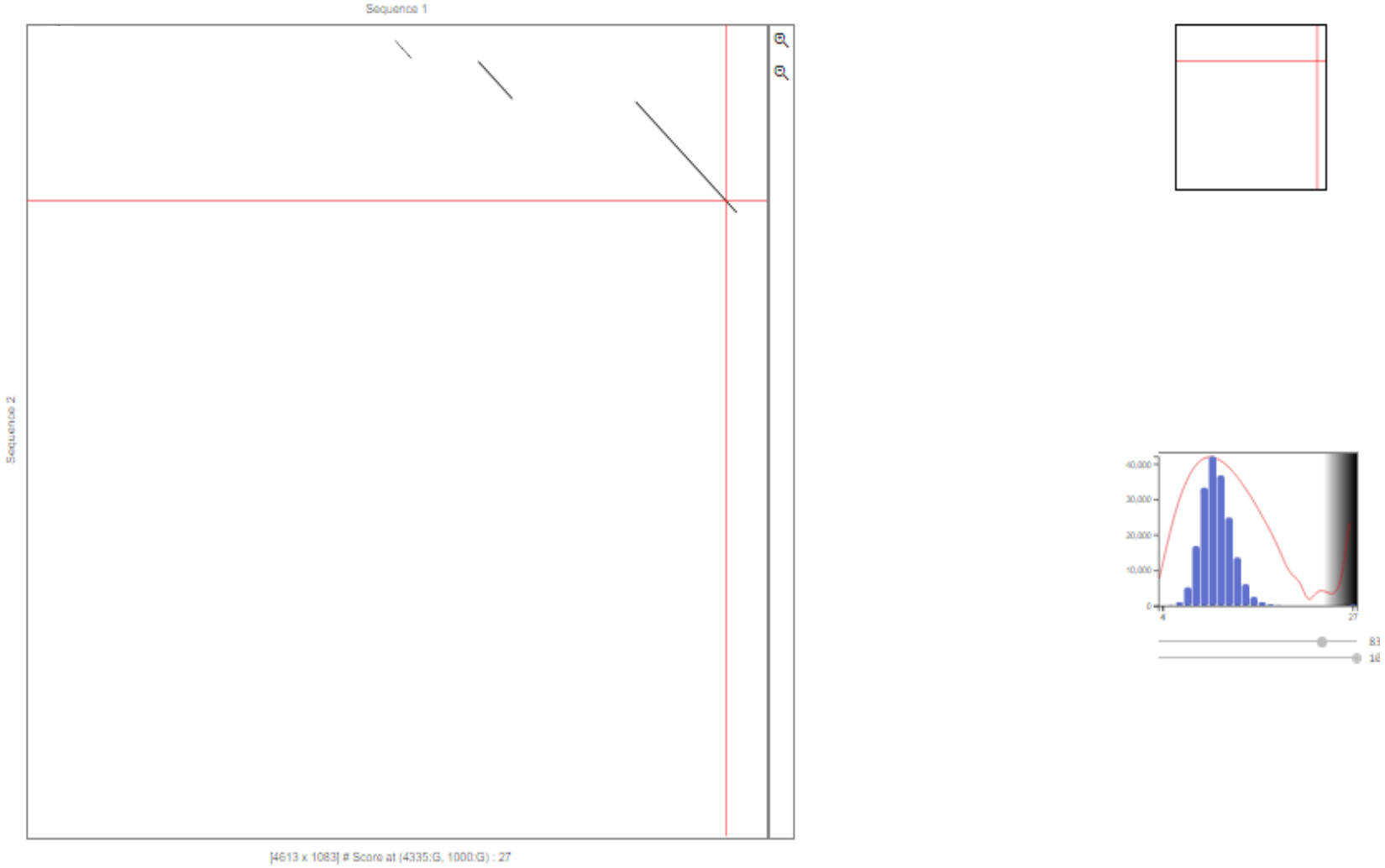
B) TaBRI1


Sequence 1
Sequence 2
[3890 x 3363] # Score at (2734:C, 2584:C) : 15
Seq1 : 2734
GAAATGGAGACCATTGGCAAGATCAAACACCGCAACCTTGTTCCGCTCCTCGGCTACTGCAAGATTGGTGAGG
GAAATGGAGACCATTGGCAAGATCAAACACCGCAACCTTGTTCCGCTCCTCGGCTACTGCAAGATTGGTGAGG
Seq2 : 2584

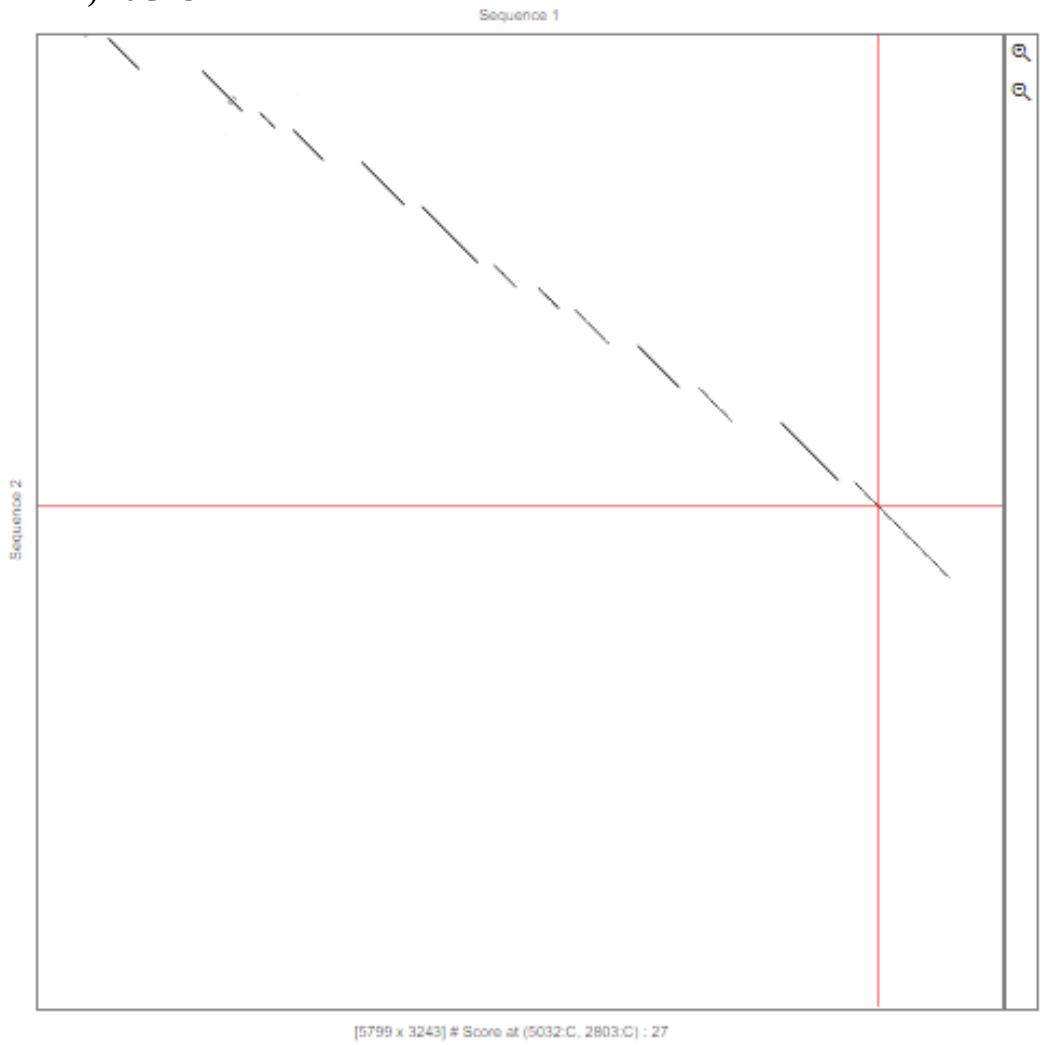
C) TaCOMT1

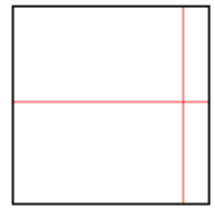

Sequence 1
Sequence 2
[3161 x 1107] # Score at (2301:A, 493:A) : 15
60,000
40,000
20,000
0
15
82%
95%
Seq1:2301
GACGGCGGCATCCCGTTCAACAAGGCGTACGGGATGACGGCGTTCGAGTACCACGGCACGGACGCCCGCTTCA
GACGGCGGCATCCCGTTCAACAAGGCGTACGGGATGACGGCGTTCGAGTACCACGGCACGGACGCCCGCTTCA
Seq2:493

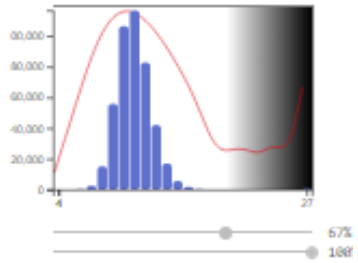
D) TaCAD8C

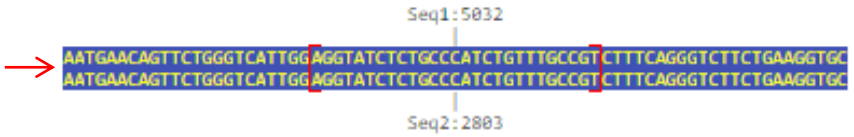

Sequence 1
Sequence 2
[4613 x 1083] # Score at (4335:G, 1000:G) : 27
40,000
30,000
20,000
10,000
0
4
27
Seq1:4335
CAGATCGAGGTCGTCAAGATGGACTACGTCAACCAGGCGTTCGAGAGGCTCGAGCGCAACGACGTGCGCTACC
CAGATCGAGGTCGTCAAGATGGACTACGTCAACCAGGCGTTCGAGAGGCTCGAGCGCAACGACGTGCGCTACC
Seq2:1000

E) TaCESA1


Sequence 1
Sequence 2
[5799 x 3243] # Score at (5032:C, 2803:C) : 27






Seq1:5032
AATGAACAGTTCTGGGTCATTGGAGGTATCTCTGCCCATCTGTTTGCCGTCTTTCAGGGTCTTCTGAAGGTGC
AATGAACAGTTCTGGGTCATTGGAGGTATCTCTGCCCATCTGTTTGCCGTCTTTCAGGGTCTTCTGAAGGTGC
Seq2:2803

F) TaCESA4

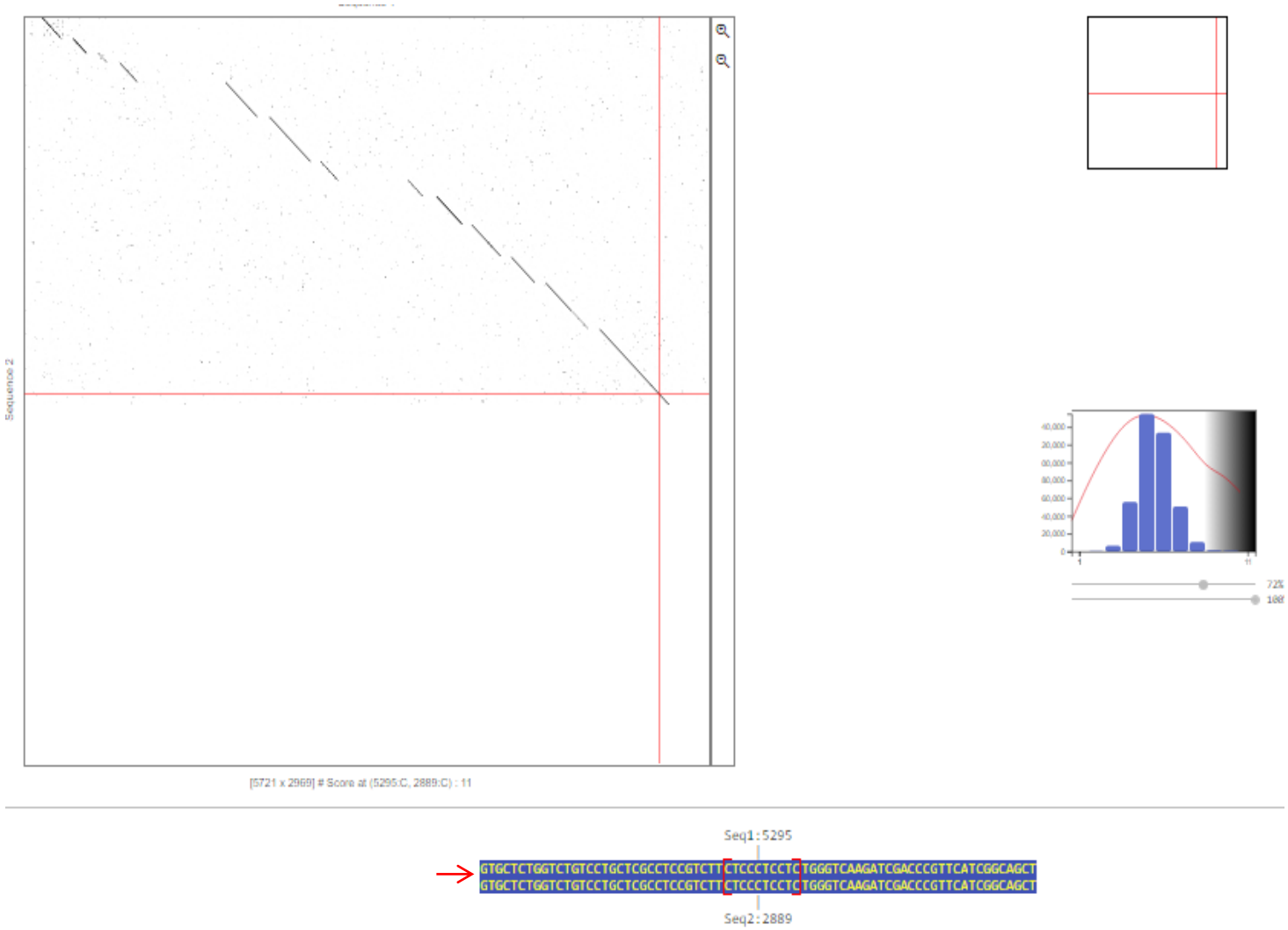


Figure 4.1. The precise location of the *in silico* probe of the six lodging-resistant genes within the reference wheat exons using Dotlet JS beta (Junier and Pagni, 2000). The x-axis represents the gene sequences, while the y-axis represents the CDS. The red arrows indicate the *in silico* probes starting site. The cross-hairs in the plot spaces highlight the region surrounding the *in silico* probes.

Table 4.2 The precise location of the *in silico* probes within wheat reference exons and the length of gene/CDS regions spanning both 5' and 3' of the *in silico* probes.

Genes	Length of the genes/CDS (in bps) 5' to the *in silico* probes		*In silico* probe's length (in bps)	Length of the genes/CDS (in bps) 3' to *in silico* probe		Wheat reference exon to which the *in silico* probe maps*
	Gene	CDS		Gene	CDS	
Rht1	1762	1571	53	730	242	1
BRI1	2705	2555	32	1153	776	1
COMT1	2272	464	36	853	607	2
CAD8C	3012	1195	38	263	68	3
CESA1	5003	2774	23	773	446	14
CESA4	5264	2858	27	436	43	13

*See Figure 4.1 for visual representation.

4.3 Chromosomal Location of Tef Homologous Candidate Lodging-Resistant Genes in *Eragrostis tef*

The six potential tef homologous candidate genes associated with lodging resistance were found within the tef subgenomes using *in silico* probes presented in Table 4.3. The tef genomic sequence used for this purpose was that of VanBuren *et al.* (2020) reposited in the NCBI database. Based on the length of the wheat gene lengths 5' and 3'of the *in silico* probes (see Table 4.3), approximately similar lengths of tef were determined 5' and 3' of the location that the *in silico* probes mapped to the tef subgenomes (Table 4.3).

Table 4.3 Mapping *in silico* probes of lodging-resistant genes in tef subgenomes.

In silico probes (Strand orientation*)	Tef subgenomes (with accession no.)	Approximate length of gene in bps	
		5' to the *in silico* probes	3' to the *in silico* probes
Rht1 (F)	4A (CM044760.1)	2100	730
BRI1 (R)	3B (CM044759.1)	3120	1500
COMT1 (R)	8B (CM044769.1)	2520	960
CAD8C (F)	2B (CM044757.1)	3060	300
CESA1 (R)	5A (CM044762.1)	5040	780
CESA4 (F)	3A (CM044758.1)	5280	480

Based on sequence of VanBuren *et al*. (2020).
* F represents 'forward' and R represents 'reverse complement'.

4.4 Computational Prediction of the Exon-Intron Structure of the Six Candidate Lodging-Resistant Genes in Tef

The retrieved tef six candidate lodging-resistant genes' sequences were computationally analyzed for various attributes, including the exon-intron structure, length, and number of exons, GC composition, and phylogenetic relationship between the candidate genes and their orthologues in wheat, barley and rice. These findings are presented in Figure 4.2 and Table 4.4.

As the tef orthologous lodging-resistant genes were obtained using *in silico* probes derived from highly conserved regions among wheat, barley and rice, it was important to confirm that the obtained tef sequences were indeed homologues. Hence, the computationally predicted

exon-intron structure for the tef lodging-resistant genes generated using SoftBerry was compared to the exon-intro structure of the corresponding genes in wheat, barley and rice. The comparison reveals that the EtRht1, EtBRI1, EtCOMT1, EtCAD8C, and EtCESA4 genes exhibit exon and intron numbers and structures coincidental to that of their counterparts in wheat, barley and rice (Figure 4.2 versus Table 3.1). The one case where a difference is observed is for EtCESA1, which is predicted to possess 11 exons, as compared to the 14 exons in wheat, barley and rice (Figure 4.2), indicating an inconsistency in the exon-intron structure compared to that of wheat, barley and rice genes as depicted in Figure 4.2 (tef) versus Table 4.1 (wheat). Figure 4.3 illustrates the computationally predicted exon-intron structure of tef CESA1's to scale comparison to its wheat counterpart (Figure 4.3).

Observing that tef orthologues follow similar gene structure as those of the three cereal crops, henceforth, the predicted tef genes are assigned the prefix Et, short form of *Eragrostis tef:* thus, EtRht1, EtBRI1, EtCOMT1, EtCAD8C, EtCESA1, and EtCESA4.

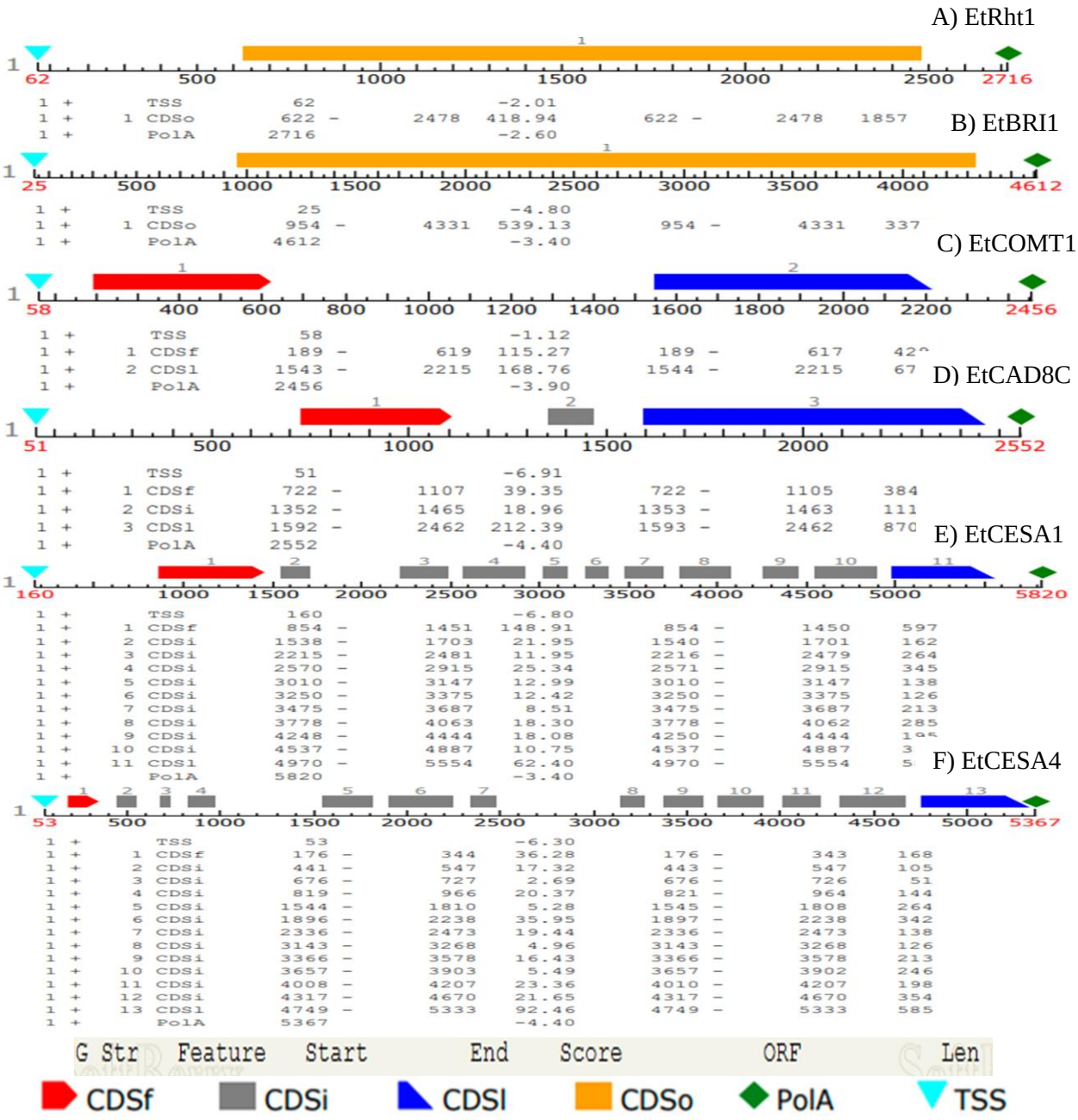


Figure 4.2. Computational prediction of the exon-intron structure of candidate homologous lodging-resistant genes in tef. A) EtRht1 gene with a sequence length of 2655 bps, with single exon in the +chain. B) EtBRI gene spanning 4588 bps, with single predicted exon in the +chain. C) EtCOMT1 gene with a sequence length of 2399 bps, in the +chain, with two exons. D) EtCAD8C gene covering 2502bps with three exons in the +chain. E) EtCESA1 gene spanning 5661 bps, with eleven exons in the +chain. F) EtCESA4 gene with a sequence length of 5661 bps, one predicted gene, with thirteen exons in the +chain. The paces between the exon represent intron. [These gene structures are predicted using FGENESH-HMM (Solovyev *et al*., 2006) gene finder].

CDSf: coding sequence in first exon,
CDSi: coding sequence in internal exon,
CDSl: coding sequence in last exon, and
CDSo: one coding sequence
G: gene
ORF: open reading frame
PolA: polyA
Str: DNA strand orientation
TSS: transcription starting site

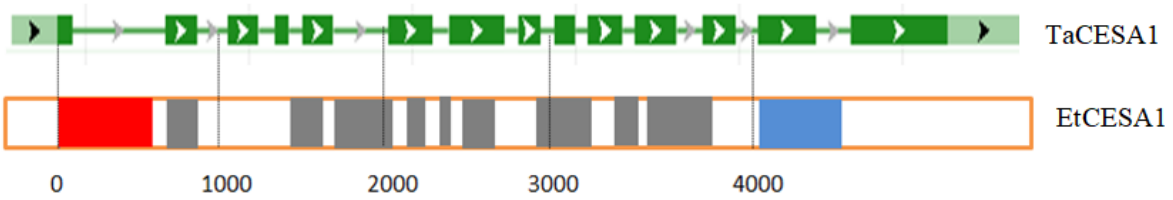


Figure 4.3. A 'to scale' comparison between the computationally predicted exon-intron structure of tef (top) and wheat (bottom) CESA1 gene. [The dotted lines represent bps starting from the translation start site. Green boxes represent exons in wheat; while grey boxes are internal exons, and red and blue boxes are first and last exons, respectively, in tef].

The comparison of six candidate lodging-resistant genes across tef, wheat, barley, and rice revealed key attributes such as gene and coding sequence (CDS) lengths, exon counts, and GC composition, summarized in Table 4.4.

EtRht1 and EtBRI1 were identified as single-exon genes, consistent with their counterparts in wheat, barley, and rice. The CDS length of EtRht1 was found to be identical to barley (1857 bps), slightly shorter than wheat (1866 bps), and significantly shorter than rice (1878 bps). Similarly, EtBRI1's CDS length (3378 bps) was slightly longer than wheat (3363 bps), rice (3357 bps), and barley (3366 bps). Despite minor variations, the GC% content of these genes was comparable across the species (Table 4.4).

EtCOMT1 exhibited a two-exon structure, consistent with wheat, barley, and rice. Its CDS length (1101 bps) was slightly longer than wheat and barley (1083 bps) and shorter than rice (1107 bps), resulting in a polypeptide length 6 amino acids longer than the shortest CDS observed in wheat and barley. The GC% content of EtCOMT1 showed slight differences compared to the other species (Table 4.4).

EtCAD8C had a CDS length of 1368 bps, longer than wheat and barley (1263 bps) and rice (1317 bps), resulting in a polypeptide 17 to 35 amino acids longer than those in the other cereals. The GC% content of EtCAD8C was comparable to wheat, barley, and rice (Table 4.4).

EtCESA4's CDS length (2904 bps) was slightly shorter than wheat (2925 bps), barley (2946 bps), and rice (2970 bps), resulting in a polypeptide 7 to 22 amino acids shorter than its counterparts in the other cereals. Minor differences in GC% content were also observed (Table 4.4).

EtCESA1, in contrast, exhibited a distinct exon-intron structure with 11 exons compared to 14 exons in wheat, barley, and rice. Despite this structural difference, EtCESA1's CDS length

(3261 bps) was comparable to wheat (3231 bps), barley (3237 bps), and rice (3231 bps). The predicted protein length of EtCESA1 was 8 to 10 amino acids longer than those of the other species, with comparable GC% content (Table 4.4).

Overall, these findings confirm the homology of the predicted tef genes with their orthologues in wheat, barley, and rice, highlighting minor variations in gene structure and sequence characteristics across these cereal species.

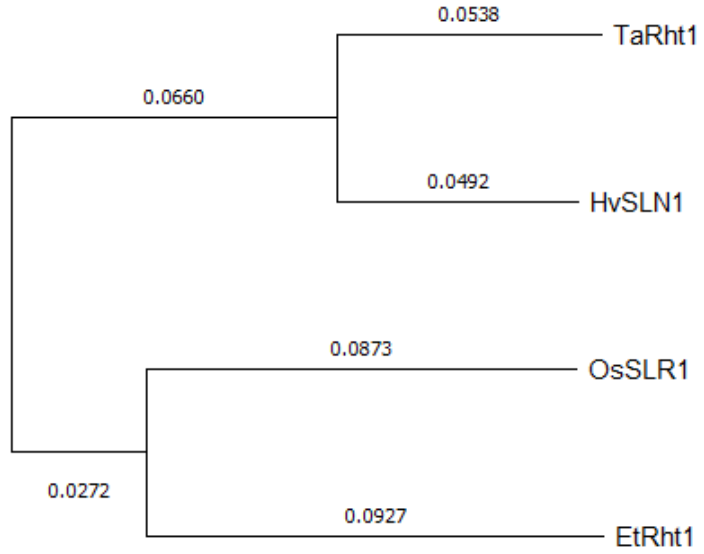
Table 4.4 Comparative presentation of the percentage of GC content, number of exons for each gene in wheat, barley, rice, and predicted tef, as well as the length of the CDS.

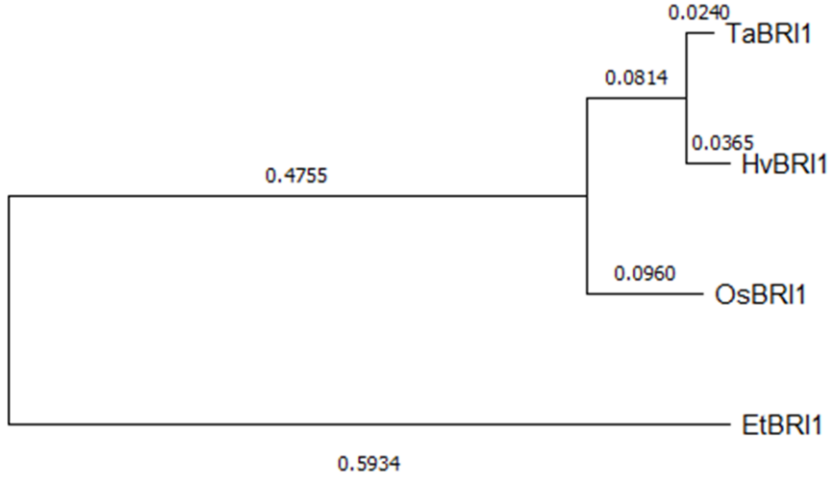
Wheat				**Barley**			**Rice**			**Tef***			
Name of the Genes	GC** %	No of exon	CDS length	GC** %	No of exon	CDS length	GC** %	No of exon	CDS length	GC** %	No of exon	CDS length	Gene length
Rht1	67.2	1	1866	65.9	1	1857	65.2	1	1878	65.8	1	1857	2655
BRI1	58.1	1	3363	57.9	1	3357	56.2	1	3366	56.7	1	3378	4588
COMT1	60.4	2	1083	62.0	2	1083	61.7	2	1107	54.9	2	1101	2399
CAD8C	52.9	3	1263	62.6	3	1263	63.6	3	1317	56	3	1368	2502
CESA1	49.0	14	3231	48.9	14	3237	56.1	14	3231	47.9	11	3261	5661
CESA4	63.7	13	2925	56.4	13	2946	52.2	13	2970	46.4	13	2904	5315

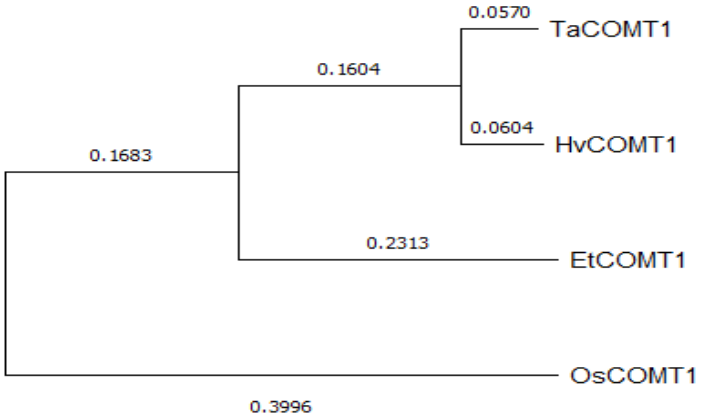
*Computationally predicted.
GC%: represents the percentage of G and C residues.

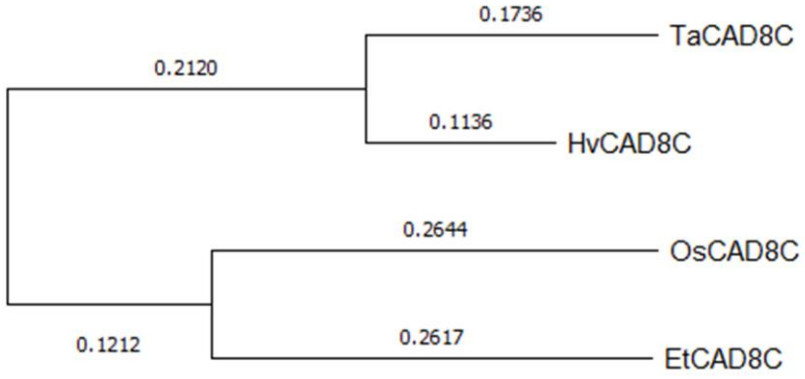
4.5 Phylogenetic Analysis of the Candidate Lodging-Resistant Genes in Tef and Their Wheat, Barley and Rice Homologous

The phylogenetic analysis in Figure 4.4 illustrates the relationships among six lodging-resistant genes in wheat, barley, rice, and their predicted homologs in tef. Among the tef genes, EtBRI1, EtCESA1 and EtCESA4 appear distantly related to the cluster formed by wheat, barley, and rice. In contrast, EtRht1, EtCOMT1 and EtCAD8C show closer to rice; and wheat and barley being more distantly related. This suggests that anthropogenic selection in wheat and barley may have accelerated their evolutionary pace compared to tef, which shows shorter estimated evolutionary distances for its homologous genes from the root.


0.0538
TaRht1
0.0660
0.0492
HvSLN1
0.0873
OsSLR1
0.0272
0.0927
EtRht1


0.0240
TaBRI1
0.0814
0.0365
HvBRI1
0.4755
0.0960
OsBRI1
EtBRI1
0.5934


0.0570
TaCOMT1
0.1604
0.0604
HvCOMT1
0.1683
0.2313
EtCOMT1
OsCOMT1
0.3996


0.1736
TaCAD8C
0.2120
0.1136
HvCAD8C
0.2644
OsCAD8C
0.1212
0.2617
EtCAD8C

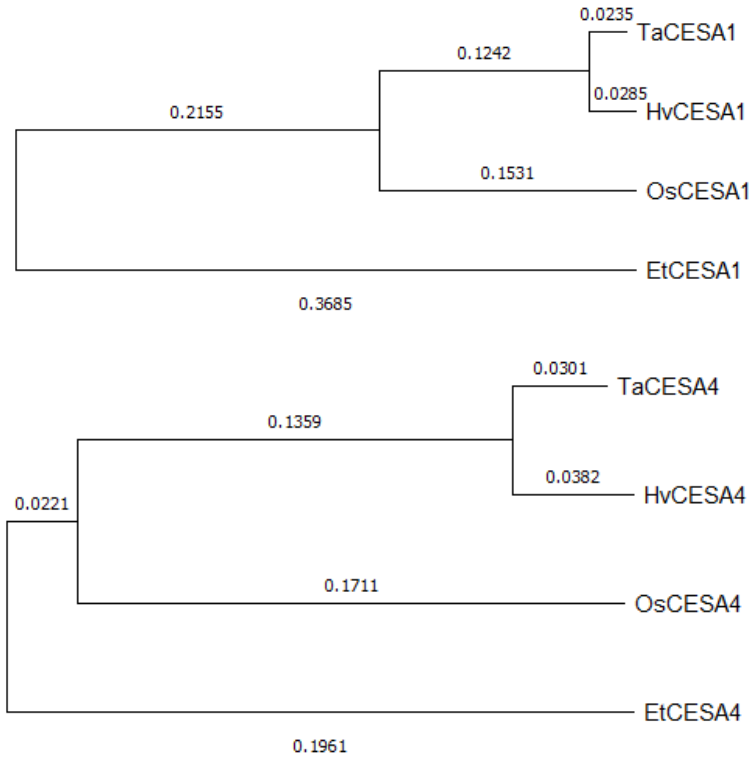


Figure 4.4. The phylogenetic relationship among six candidate tef lodging-resistant genes with their homologous in wheat, barley and rice by ML method.
Ta: wheat (*T. aestivum):* Hv: barley (*H. vulgare*); Os: rice ***(****O. sativa****)***; and, Et: tef (*E. tef).*
[*Branch lengths are shown above branches*].

4.6 Sequence Identity Comparison of Predicted Tef Proteins with Wheat, Barley and Rice

The predicted tef proteins EtRht1, EtBRI1, EtCOMT1, EtCAD8C, EtCESA1, and EtCESA4 exhibit high sequence identity with those of wheat, barley and rice (grey shaded in Table 4.5).

The percent identity matrix, generated by Clustal 2.1 using the BLOSUM62 amino acid substitution matrix, illustrates share hign identity percentages between the predicted six tef lodging-resistant polypeptide sequences and those found in wheat, barley and rice. The predicted EtRht1 shares a sequence identity of about 90% with the other three species. Moreover, the aa sequence identity of predicted EtBRI1 polypeptide is approximately 84% compared to wheat, barley and rice. Additionally, the aa sequence identity of predicted EtCAD8C polypeptide is approximately 74% compared to wheat, barley and rice. The amino acid sequence identity of predicted EtCOMT1 is about 83% against all three species; and, EtCESA4's identity is about 66% with wheat and barley, and 93% with rice. EtCESA1 with its predicted fewer number of exons exhibits an overall average sequence identity of about 68% with its three orthologous (see Table 4.5).

Table 4.5 Percent identity matrix among wheat, barley, rice and putative tef polypeptide sequences.

Rht1		**Wheat**	**Barley**	**Rice**	**Tef**
	Wheat	100	96.91	87.46	89.49
	Barley	96.91	100	87.75	89.84
	Rice	87.46	87.75	100	89.03
	Tef	89.49	89.84	89.03	
BRI1		**Wheat**	**Barley**	**Rice**	
	Wheat	100	96.91	83.50	83.29
	Barley	96.06	100	83.02	82.81
	Rice	83.50	83.02	100	83.50
	Tef	83.29	82.81	83.50	
COMT1		**Wheat**	**Barley**	**Rice**	
	Wheat	100	98.03	82.22	82.22
	Barley	98.03	100	82.78	82.78
	Rice	83.06	82.76	100	82.51
	Tef	82.22	82.51	82.51	
CAD8C		**Wheat**	**Barley**	**Rice**	
	Wheat	100	96.11	88.30	74.64
	Barley	96.11	100	88.86	73.68
	Rice	88.30	88.86	100	75.54
	Tef	74.64	73.86	75.54	
CESA1		**Wheat**	**Barley**	**Rice**	
	Wheat	100	99.17	70.54	67.68
	Barley	99.17	100	70.35	67.72
	Rice	70.54	70.35	100	67.68
	Tef	67.68	67.72	67.68	
CESA4		**Wheat**	**Barley**	**Rice**	
	Wheat	100	99.33	65.41	65.71
	Barley	99.33	100	65.44	65.43
	Rice	65.41	65.44	100	92.66
	Tef	65.71	65.43	92.66	

4.7 Unique Variations in *Eragrostis tef* Predicted Polypeptide Sequences as Compared to Their Corresponding Sequences in Wheat, Barley and Rice

The putative polypeptide sequences of six tef lodging-resistant genes were obtained from their predicted CDS (see Figure 4.2). The output from the MSA was examined to identify variations that are uniquely present in tef, but otherwise conserved in the other three cereal crops (Figure 4.5). Identified variations in known conserved functional and structural domains were collated and catalogued (Table 4.6).

The comparative analysis yielded unique variations in the six putative proteins in tef, as compared to their counterparts in wheat, barley and rice. Yielding valuable insights into the distinct differences observed in tef among these proteins.

The putative amino acid sequence of EtRht1 exhibits notable variations in the polar residues site and GRAS domain particularly, in the Leucine Heptad Repeat 1 (LHR1), short alpha-helical region and winged-helix (SAW) and Proline-Phenylalanine-Tyrosine-Arginine-Glutamate (PFYRE) subdomains, when compared to wheat, barley and rice (Figure 4.5A and Table 4.6).

The EtBRI1 protein exhibited unique differences in multiple regions when compared to its equivalents in wheat, barley and rice. Variations were observed in the signal peptide, cysteine pair 1 region (Cys 1), the leucine-rich repeat domain (LRR), transmembrane domain (TMD), and the kinase domain (KD), as illustrated in Figure 4.5B and elaborated upon in Table 4.6.

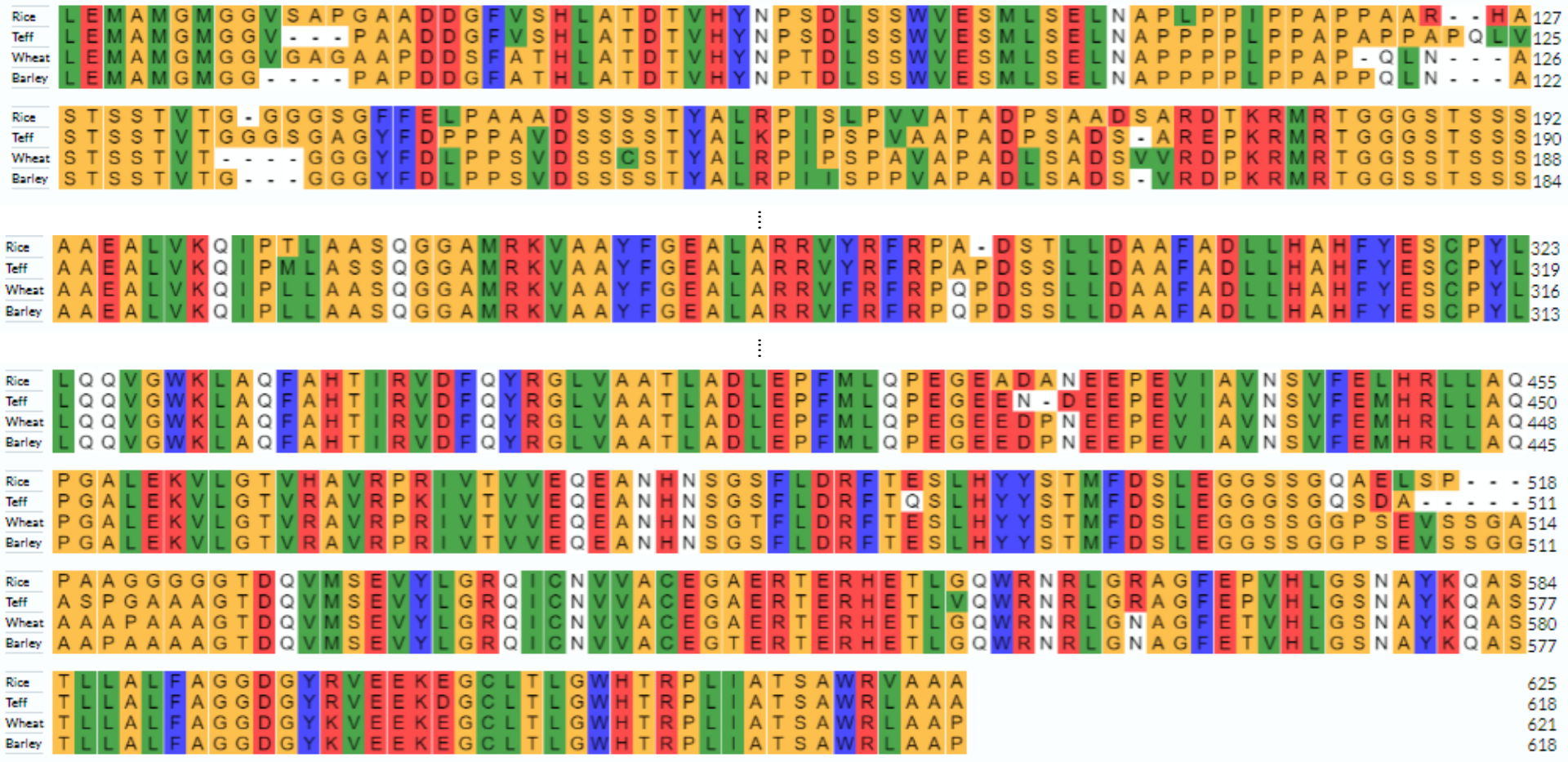
The EtCOMT1 protein has unique variations identified within the unique N-terminal dimerization domain, and the C-terminal methyltransferase domain. These differences distinguished it from the alignments observed in wheat, barley and rice, as represented Figure 4.5C and Table 4.6.

The EtCAD8C protein exhibited distinct amino acid disparity within the nicotinamide adenine dinucleotide phosphate (NADP+) region in comparison to its homologous in wheat, barley and rice, as outlined in Figure 4.5D and Table 4.6.

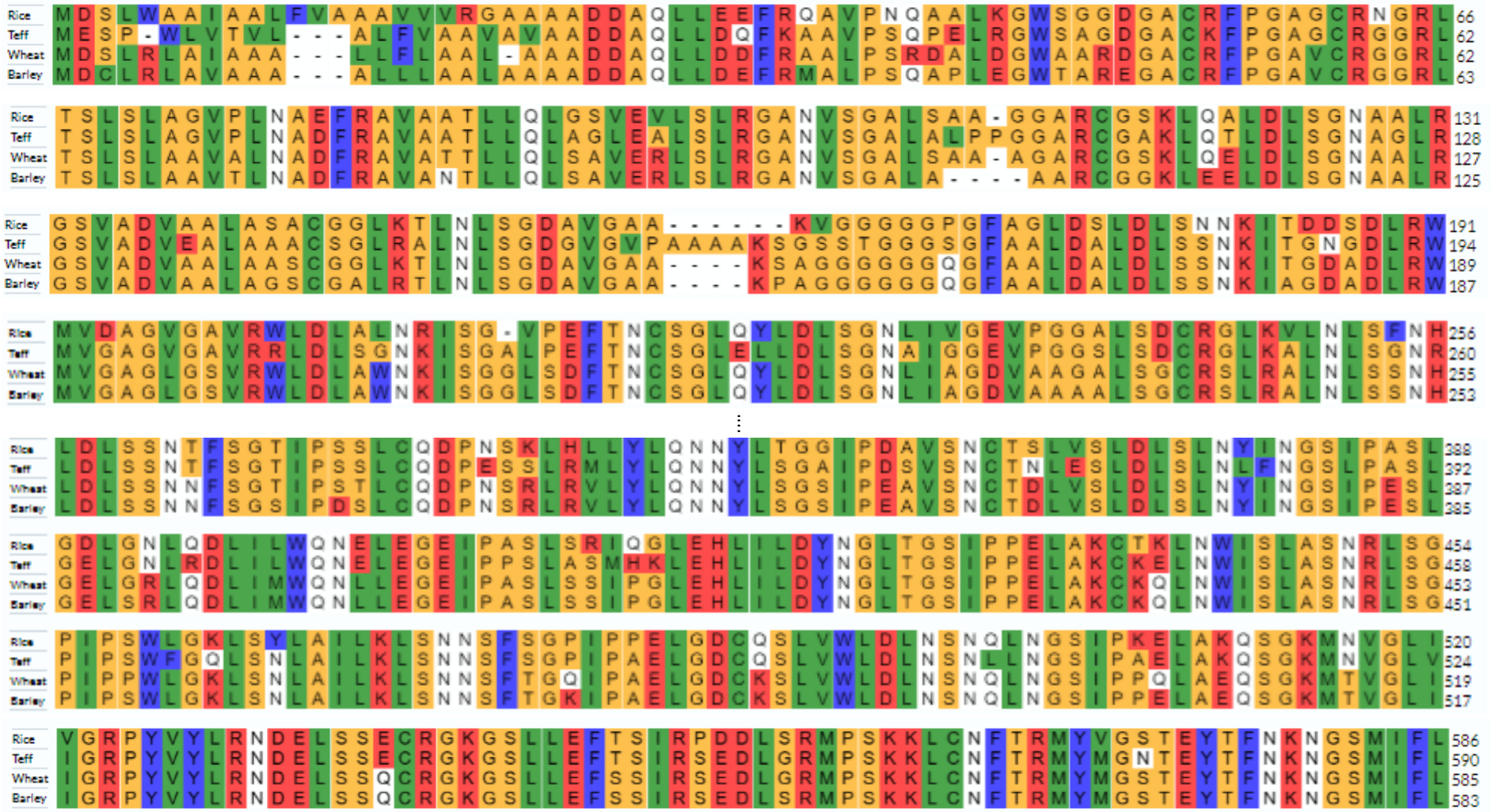
Moreover, variations were observed in the EtCESA1 protein particularly, in the zinc finger domain, coil-coiled region, modified phosphoserine residues, the transmembrane domains (TMDs), cytoplasmic domains, and N-linked glycosylation site, when compared to its homologous in wheat, barley and rice, as illustrated in Figure 4.5E and Table 4.6.

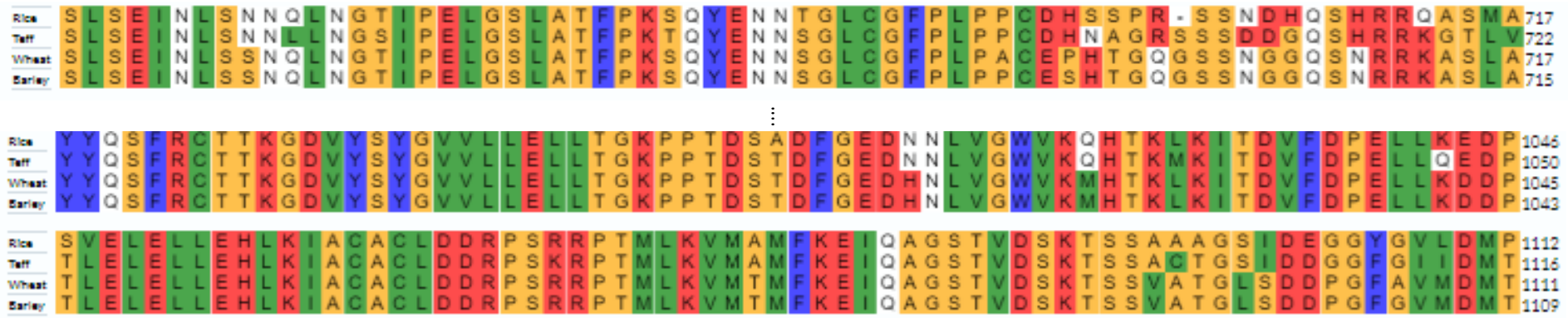
Furthermore, the EtCESA4 protein displayed unique amino acid variations in the TMD and extracellular domain distinguishing it from other three cereal crops (Figure 4.5F and Table 4.6).

A) Rht1

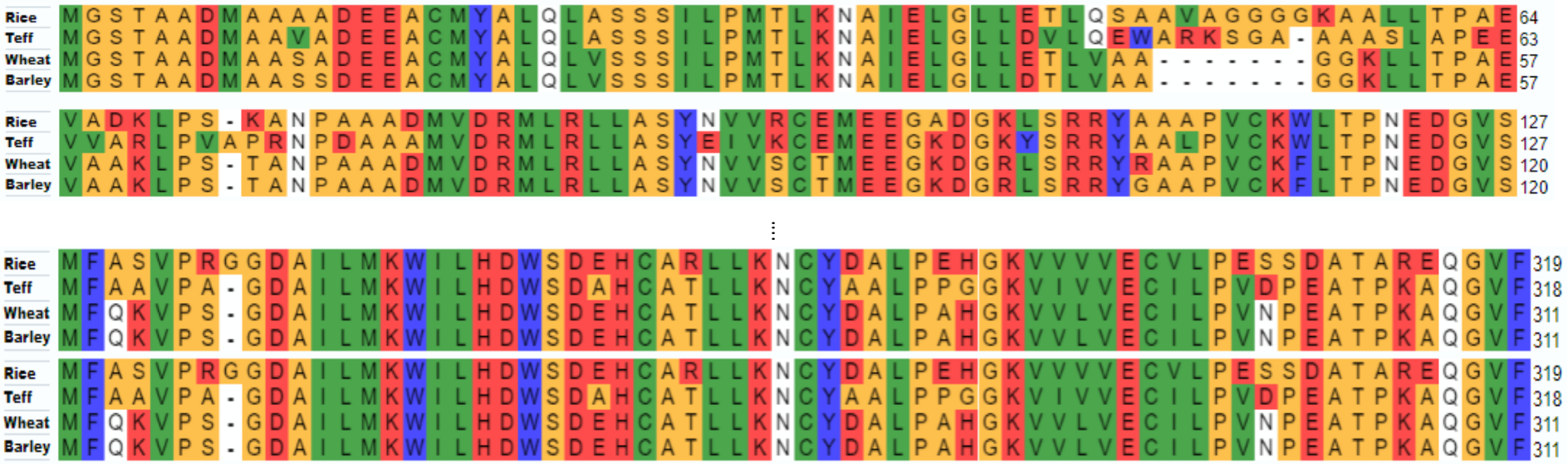

Rice LEMAMGMGGVSAPGAADDGFVSHLATDTVHYNPSDLSSWVESMLSELNAPLPPIPPAPPAARHA 127
Teff LEMAMGMGGV---PAADDGFVSHLATDTVHYNPSDLSSWVESMLSELNAPPPPLPPAPAPPAPQLV 125
Wheat LEMAMGMGGVGAGAAPDDSFATHLATDTVHYNPTDLSSWVESMLSELNAPPPPLPPAP-QLN---A 126
Barley LEMAMGMGG----PAPDDGFATHLATDTVHYNPTDLSSWVESMLSELNAPPPPLPPAPPQLN---A 122
Rice STSSTVTG-GGGSGFFELPAAADSSSSTYALRPISLPVVATADPSAADSARDTKRMRTGGGSTSSS 192
Teff STSSTVTGGGSGAGYFDPPPAVDSSSSTYALKPIPSPVAAPADPSADS-ARERPKRMRTGGGSTSSS 190
Wheat STSSTVT----GGGYFDLPPSVDSSCSTYALRPIPSPAVAPADLSADSVVRDPKRMRTGGSSTSSS 188
Barley STSSTVTG---GGGYFDLPPSVDSSSSTYALRPITSPPVAPADLSADS-VRDPKRMRTGGSSTSSS 184
Rice AAEALVKQIPTLAASQGGAMRKVAAYFGEALARRVYRFRPA-DSTLLDAAFADLLHAHFYESCPYL 323
Teff AAEALVKQIPMLASSQGGAMRKVAAYFGEALARRVYRFRPAPDSSLLDAAFADLLHAHFYESCPYL 319
Wheat AAEALVKQIPLLAASQGGAMRKVAAYFGEALARRVFRFRPQPDSSLLDAAFADLLHAHFYESCPYL 316
Barley AAEALVKQIPLLAASQGGAMRKVAAYFGEALARRVFRFRPQPDSSLLDAAFADLLHAHFYESCPYL 313
Rice LQQVGWKLAQFAHTIRVDFQYRGLVAATLADLEPFMLQPEGEADANEEPEVIAVNSVFELHRLLAQ 455
Teff LQQVGWKLAQFAHTIRVDFQYRGLVAATLADLEPFMLQPEGEEN-DEEPEVIAVNSVFEMHRLLAQ 450
Wheat LQQVGWKLAQFAHTIRVDFQYRGLVAATLADLEPFMLQPEGEEDPNEEPEVIAVNSVFEMHRLLAQ 448
Barley LQQVGWKLAQFAHTIRVDFQYRGLVAATLADLEPFMLQPEGEEDPNEEPEVIAVNSVFEMHRLLAQ 445
Rice PGALEKVLGTVHAVRPPRIVTVVEQEANHNSGSFLDRFTESLHYYSTMFDSLEGGSSGQAELSP--- 518
Teff PGALEKVLGTVRAVRPPKIVTVVEQEANHNSGSFLDRFTQSLHYYSTMFDSLEGGGSGQSDA----- 511
Wheat PGALEKVLGTVRAVRPPRIVTVVEQEANHNSGTFLDRFTESLHYYSTMFDSLEGGSSGGPSEVSSGA 514
Barley PGALEKVLGTVRAVRPPRIVTVVEQEANHNSGSFLDRFTESLHYYSTMFDSLEGGSSGGPSEVSSGG 511
Rice PAAGGGGGTDQVMSEVYLGRQICNVVACEGAERTERHETLGQWRNRLGRAGFEPVHLGSNAYKQAS 584
Teff ASPGAAAAGTDQVMSEVYLGRQICNVVACEGAERTERHETLVQWRNRLGRAGFEPVHLGSNAYKQAS 577
Wheat AAAPAAAAGTDQVMSEVYLGRQICNVVACEGAERTERHETLGQWRNRLGNAGFETVHLGSNAYKQAS 580
Barley AAPAAAAGTDQVMSEVYLGRQICNVVACEGTERTERHETLGQWRNRLGNAGFETVHLGSNAYKQAS 577
Rice TLLALFAGGDGYRVEEKEGCLTLGWHTRPLIATSAWRVAAA 625
Teff TLLALFAGGDGYRVEEKDGCLTLGWHTRPLIATSAWRLAAA 618
Wheat TLLALFAGGDGYKVEEKEGCLTLGWHTRPLIATSAWRLAAP 621
Barley TLLALFAGGDGYKVEEKEGCLTLGWHTRPLIATSAWRLAAP 618

B) BRI1


Rice MDSLWAAIAALFVAAAVVVRGAAAADDAQLLEEFRQAVPNQAALKGWSGGDGACRFPGAGCRNGRL 66
Teff MESP-WLVTVL---ALFVAAVAVAADDAQLLDQFKAAVPSQPELRGWSAGDGACKFPGAGCRGGRL 62
Wheat MDSLRLAIAAA---LLFLAALA-AAADDAQLLDDFRAALPSRDALDGWAARDGACRFPGAVCRGGRL 62
Barley MDCLRLAVAAA---ALLLAALAAAADDAQLLDEFRMALPSQAPLEGWTARDGACRFPGAVCRGGRL 63
Rice TSLSLAGVPLNAEFRAVAATLLQLGSVEVLSLRGANVSGALSAA-GGARCGSKLQALDLSGNAALR 131
Teff TSLSLAGVPLNADFRAVAATLLQLAGLEALSLRGANVSGALALPPGGARCGAKLQTLDLSGNAGLR 128
Wheat TSLSLAAVALNADFRAVATTLLQLSAVERLSLRGANVSGALSAA-AGARCGSKLQELDLSGNAALR 127
Barley TSLSLAAVTLNADFRAVANTLLQLSAVERLSLRGANVSGALA----AARCGGKLEELDLSGNAALR 125
Rice GSVADVAALASACGGLKTLNLSGDAVGAA------KVGGGGGPGFAGLDSLDLSNNKITDDSDLRW 191
Teff GSVADVEALAAACSGLRALNLSGDGVGVPAAAAKSGGSSTGGGSGFAALDALDLSSNKITGNGDLRW 194
Wheat GSVADVAALAASCGGLKTLNLSGDAVGAA----KSAGGGGGGQGFAALDALDLSSNKITGDADLRW 189
Barley GSVADVAALAGSCGALRTLNLSGDAVGAA----KPAGGGGGGQGFAALDALDLSSNKIAGDADLRW 187
Rice MVDAGVGAVRWLDLALNRISG-VPEFTNCSGLQYLDLSGNLIVGEVPGGALSDCRGLKVLNLSFNH 256
Teff MVGAGVGAVRRLDLSGNKISGALPEFTNCSGLELLDLSGNAIGGEVPGGSLSDCRGLKALNLSGNR 260
Wheat MVGAGLGSVRWLDLAWNKISGGLSDFTNCSGLQYLDLSGNLIAGDVAAGALSGCRSLRALNLSSNH 255
Barley MVGAGLGSVRWLDLAWNKISGGLSDFTNCSGLQYLDLSGNLIAGDVAAAALSGCRSLRALNLSSNH 253
Rice LDLSSNTFSGTIPSSLCQDPNSKLHLLYLQNNYLTGGIPDAVSNCTSLVSLDLSLNYINGSIPASL 388
Teff LDLSSNTFSGTIPSSLCQDPESSLRMLYLQNNYLSGAIPDSVSNCTNLESLDLSLNLFNGSLPASL 392
Wheat LDLSSNNFSGTIPSTLCQDPNSRLRVLYLQNNYLSGSIPEAVSNCTDLVSLDLSLNYINGSIPESL 387
Barley LDLSSNNFSGSIPDSLCQDPNSRLRVLYLQNNYLSGSIPEAVSNCTDLVSLDLSLNYINGSIPESL 385
Rice GDLGNLQDLILWQNELEGEIPASLSRIQGLEHLILDYNGLTGSIPPELAKCTKLNWISLASNRLSG 454
Teff GELGNLRDLILWQNELEGEIPPSLASMHKLEHLILDYNGLTGSIPPELAKCKELNWISLASNRLSG 458
Wheat GELGRLQDLIMWQNLLEGEIPASLSSIPGLEHLILDYNGLTGSIPPELAKCKQLNWISLASNRLSG 453
Barley GELSRLQDLIMWQNLLEGEIPASLSSIPGLEHLILDYNGLTGSIPPELAKCKQLNWISLASNRLSG 451
Rice PIPSWLGKLSYLAILKLSNNSFSGPIPPELGDCQSLVWLDLNSNQLNGSIPKELAKQSGKMNVGLI 520
Teff PIPSWFGQLSNLAILKLSNNSFSGPIPAELGDCQSLVWLDLNSNLLNGSIPAELAKQSGKMNVGLV 524
Wheat PIPPWLGKLSNLAILKLSNNSFTGQIPAELGDCKSLVWLDLNSNQLNGSIPPQLAEQSGKMTVGLI 519
Barley PIPSWLGKLSNLAILKLSNNSFTGKIPAELGDCKSLVWLDLNSNQLNGSIPPELAEQSGKMTVGLI 517
Rice VGRPYVYLRNDELSSECRGKGSLLEFTSIRPDDLSRMPSKKLCNFTRMYVGSTEYTFNKNGSMIFL 586
Teff IGRPYVYLRNDELSSECRGKGSLLEFTSIRSEDLGRMPSKKLCNFTRMYMGNTEYTFNKNGSMIFL 590
Wheat IGRPYVYLRNDELSSQCRGKGSLLEFSSIRSEDLGRMPSKKLCNFTRMYMGSTEYTFNKNGSMIFL 585
Barley IGRPYVYLRNDELSSQCRGKGSLLEFSSIRSEDLSRMPSKKLCNFTRMYMGSTEYTFNKNGSMIFL 583

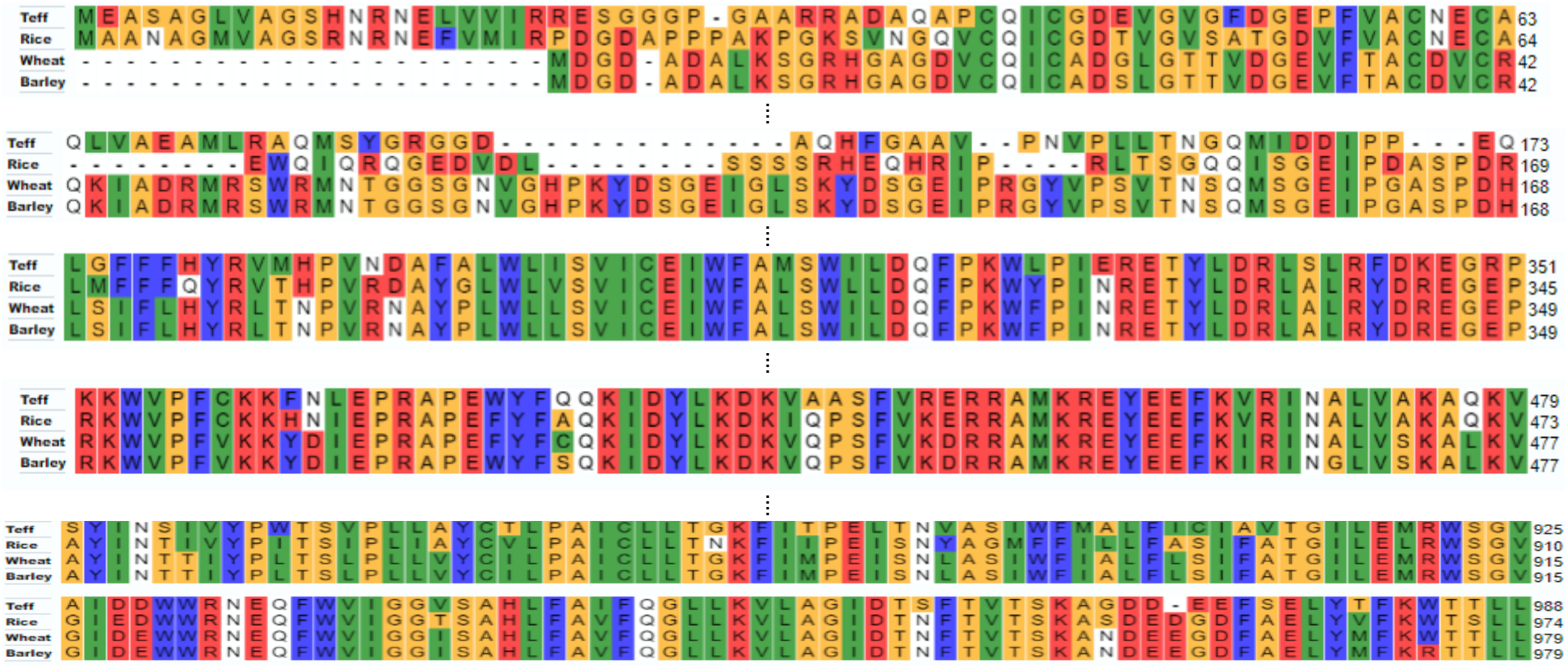


C) COMT1



D) CAD8C

E) CESA1



F) CESA4

Figure 4.5. Multiple sequence alignment shows the critical domains of the six lodging-resistant genes in the four species. A) Highlights amino acid variations and conservations among tef (EtRht1) and wheat (TaRht1), barley (SLN1) and rice (SLR1). B) Shows amino acid variations and conservations of BRI1 among tef, wheat, barley and rice. C) Highlights amino acid the variations and conservations of COMT1 among four cereal crops. D) Illustrates amino acid variations and conservations of CAD8C among the four species. E) Displays amino acid variations and conservation of CESA1 among the four species. F) Shows amino acid variation and conservations of CESA4 in tef, wheat, barley and rice. MSA was performed using Clustal Omega (Madeira *et al.*, 2019) enabling the identification of variations across these important genes in tef compared to the reference cereal crops.

⁞ Represents break in the MSA.

(-) Minus sign represents deletions/insertions within the amino acid sequence alignments.

Amino acids with the same colors possess the same physico-chemical property.

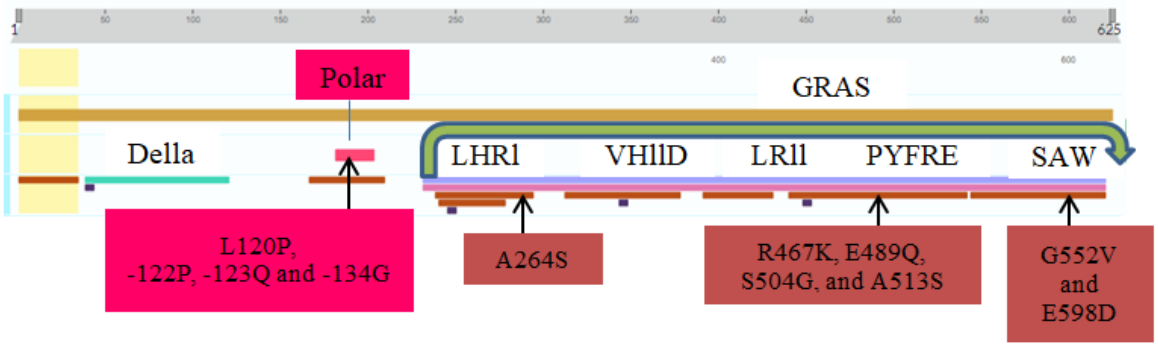
Table 4.6 Unique variations in putative tef polypeptide sequences compared to wheat, barley and rice proteins.

Proteins	Conserved amino acids in wheat, barley and rice; unique variations in the predicted tef at a given position in tef numbering.						
EtRht1	Polar residues		GRAS domain				
			LHR1 subdoman	PFYRE subdomain		SAW subdomain	
	L120P, -122P, -123Q and -134G		A267S	R467K, E489Q, S504G, and A513S		G552V, and E598D	
EtBRI1	Signal peptide	Cysteine pair 1	LRR domain		Island domian	Transmembrane domain	Kinase domain
	D2E, L4P, A6L, A8T, A9V and A19V	R51K	-107P, A135E, G142S, T146A, A153G, A156V, A157P, ----(158-161)AAAA, G166S,G167T, D189N, W205R, A209S, Q227E,Y228L, L235A, N347E, V375E, Y383L,L384F, Q399R, G421K, L464F, K466Q, Q553I, and Q667I		I520V and S576N	S719T, L724I, and A738F	K1047Q, R1073K
EtCOMT1	Dimerization domain			Methyltransferase domain			
	T44V, A48W, L58S, A62E, A65V, S70V, -71A, A73R, A76D and D79A			E278A, D289A and H294G			
EtCAD8C	NADP+ binding site C400S						
EtCESA1	Zinc/Ring finger domia	Coiled coil	Modified residues (phosphoserine)	Transmembrane domains	Cytoplasimic domains		N-linked Glycosylation
	P69A and K76R	K453R	S161T and S165 to 166-	T297M, R301N, Y304F and S898T	N333E, Y345F, R347K, E350R S911C, F913A and A914V		N963S
EtCESA4	Extracellular domains			Transmembrane domains			
	A863S			D900A			

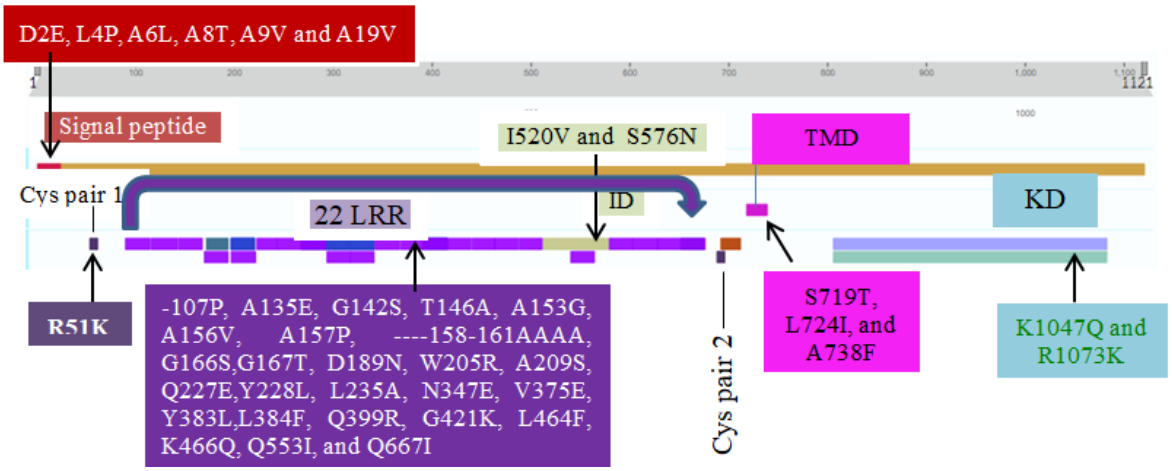
*Minus (-) sign indicates amino acid deletion.

The identified unique variations within the critical domains of the EtRht1, EtBRI1, EtCOMT1, EtCAD8C, EtCESA1, and EtCESA4 polypeptide sequences, compared to their corresponding amino acids, perfectly conserved across wheat, barley and rice, as depicted in Figure 4.6.

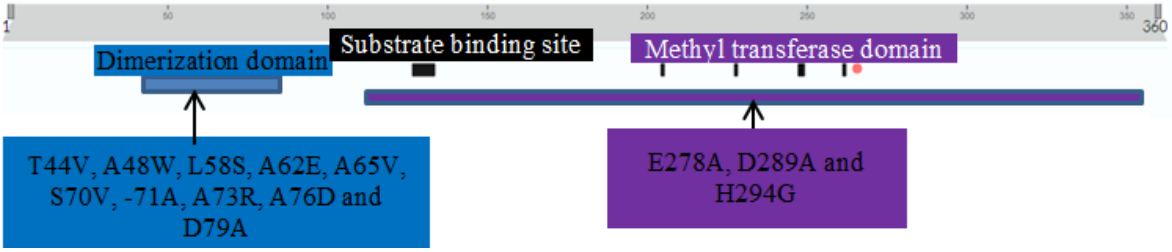
A) SLR1



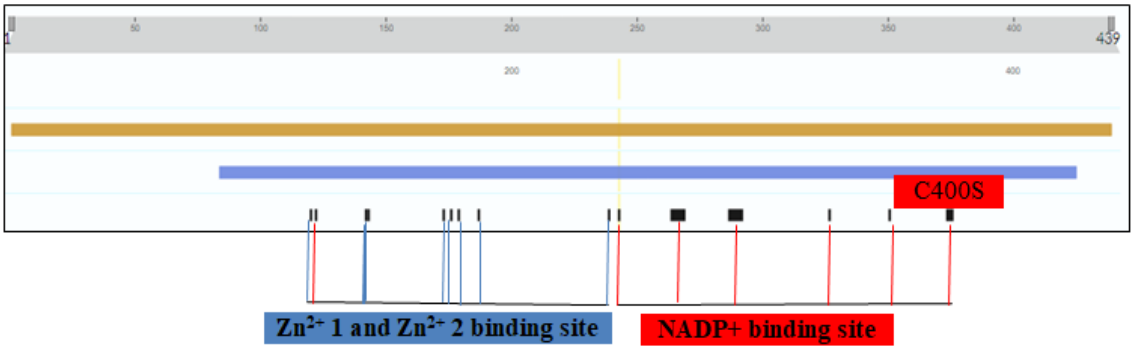
B) BRI1



C) COMT1



D) CAD8C



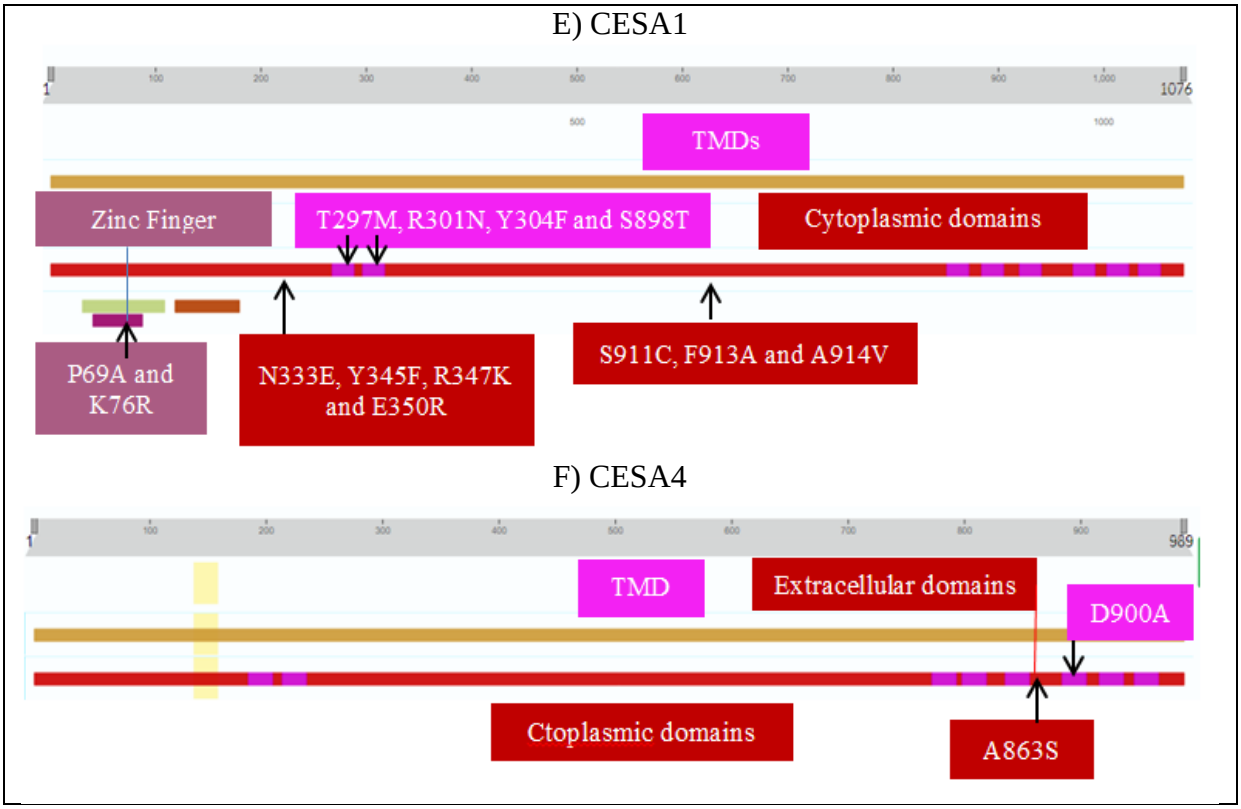


Figure 4.6. Amino acid substitutions and their locations within the domains, subdomains and at various binding sites of the six lodging-resistant polypeptide sequences.

All the identified unique variations in tef located in important domains were interrogated using a counterfactual approach in SIFT of which six counterfactual mutations, in four of the six proteins, were predicted to cause damage with SIFT score that is significant or higher. The predicted EtRht1, EtBRI1, EtCOMT1 and EtCESA1 proteins' domains, significant counterfactual mutations, SIFT score effects, and the corresponding sequence representations (Seq rep) - obtained from the UniProt database - are presented in Table 4.7.

Table 4.7 SIFT identified amino acid substitutions in tef that are not tolerated.

Tef proteins	Domains	Counterfactual mutation	Seq rep	SIFT score effect
EtRht1	PFYRE	E489Q	22	0.00***
	SAW	G552V	22	0.04**
EtBRI1	Kinase	K1047Q	8	0.10*
EtCOMT1	Dimerization	A48W	22	0.00***
		S70V	28	0.07*
EtCESA1	Transmembrane	T297M	44	0.06*

SIFT (sorting intolerant from tolerant); more deleterious mutations have values closer to 0.
* indicates significance at <0.1
** indicates high significance at <0.05
*** indicates very high significance at <=0.0

4.8 Structural Characterization of Tef Polypeptide Sequences Compared to Their Corresponding Rice Sequences

The predicted 3D structures of EtBRI1, EtCOMT1, EtCAD8C, EtCESA1, and EtCESA4 were generated using the SWISS MODEL platform, and are presented in Figure 4.7B, C, D, E & F. The predicted 3D structure of EtRht1 was already available in the UniProt database (Figure 4.7A).

Those polypeptides with uniquely tef amino acid variations that also had significant SIFT scores (Table 4.6) were further considered for structural comparison with their counterpart 3D structures of rice polypeptides (rice 3D structures were obtained from UniProt database). Accordingly, the 3D structural variations of four lodging-resistant proteins (EtRht1, EtBRI1, EtCOMT1, and EtCESA1) were compared to their corresponding rice 3D structures, as depicted in Figure 4.7A, B, C, & D.

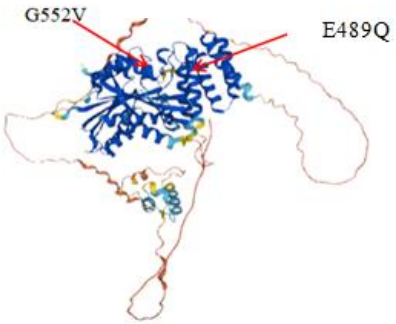
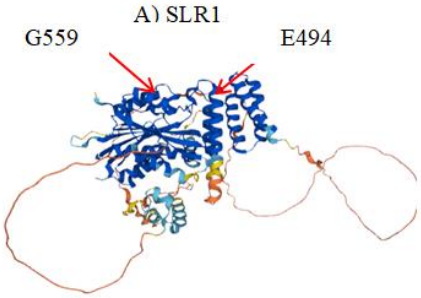
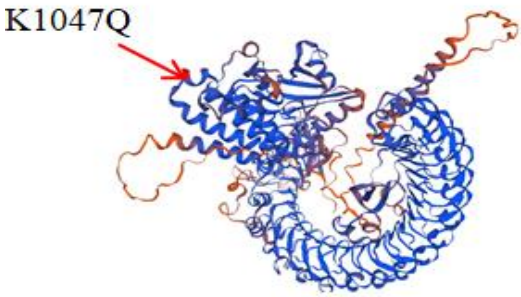
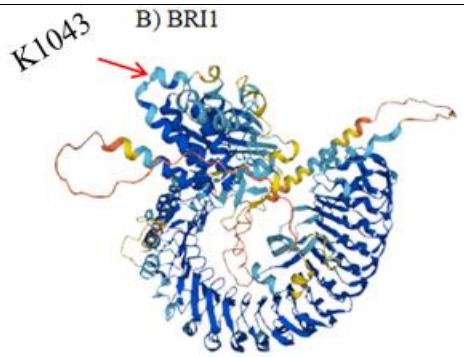
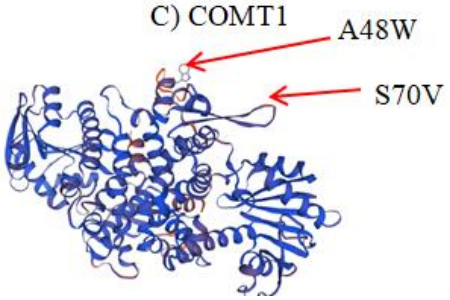
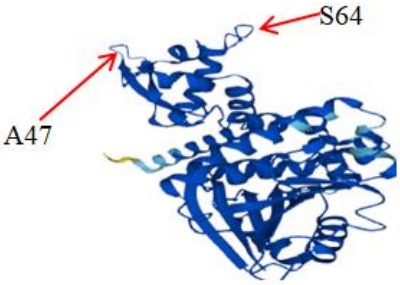
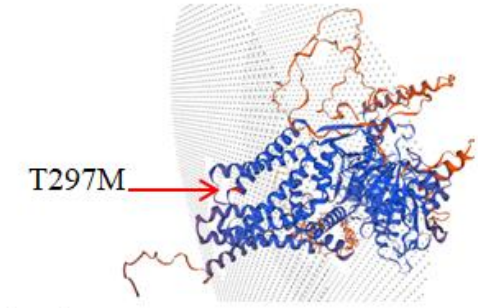
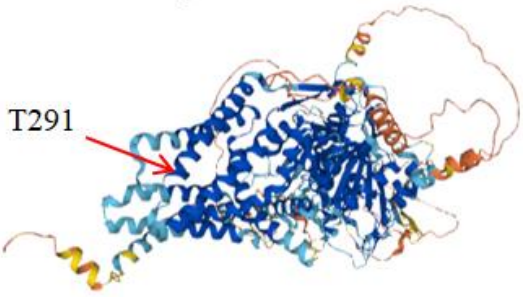
Variations in the 3D protein structure of EtRht1 within the GRAS domain were examined at two positions with significant SIFT scores: namely, Glu489Gln and Gly552Val, located within the PFYRE and SAW subdomains, respectively. In the EtRht1 PFYRE subdomain, the Glu489Gln is the 7th amino acid from the start of the predicted alpha-helix that extends from 483 to 503 aas and has 5.8 turns. In the corresponding OsSLR1 3D structure, the Glu494 is also the 7th amino acid in the alpha-helix that extends from 488 to 516 aas, but in this instance, the alpha-helix has 8.0 turns. In contrast in the SAW subdomain, both the EtRht1 Gly552Val and OsSLR1 Gly559 are the 2nd amino acids in the alpha-helix of 2.7 turns that extends from 551 to 560 aas and 558 to 567 aas, respectively.

A Lys1047Gln variation with significant SIFT score is observed in the kinase domain of EtBRI1 that resides within a short 1.9 turn alpha-helix which stretches from 1043 to 1049 aas and is book-ended by proline residues at both ends. The proline residues are immediately followed by alpha-helices and may thus be interrupting what would otherwise be a continuous alpha-helical structure. The position in OsBRI1 corresponding to the variation in tef is Lys1043 which resides in the alpha-helix that stretches from 1039 to 1044 with a 1.6 turn. Interestingly, proline residue is present (1039) on one end of the alpha-helix but the other end of the alpha-helix borders to an Asp1045 residue with the subsequent alpha-helix beginning with a Pro1046 residue (Figure 4.7B).

EtCOMT1 has two unique variations in its N-terminal dimerization domain with significant SIFT scores, namely, Ala48Trp and Ser70Val. The Ala48Trp resides near the edge of a

structured alpha-helix extending from 41 to 49 aas with a 2.5 turn. The corresponding wheat (because rice 3D structure was unavailable in UniProt) amino acid is Ala47 which resides at the very edge of a short 1.9 turn alpha-helix. Also in EtCOMT1, the variation Ser70Val is found in an irregular loop stretching from 67 to 76 aas that links two alpha-helices on either side. Similarly, in TaCOMT1, the corresponding residue is Ser64 which is found in an irregular loop stretching from 60 to 69 aas (Figure 4.7C and Table 4.7).

EtCESA1 exhibits the variation T297M with a significant SIFT score and that is located within an alpha-helix that forms the first TMD. This alpha-helix is 27 aas long (271-297) and makes a 7.5 turn. The corresponding position in OsCESA1 is Thr291 which is similarly located at the edge of a 27 aas long alpha-helix (265-291) with 7.5 turns (Figure 4.7D).

1) Tef	2) Rice
A) Rht1 	A) SLR1 
B) BRI1 	B) BRI1 
C) COMT1 	C) COMT1 (wheat) 
D) CESA1 	D) CESA1 

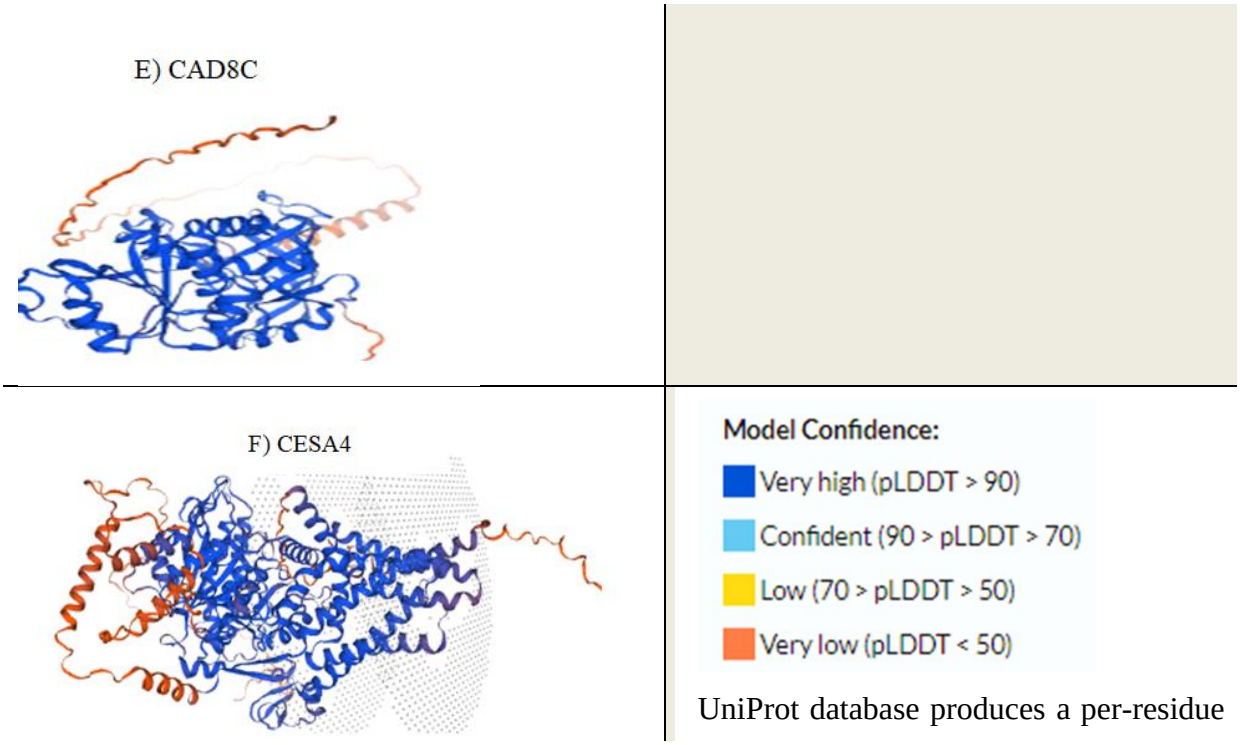


Figure 4.7. Predicted 3D structure of tef lodging-resistant proteins and their comparison to that of rice. Rice and wheat 3D structures were obtained from UniProt; and, only counterfactual variations with significant SIFT scores are considered. A) Locations of two amino acid variations in the GRAS domain of EtRht1 and OsSLR1: 494/489 PFYRE subdomain and 552/559 SAW subdomain (Both tef and rice 3D structures obtained from UniProt). B) Location of an amino acid variation in the kinase domain of the predicted 3D structure of EtBRI1 (1047) and OsBRI1 (1040). C) Locations of two amino acid variations in the N-terminal dimerization domain of predicted EtCOMT1 (48 & 70) and TaCOMT1 (47 & 64). [EtCOMT1 presented as a dimer]. D) Location of an amino acid variation in the TMD of predicted 3D structure of EtCESA1 (297) and OsCESA1 (291). (E) and (F) show predicted 3D structures of EtCAD8C and ETCESA4, respectively, but they have no counterfactual variations with significant SIFT score. Red arrow indicates the counterfactual variations sites with significant SIFT score. Column 1) tef polypeptide 3D structure, except Rht1, generated using the SWISS MODEL platform. Column 2) rice polypeptide 3D structure obtained from UniProt database. [*Panel C represents wheat 3D structure*].

5. DISCUSSION

The six lodging-resistant proteins of tef, as compared to wheat, barley and rice, show unique amino acid variations at critical positions that may likely compromise the function and structure thereby potentially affecting tef lodging-resistant attributes.

The examination of the sequences (gene, CDS, and protein), exon-intron structure, phylogenetic relationship, and percent GC content show that there is cross species overall conservation among the six lodging-resistant genes across all four species. This lends support that the computationally obtained tef sequences are indeed the orthologous of these genes in *E. tef.* Moreover, the high level of amino acid level identity of the tef polypeptides *vis-à-vis* their orthologous in the other three species (Table 4.5) is indicative that any possible changes to the polypeptide sequences in tef, in contrast to the other three species, may have biological significance (Pearson, 2013).

5.1 EtRht1: Unique Amino Acid Variations in the GRAS Domain and Their Potential Implication on Gibberellin Signaling

In wheat, the GRAS domain extends from position from 225 to 617 as reported by Peng *et al*. (1999) which coincides with predicted EtRht1's GRAS domain of 227 to 614. As per Peng *et al*. (1999), the polar rich residues region in wheat, which rest in between DELLA and GRAS domains, spans from 122 to 138, with a similar stretch also found in barley and rice (but in rice the stretch is one aa shorter). The predicted tef protein has three insertions at positions 122, 123 and 134 of proline (non-polar), glutamine (polar) and glycine (non-polar) residue, respectively, that are otherwise not present in wheat, barley and rice. This entails that the stretch rich in polar residues in tef is three aas longer and its proportion of polar residues drops slightly to 60%, as compared to 65% in wheat and barley. These variations in length and composition of EtRht1's polar rich region suggest that it may alter the property of the region, and hence, affect its hydrophobicity/hydrophilicity and its folding characteristics.

Despite perfect conservation in the two of the five GRAS subdomains, the variations observed in LHRI, PFYRE, and SAW subdomains (Figures 4.5 and 4.6A, and detailed in Table 4.6) indicate possible variation effects. The transcription regulation function of Rht1 resides in the LHR1 and PFYRE subdomains. Huang *et al*. (2023) reported that LHRI subdomain is essential for binding various DELLA-interacting transcription factors like Brassinazole-Resistant 1 (BZR1), Phytochrome-Interacting Factors (PIF3), Indeterminate Domain 3 (IDD3), and Teosinte Branched 1, Cycloidea, and Proliferating Cell Factor 1

(TCP14). Work done by Huang *et al*. (2023) in arabidopsis provides indirect evidence that mutation in the LHR1 subdomain results in significantly reduced interaction with the target TFs. In addition, the PFYRE subdomain interacts with H2A histone stabilizing the TF–Rht1–H2A complex at the target chromatin. Hence, both the LHR1 and PFYRE subdomains play a crucial role in Rht1-mediated transcription repression, as reported by Huang *et al*. (2023).

Unique tef Ala264Ser variation (Figure 4.5 and 4.6A and detailed in Table 4.6), which is a substitution of a non-polar to polar residue, is suggestive that the substitution may compromise LHRI subdomain's affinity to its interacting TFs BZR1, PIF3, IDD3 or TCP14.

The C-terminal PFYRE and SAW subdomains contribute to the structural or functional integrity of the proteins' GRAS domain (Hofmann, 2016). In wheat, PFYRE and SAW subdomains regulate GA responses by interacting with GA-Insensitive Dwarf 1 (GID1). Phenotypic alterations are demonstrated to be caused by a variety of mutations in these subdomains. Single aa change modifications within these subdomains, including the complete removal of the GRAS domain, induces of a loss-of-function slender phenotype due to growth suppression by the GRAS domain as reported by Phokas and Coates (2021).

Similarly, experimental evidence in SLR1 (the rice homologue of Rht1) of point mutations of both Ser196 (located in the polar residue rich area) and Ser510 (located in the PFYRE subdomain) to either Ala or Asp results in a defective phosphorylation site or constitutive phosphorylation site, respectively. GA signaling of plants containing the corresponding mutated SLR1 was examined by detecting the transcripts of GA synthesis genes GA20ox2 and GA3ox2 demonstrated that, after GA_3 treatment, constitutive phosphorylation of SLR1 at Ser510 or both Ser196/Ser510 would result in the suppressed GA signaling (enhanced expression of GA20ox2 and GA3ox2), whereas suppressed phosphorylation of SLR1 at either Ser196, Ser510, or both Ser196/Ser510 would, in fact, result in the significantly enhanced GA signaling (evidenced by suppressed expression of GA20ox2 and GA3ox2). These findings provide more evidence for the significance of SLR1 phosphorylation, particularly at Ser196 or Ser510, on the actions of SLR1 in GA signaling in rice, as demonstrated by Dai and Xue (2010). Position Ser196 in rice corresponds to position Ser194 in tef, a conserved residue. However, the phosphorylation site of Ser510 in rice corresponds to Gly504 in tef, a non-phosphorylate-able amino acid. The evidence from the rice study points to lack of a phosphorylation site in the corresponding tef position, therefore, potentially increasing GA signaling resulting in suppressed expression of GA20ox2 and GA3ox2. This finding is

indicative that the poor lodging-resistant trait of tef is not significantly affected by the loss of a phosphorylation site at position 504.

Aside of the single point mutagenesis study discussed above, Zhu *et al*. (2012), conversely employed a bulk mutagenesis approach using ethyl methanesulfonate (EMS) to induce transition mutations followed by high throughput sequencing to precisely identify the induced mutations. Those mutants with low SIFT scores were planted and evaluated for height phenotype. One mutation, with SIFT score of 0.00, Pro120Ser (in a Pro biased region) results in the substitution of a non-polar residue with a polar residue possibly affecting its function by altering its 3D folding.

The current study has identified a single significant counterfactual variation of Glu489Gln (charged to uncharged) (Figure 4.5 and 4.6A, and detailed in Table 4.6) which is predicted to severely affect protein, probably due to the alteration of the structural or functional integrity of the protein's GRAS domain. Of the two variations in the tef SAW domain, Gly552Val and Glu598Asp, only the former exhibited a significant counterfactual SIFT score (Table 4.7) signifying it may alter the maintenance of GRAS domain's structure. Additionally, a qualitative examination of the variation of Glu598Asp (in tef), which is conserved as Glu in wheat, barley and rice is shown in Figures 4.5 and 4.6A, and Table 4.6. Glutamic acid is characterized by a large size and high side chain flexibility, while aspartic acid is medium-sized with moderate side chain flexibility, but both are negatively charged. This contrasting physico-chemical property of the amino acids at this position suggests a deviation in function and/or structure of the protein in tef from that of the Rht1+- protein in the other three species.

In totality, the EtRht1 variations that are unique to tef, when considered in unison, their additive effect cannot be underestimated. These variations will likely have an effect on the GA mediated signaling pathway and its attendant phenotypic manifestations related to lodging attributes in tef.

5.2 EtBRI1's: Unique Amino Acid Variations in Multiple Domains and Their Potential Impact on Brassinosteroid Signaling Pathways

The many unique tef amino acid substitutions and deletions in multiple regions point to potential effects. A mutation in the Cys pair 1, which is known for mediating heterodimeric protein-protein interaction with BAK1 (Kwezi *et al*., 2007) results in a weak BRI1-dwarf phenotype in arabidopsis (Noguchi *et al*., 1999; Nam and Li, 2002). In this current study, variation in the Cys pair 1of tef is observed at position Arg51Lys (second residue in the

domain) may disrupt the formation of some heterodimers as in the case with BAK1, aligning with the findings of Kwezi *et al*. (2007).

In rice and arabidopsis, LRR mutations of the BRI1 cause a severe phenotype (Li and Chory 1997; Nakamura *et al*., 2006). Numerous unique variations are observed within the LRR domain and ID of EtBRI1, as depicted in Figures 4.5 and 4.6B, and detailed in Table 4.6. Of the 22 LRR domains, LRR3 and LRR6 exhibit multiple variations. Particularly, the eight variations observed in LRR3 (including 4 consecutive Ala insertions) are suspected will alter the secondary and tertiary structure, or both, of the LRR domain.

These substitutions may alter the secondary and tertiary structure or both of LRR and severely reduce its function in EtBRI1. The Leu464Phe substitution is proximal to the ID and may likely affect the ID. This is supported by findings by Nakamura *et al*. (2006) who reported that a Val491Met substitution in OsBRI1 of the LRR, just prior to the ID, causes an intermediate phenotype. Two substitutions in the ID, between LRR18 to LRR19, of EtBRI1 can potentially compromise signaling because it is a BR binding site. Research in rice shows that the substitutions in ID of Gly522Glu and Gly539Asp in rice exhibited a mild phenotype, while a corresponding substitution in arabidopsis resulted in the loss-of-function (Li and Chory, 1997; Nakamura *et al*., 2006).

Within the TMD domain, of the three variations in EtBRI1: Ser719Thr, Leu724Ile, and Ala738Phe two exhibit changes in the amino acids' physico-chemical properties (Figures 4.5 and 4.6B, and Table 4.6). The Ser719Thr substitution replaces a small-sized serine with a medium-sized threonine. Additionally, in Ala738Phe Alanine is small-sized with limited side chain flexibility, while phenylalanine is large-sized with moderate side chain flexibility. In the same domain in rice, a single nucleotide substitutions that result in an Asp759STOP causes a mutant, alters OsBRI1 leading to the most severe phenotype (Nakamura *et al*., 2006). It is thus conjectured that variations in the EtBRI1 TMD may seriously compromise its function.

In the KD domain in tef there is single unique mutation of Lys1047Gln (Figures 4.5 and 4.6B, and Table 4.6) which is predicted to significantly affect protein function, indicated by the probability SIFT score of 0.10 (Table 4.7). This substitution may potentially alter the kinase enzymatic activity and affect its trans-phosphorylation of BRI1 functionality. In the same domain, Nakamura *et al*. (2006) observed that a single amino acid substitution in rice caused a mild phenotype.

5.3 EtCOMT1: Unique Amino Acid Variations in N-Terminal Dimerization and C-Terminal Methyltransferase Domains and Their Potential Impact on Lodging Resistance

In tef, ten unique amino acid variations are observed within the N-terminal dimerization domain of the EtCOMT1 protein (Figure 4.5C and Table 4.4). Of note is that four substitutions at positions 58, 62, 73 and 76 all represent physico-chemical marked changes from a non-polar to polar amino acid. Further, the counterfactual examination of two other substitutions, Ala48Trp and Ser70Val, are both significant (Table 4.7) and probably compromise EtCOMT1 function.

As all six substitutions are coincidental to EtCOMT1's N-terminal dimerization domain, it increases the possibility that, in conjunction, they may compromise the dimerization capability, and, hence, negatively affect the function of the protein. Dimerization was shown to be key for the proper formation of the core stable structure and, that the dimerization, in turn, is important for the catalytic function of COMT1 and other related proteins in the same family (Zhou *et al*., 2010; Sattler *et al*., 2012).

In the C-terminal methyltransferase domain, the substitutions Glu271Ala, Asp282Ala and His294Gly (Figures 4.5 and 4.6C, and detailed in Table 4.6) all exhibits a change from charged to non-charged amino acids; and, in the first two, medium-to-large sized polar amino acids are changed to small non-polar amino acids. All three variations reside in the substrate and SAM binding regions as well as the enzymatic active site. Though the proton acceptor site, His273, required for the methyl transfer activity is conserved in tef, it is here suggested that these three identified variations may perturb the enzymatic activity of EtCOMT1 by affecting the binding of the substrate and/or SAM. Several studies show that variations in the methyltransferase domain of COMT1 greatly reduce enzymatic activity resulting in reduced lignin composition (Vignols *et al*., 1995; Zhou *et al*., 2010; Sattler *et al*., 2012).

5.4 EtCAD8C: Single Unique Amino Acid Variation in the $NADP^+$ Binding Site and Its Potential Implication for Lignin Biosynthesis

In wheat, barley and rice, the CAD8C protein region spanning amino acids 374-376 (NCV) is known to serve as an NADP+ binding site (UniProt manual assertion by sequence similarity). The computationally predicted EtCAD8C has a Cys400Ser variation in the center of the triad NADP+ binding site (Figures 4.5 and 4.6D, and Table 4.6). Substitution of the conserved cysteine (hydrophophic) by serine (hydrophilic) in EtCAD8C may compromise the catalytic

activity of the encoded protein, leading to changes in lignin content, structure, and composition, ultimately resulting in higher saccharification efficiency. Several studies also show that variations in the $NADP^+$ binding region of CAD proteins affect its binding affinity for $NADP^+$ cofactor, as well as alter the interaction between the flexible motif of CAD and $NADP^+$ phosphate backbone (Sattler *et al*., 2009; Xiong *et al*., 2020).

5.5 EtCESA1: Unique Amino Acid Variations in the Zinc-Finger, TMDs, and Cytoplasmic Domains and Their Potential Implication on Cellulose Biosynthesis

CESA1 is plasma membrane protein with 8 TMDs (Kikuchi *et al*., 2003) which are thought to form a channel in the plasma membrane, facilitating the extrusion of the newly synthesized glucan chains (Kaur *et al*., 2016) towards formation of the primary cell wall. In addition to its TMDs, it comprises two important cytoplasmic domains; namely, zinc-finger and coiled-coil. In rice, an extracellular domain is known to possess one GlcNAc glycosylation site at position 948, and seven phosphoserine sites, of which the first five are located in the first N-terminus cytoplasmic region. The zinc-finger domain (spanning from position 41 to 87 - rice numbering) is described by Cheng *et al*. (2005) as a small, functionally autonomous, folded region that utilizes cysteine and/or histidine residues to coordinate one or more zinc ions, stabilizing its structure.

Recessive mutations in the zinc-finger domain of CESA1, or TMDs, results in a phenotype that causes the failure of CESA1 from interacting with other subunits to form the CSC (Ma *et al*., 2021). Of the two unique variations in zinc-finger domain of tef, Pro69Ala and Lys76Arg (Figures 4.5 and 4.6E, and detailed in Table 4.6), the latter is not expected to have register significant phenotypic alterations as both amino acids have very similar physico-chemical properties. However, in the former, the substitution of the hydrophilic and medium in sized proline is substituted in tef by the hydrophobic and small in size alanine which may potentially disrupt the coordination of zinc ions and, hence the stability of the protein structure, compromising the interaction with other CESA subunits to form a CSC (Ma *et al*., 2021).

In EtCESA1, a substitution of Ser161Thr and a deletion of serine (between positions 165-166 - tef numbering) (Figures 4.5 and 4.6E, and Table 4.6) are otherwise experimentally demonstrated phosphoserine sites in rice. This deletion between positions 165-166 in conjunction with the Ser161Thr substitution in EtCESA1 may alter the complex phosphorylation process potentially influencing subcellular trafficking and stability, and ultimately CSC activity itself, aligning with the findings of Speicher and Wallace (2018).

In this study, a substitution in EtCESA1 of Ser161Thr is observed. Additionally, a deletion of serine is evident in tef between position 165-166 (tef numbering) (Figures 4.5 and 4.6E, and Table 4.6) which is otherwise conserved in wheat, barley and rice; and, has been experimentally demonstrated in rice as a phosphoserine site. This deletion between positions 165-166 in conjunction with the Ser161Thr substitution in EtCESA1 may alter the complex phosphorylation process potentially influencing subcellular trafficking and stability, and ultimately CSC activity itself, aligning with the findings of Speicher and Wallace (2018).

The three unique tef variations, namely, Thr297Met, Arg301Asn and Tyr304Phe occur in the first TMD, and a Ser898Thr occurs in the fourth TMD (Figures 4.5 and 4.6E, and Table 4.6). The substitutions at positions 297 and 304 are polar to non-polar, while Arg301Asn and Ser898Thr yields polar to polar aa substitutions. That said, the helical domains that traverse the lipid bi-layer are expected to favor to be non-polar and hence, no significant effect is expected from these substitutions. However, a counterfactual examination of Thr297Met substitution in tef, *vis-à-vis* the perfectly conserved Thr of this site in the other three cereals predicts a different outcome.

Threonine is characterized as polar, of medium size, and with low side chain flexibility, while methionine is non-polar, large in size, and with high side chain flexibility. This substitution is predicted to significantly affect protein function (Table 4.7). This probably alters the ability of EtCESA1 to properly discharge its function in facilitating the extrusion of newly synthesized glucan chain, thereby negatively impacting the synthesis of cellulose microfibrils in the primary cell wall. A study by Harris *et al*. (2012) showed that, in the same TMD4 of arabidopsis, a Ala903Val (a non-polar to non-polar) substitution increases CSC mobility and sacchirification, and decreases cellulose crystalinity resulting in the increased rate of cellulose polymerization. This improved property in arabidopsis where the substituted aa remains non-polar contrasts with that of a polar-by-polar substitution predicted in EtCESA1.

With respect to the cytoplasmic domains of CESA1, of the total of seven unique variations in EtCESA1, four, namely, Asn333Glu, Tyr345Phe, Arg347Lys and Glu350Arg occur in the second cytoplasmic domain. All except the Tyr345Phe are polar to polar variations, whereas the Tyr345Phe is a polar-to-non-polar substitution. The second cytoplasmic domain is the largest such domain (sometimes also called central cytoplasmic domain) and, based on computational analysis, believed to be involved in interactions between CESA proteins that

are likely to be required for oligomerization to assemble into the larger CSC required for cellulose synthesis (Vandavasi, *et al*. 2016). Also, a study in *Arabidopsis thaliana* AtCESA1 of a single missense mutation in the second cytoplasmic domain of Ala549Val (both non-polar) has been demonstrated to exhibit a phenotype of impaired root elongation due to less organized cellulose microfibrils (Slabaugh *et al*. and Kaur *et al*., 2016). Given the prominence of this domain, it is here speculated that a polar-to-non-polar change in this domain may result in potentially affecting its regular function.

CESA1 is determined, as per the rules that define candidate N-linked glycosylation sites (UniProt), to possess such a site for the stated post-translational modification at Asn948 (rice numbering) in an extracellular domain that is also conserved in wheat and barley. The EtCESA1 has an Asn963Ser substitution that potentially abrogates N-linked glycosylation, by instead having an amino acid that can only be O-linked glycosylated. However, Gillmor *et al*. (2002), using large scale screen for mutations, demonstrate that, though proper N-linked glycosylation is important for cellulose synthesis that this effect does not rely on the N-linked glycosylation of the CESA1. Hence, the loss of an N-linked glycosylation site in EtCESA1 is not suspected to have a major impact on the cellulose synthesis process.

5.6 EtCESA4: Two Unique Amino Acid Variations in the Extracellular and TMD Domains and Their Potential Implication on Cellulose Biosynthesis

EtCESA4 is found to exhibit variations of Ala869Ser and Asp907Ala (numbering in tef) (illustrated in Figures 4.5 and 4.6F, and Table 4.6) that reside in an extracellular and TMD6 domains, respectively (see Figure 4.6F). TMD6 is one of eight TMDs that act as catalytic subunit within the CSC, playing a pivotal role in glucan chain elongation. Studies of missense mutations in OsCESA4 (Scheible *et al.*, 2001; Chen *et al*., 2005: Zhang *et al.*, 2009) of Leu802Phe, Gly858Arg and Gly998Asp, respectively, residing in TMD4, extracellular domain immediately after TMD5 and TMD7 were previously reported to influence such phenotypes as decreased cellulose contents in the membrane and reducing the mechanical properties of the plant. This phenotypic effect can be attributed to the reduced presence of the OsCESA4 in the plasma membrane (likely due to a defect in the secretion process of the CSC) and, modification of its normal conformational structure.

The variations observed in EtCESA4 are not spatially coincidental to the missense mutations described in rice and hence, may not (if at all) manifest the same phenotypic alterations described. Nonetheless, the Ala869Ser and Asp907Ala in EtCESA4 point to a major change in the physico-chemical properties of the aas. Alanine is a hydrophobic, non-polar and with

limited side chain flexibility amino acid; while serine is hydrophilic, polar and a low side chain flexibility amino acid. Aspartic acid, on the other hand, is polar, negatively charged, medium-sized, hydrophilic, and possesses moderate side chain flexibility; while alanine is non-polar, small in size, hydrophobic, and has limited side chain flexibility. This leads us to conjecture that these variations may exert some alteration in the protein function/structure, particularly with respect to the Asp907Ala that is found in a TMD.

5.7 Structural Analysis of Significant Counterfactual Variations in Four Tef Lodging-Resistant Proteins

The computationally predicted 3D structures of the five predicted polypeptide sequences, except EtRht1, in tef (illustrated in Figure 4.7) are reported here for the first time: the predicted 3D structure of EtRht1 was obtained from UniProt. To the best of my knowledge, there is no experimentally established 3D structure of all the six tef lodging-resistant proteins.

Only EtRht1, EtBRI1, EtCOMT1, and EtCESA1 proteins' predicted 3D structures were chosen (because of their counterfactual significant SIFT scores) for further structural level comparison with their homologues in rice (or wheat for COMT1) 3D structures reposed in the UniProt database (Figure 4.7 and Table 4.7).

In EtRht1, of the two counterfactual variations, Glu489Gln found in the PFYRE subdomain of the GRAS domain *vis-à-vis* OsSLR1, exhibits a markedly shorter alpha-helix (see Figure 4.7A) with important implications on the integrity of the GRAS domain (Hofmann, 2016) as well as interaction with H2A histone stabilizing the TF-Rht1-H2A complex at the target chromatin (Huang *et al*., 2023). Study by Jobson *et al.* (2021) in the same subdomain in wheat, but at Gly543Arg likely had a detrimental impact on the structure and function of the Rht1 protein by inhibiting an already weak DELLA/GID1 binding interaction. This is suggested by values closer to zero within the SIFT score range of 0 to 0.62. It is thus very likely such a dramatic change in the length of the alpha-helix in the EtRht1's PFYRE subdomain may affect its role as a bridge between the chromatin and TF thereby influencing the expression/repression level of GA. When not repressed by Rht1, expression of GA results in a tall phenotype (Tong *et al*., 2014).

EtBRI1 contains one significant counterfactual variation of Lys1047Gln in its KD. Structurally, this variation resulted in an only slightly longer alpha-helical structure in tef as compared to rice. However, it is notable that EtBRI1 α-helix has rigid borders set by the proline residues, while in its rice counterpart; only one end is bounded by a proline. The

presence of these prolines in EtBRI1 may provide a structural constraint that entails less flexibility in the kinase domain which is required for substrate binding. Work by Li and Chory (1997) in arabidopsis shows a Gly1048Asp in the same kinase domain's activation loop results in potential disruption of substrate binding or displacement of the catalytic Asp1009.

Similar to the other KD in the three cereal crops, the arabidopsis position 1047 is lysine, which in turn corresponds to the identified EtBRI1 variation Lys1047Gln. Study by Sun *et al*. (2017) showed that two mutations in Glu1078Lys and Asp1139Asn near the end of the KD result in strong dwarfing phenotype. This finding suggests that point mutations in this domain may involve structural modifications that dramatically affect the KD's function. It is plausible that the EtBRI1 1047Gln, which is a non-charged polar residue, in contrast to the positive charged polar lysine, might contribute to structural changes that affect the KD creating increased sensitivity to brassenosteroid and, hence, resulting in non-dwarf phenotype in tef.

The two significant counterfactual variations in EtCOMT1 (Ala48Trp and Ser70Val) reside in its N-terminal dimerization domain (see Table 4.7). The EtCOMT1 Ala48Trp is found within a short α-helix and comparing it to the same region in wheat shows it has an even shorter α-helix that ends with an Ala47 that corresponds to the 48^{th} position in tef. Further, EtCOMT1's Ser70Val is found in an irregular loop between two α-helical structures and is one amino acid shorter than the same irregular loop found in the wheat polypeptide (see Table 4.7 and Figure 4.7C). Together, the presence of these two unique variations in the N-terminal dimerization domain in tef point to their possible disruptive effect on the secondary structure required for proper dimer interface. This conjecture is supported by the finding in sorghum that showed that Ala71Val in its N-terminal dimerization domain resulted in reduced secondary structure resulting in reduced lignin content and altered lignin sub-unit composition (Green *et al*., 2014). Therefore, it is likely that the two unique tef variations could cause the same phenotype resulting in weaker primary cell wall, contributing to susceptibility to lodging.

EtCESA1 exhibits a Thr297Met variation that is located within an α-helix of 7.5 helical turns which makes up the first of eight TMDs. Structurally, this is similar to the first α-helix of OsCESA1 (as depicted in Figure 4.7D). However, noting that this is a significant counterfactual variation there is high probability that it will cause intolerable damage. A closer examination of the specific variation shows that substitution of threonine (polar) by

methionine (non-polar) will likely increase the hydrophobic property of this TMD and thereby increase the abundance of CESA1 in the TMD. Whether, and how, the relative abundance of CESA1 in the plasma membrane affects cellulose synthesis and the structure of the secondary cell wall remains to be investigated.

6. CONCLUSION

Tef (*Eragrostis tef*), an indigenous cereal crop, holds a crucial role in ensuring food security for Ethiopia. However, it is vulnerable to lodging. *In silico* comparison against wheat, barley and rice homologues shows many unique tef variations in the studied six lodging-resistant genes; some of which are present in functionally or structurally important domains. Qualitative, quantitative and structural analyses of these unique variations indicate that these variations may contribute to tef lodging susceptibility with varying degrees of magnitude. The aggregate effect of multiple identified unique variations can potentially impact the genetic component of quantitative traits associated with lodging resistance. Lodging resistance in crops is influenced by multiple quantitative traits, such as stem strength and thickness, root system architecture, plant height, stem diameter, culm anatomy, leaf angle, grain weight, and tiller distribution. These traits are controlled by multiple genes and affected by environmental factors, making them complex and continuously variable.

7. RECOMMENDATIONS

It is necessary to experimentally validate the effects of the discovered unique tef variations in the six lodging-resistant genes on the morphological and biochemical characteristics of tef as it relates to its lodging resistance/susceptibility. Experimental approaches such as site-directed mutagenesis using CRISPR-CAS9 may be employed to substitute the unique tef variations to their conserved amino acids of wheat, barley and rice to investigate the phenotypic characteristics of the mutated tef variants. Given the number of genes and substantial number of unique tef variations per gene, an experimental validation, as suggested, may require investigating one, a few or all genes simultaneously.

Also, it still remains to be shown how these multiple unique variations interact with the growth conditions found in traditionally tef growing regions of Ethiopia. Furthermore, those promising results under laboratory conditions should be supplemented by field trials to assess the effects of the experimental interventions in real-world farming scenarios.

Finally, as means to gain an in-depth understanding of how these proteins function at the molecular level, further research will be required to elucidate the precise mechanism of how these lodging-resistant proteins molecules confer lodging resistance in tef.

REFERENCES

Achard, P., & Genschik, P. (2009). Releasing the brakes of plant growth: how GAs shutdown DELLA proteins. *Journal of experimental botany*, *60*(4), 1085-1092.

Alaunyte, I., Stojceska, V., Plunkett, A., Ainsworth, P., & Derbyshire, E. (2012). Improving the quality of nutrient-rich Tef (*Eragrostis tef*) breads by combination of enzymes in straight dough and sourdough breadmaking. *Journal of Cereal Science*, *55*(1), 22-30.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Appenzeller, L., Doblin, M., Barreiro, R., Wang, H., Niu, X., Kollipara, K., & Dhugga, K. S. (2004). Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. *Cellulose*, *11*, 287-299.

Ashikari, M., Sasaki, A., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Datta, S., & Matsuoka, M. (2002). Loss-of-function of a rice gibberellin biosynthetic gene, GA20 oxidase (GA20ox-2), led to the rice 'green revolution'. *Breeding Science*, *52*(2), 143-150.

Assefa, K., Chanyalew, S., & Metaferia, G. (2011). Conventional and molecular tef breeding. In *Achievements and prospects of tef improvement, Proceedings of the second international workshop* (pp. 33-51).

Assefa, K., Chanyalew, S., & Tadele, Z. (2017). Tef, *Eragrostis tef* (Zucc.) trotter. *Millets and sorghum: Biology and genetic improvement*, 226-266. Wiley.

Assefa, K., Yu, J. K., Zeid, M., Belay, G., Tefera, H., & Sorrells, M. E. (2011). Breeding tef [*Eragrostis tef* (Zucc.) trotter]: conventional and molecular approaches. *Plant breeding*, *130*(1), 1-9.

Bai, M. Y., Zhang, L. Y., Gampala, S. S., Zhu, S. W., Song, W. Y., Chong, K., & Wang, Z. Y. (2007). Functions of OsBZR1 and 14-3-3 proteins in brassinosteroid signaling in rice. *Proceedings of the National Academy of Sciences*, *104*(34), 13839-13844.

Bayable, M., Tsunekawa, A., Haregeweyn, N., Ishii, T., Alemayehu, G., Tsubo, M., & Masunaga, T. (2020). Biomechanical properties and agro-morphological traits for improved lodging resistance in Ethiopian teff (*Eragrostis tef* (Zucc.) Trottor) accessions. *Agronomy*, *10*(7), 1012.

Belay, G., Tefera, H., Getachew, A., Assefa, K., & Metaferia, G. (2008). Highly client-oriented breeding with farmer participation in the Ethiopian cereal tef [Eragrostis tef (Zucc.) Trotter]. *African Journal of Agricultural Research*, *3*(1), 022-028.

Bennetzen, J. L., Smith, S. M., Yuan, Y. N., & Groth, D. (2009). Opening new avenues for the improvement of orphan crops in a time of rapid and potentially catastrophic change in worldwide agriculture. *In Proceedings of the New Approaches to Plant Breeding of Orphan Crops in Africa. Proceedings of an International Conference, Bern, Switzerland*, 11–19.

Berehe, T. (1975). Breakthrough in tef breeding technique. FAO Inf. *Bull., CerealImprovement and Production, Near East Project (3)*, 11-23.

Berry, P. M., Spink, J. H., Gay, A. P., & Craigon, J. (2003). A comparison of root and stem lodging risks among winter wheat cultivars. *The Journal of Agricultural Science*, *141*(2), 191-202.

Berry, P. M., Sterling, M., Spink, J. H., Baker, C. J., Sylvester-Bradley, R., Mooney, S. J., & Ennos, A. R. (2004). Understanding and reducing lodging in cereals. *Advances in agronomy*, *84*(04), 215-269.

Beyene, G., Chauhan, R. D., Villmer, J., Husic, N., Wang, N., Gebre, E., & MacKenzie, D. J. (2022). CRISPR/Cas9-mediated tetra-allelic mutation of the 'Green Revolution'semi-

dwarf-1 (SD-1) gene confers lodging resistance in tef (*Eragrostis tef*). *Plant biotechnology journal*, *20*(9), 1716-1729.

Bi, C., Chen, F., Jackson, L., Gill, B. S., & Li, W. (2011). Expression of lignin biosynthetic genes in wheat during development and upon infection by fungal pathogens. *Plant Molecular Biology Reporter*, *29*, 149-161.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature protocols*, *4*(1), 1-13.

Bouma, J., & Ohnoutka, Z. (1991). Importance and application of the mutant 'Diamant' in spring barley breeding. In *Plant mutation breeding for crop improvement, 1(3)*, 127-133.

Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology*, *268*(1), 78-94.

Cannarozzi, G., Plaza-Wüthrich, S., Esfeld, K., Larti, S., Wilson, Y. S., Girma, D., & Tadele, Z. (2014). Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). *BMC genomics*, *15*(1), 1-21.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., & Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research*, *18*(1), 188-196.

Central Statistical Authority (CSA), 2022. Central Statistical Agency. The Federal Democratic Republic of Ethiopia, Central Statistical Agency, Agricultural Sample Survey 2021/22 (2014 E.C.),Volume I, Report on Area and Production of Major Crop.

Chandler, P. M., & Harding, C. A. (2013). ‘Overgrowth ’mutants in barley and wheat: new alleles and phenotypes of the ‘Green Revolution’ DELLA gene. *Journal of Experimental Botany*, *64*(6), 1603-1613.

Chen, X., Shi, C., Yin, Y., Wang, Z., Shi, Y., Peng, D., & Cai, T. (2011). Relationship between lignin metabolism and lodging resistance in wheat. *Acta Agronomica Sinica*, *37*(9), 1616-1622.

Chen, Z., Hong, X., Zhang, H., Wang, Y., Li, X., Zhu, J. K., & Gong, Z. (2005). Disruption of the cellulose synthase gene, AtCesA8/IRX1, enhances drought and osmotic stress tolerance in Arabidopsis. *The Plant Journal*, *43*(2), 273-283.

Cheng, C. H., Chung, M. C., Liu, S. M., Chen, S. K., Kao, F. Y., Lin, S. J., & Chow, T. Y. (2005). A fine physical map of the rice chromosome 5. *Molecular Genetics and Genomics*, *274*, 337-345.

Chikkagoudar, S. (2010). *Algorithms in comparative genomics*. New Jersey Institute of Technology.

Chono, M., Honda, I., Zeniya, H., Yoneyama, K., Saisho, D., Takeda, K., & Watanabe, Y. (2003). A semi-dwarf phenotype of barley uzu results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. *Plant physiology*, *133*(3), 1209-1219.

Colebrook, E. H., Thomas, S. G., Phillips, A. L., & Hedden, P. (2014). The role of gibberellin signalling in plant responses to abiotic stress. *Journal of experimental biology*, *217*(1), 67-75.

Costanza, S. H., Dewet, J. M. J., & Harlan, J. (1979). Literature review and numerical taxonomy of *Eragrostis tef* (T’ef). *Economic Botany*, *33*, 413-424.

Crook, M. J., & Ennos, A. R. (1994). Stem and root characteristics associated with lodging resistance in four winter wheat cultivars. *The Journal of Agricultural Science*, *123*(2), 167-174.

Cufodontis, G. (1974). Enumeration Planetarium Aethopiae Spermatophyta, *Bulletin du Jardin Botanique*, Brussels.

Dai, C., & Xue, H. W. (2010). Rice early flowering1, a CKI, phosphorylates DELLA protein SLR1 to negatively regulate gibberellin signalling. *The EMBO Journal*, *29*(11), 1916-1927.

Del Río, J. C., Rencoret, J., Prinsen, P., Martínez, Á. T., Ralph, J., & Gutiérrez, A. (2012). Structural characterization of wheat straw lignin as revealed by analytical pyrolysis, 2D-NMR, and reductive cleavage methods. *Journal of agricultural and food chemistry*, *60*(23), 5922-5935.

Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, *23*(6), 673-679.

Dong, X. C., Qian, T. F., Chu, J. P., Zhang, X., Liu, Y. J., Dai, X. L., & He, M. R. (2023). Late sowing enhances lodging resistance of wheat plants by improving the biosynthesis and accumulation of lignin and cellulose. *Journal of Integrative Agriculture*, *22*(5), 1351-1365.

Ebba, T. (1975). Tef (*Eragrostis tef*) cultitabs: morphology and classification. Eiar.gov.et .

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, *17*, 368-376.

Gale, M. D., & Marshall, G. A. (1976). The chromosomal location of Gai1 and Rht1, genes for gibberellin insensitivity and semi-dwarfism, in a derivative of Norin 10 wheat. *Heredity*, *37*(2), 283-289.

Gayathiri, E., Prakash, P., Kumaravel, P., Jayaprakash, J., Ragunathan, M. G., Sankar, S., & Govindasamy, R. (2023). Computational approaches for modeling and structural design of biological systems: A comprehensive review. Progress in Biophysics and Molecular Biology.

Gillmor, C. S., Poindexter, P., Lorieau, J., Palcic, M. M., & Somerville, C. (2002). α-Glucosidase I is required for cellulose biosynthesis and morphogenesis in Arabidopsis. *The Journal of cell biology*, *156*(6), 1003-1013.

Green, A. R., Lewis, K. M., Barr, J. T., Jones, J. P., Lu, F., Ralph, J., & Kang, C. (2014). Determination of the structure and catalytic mechanism of Sorghum bicolor caffeic acid O-methyltransferase and the structural impact of three brown midrib12 mutations. *Plant physiology*, *165*(4), 1440-1456.

Gruszka, D. (2020). Exploring the brassinosteroid signaling in monocots reveals novel components of the pathway and implications for plant breeding. *International Journal of Molecular Sciences*, *21*(1), 354.

Gugsa, L., Sarial, A. K., Lörz, H., & Kumlehn, J. (2006). Gynogenic plant regeneration from unpollinated flower explants of *Eragrostis tef* (Zuccagni) Trotter. *Plant cell reports*, *25*, 1287-1293.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, *52*(5), 696-704.

Günther, T., & Schmid, K. J. (2010). Deleterious amino acid polymorphisms in Arabidopsis thaliana and rice. *Theoretical and Applied Genetics*, *121*, 157-168.

Guryev, V., Berezikov, E., Malik, R., Plasterk, R. H., & Cuppen, E. (2004). Single nucleotide polymorphisms associated with rat expressed sequences. *Genome research*, *14*(7), 1438-1443.

Harris, D. M., Corbin, K., Wang, T., Gutierrez, R., Bertolo, A. L., Petti, C., & DeBolt, S. (2012). Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903V and CESA3T942I of cellulose synthase. *Proceedings of the National Academy of Sciences*, *109*(11), 4098-4103.

Hayat, S., Irfan, M., & Ahmad, A. (2011). Brassinosteroids: under bioti stress. brassinosteroids: a class of plant hormone, 345-360. Springer link.

Hedden, P. (2003). The genes of the Green Revolution. *TRENDS in Genetics*, *19*(1), 5-9.

Hofmann, N. R. (2016). A structure for plant-specific transcription factors: the GRAS domain revealed. *The Plant Cell*, 28(5), 993–994.

Hothorn, M., Belkhadir, Y., Dreux, M., Dabi, T., Noel, J. P., Wilson, I. A., & Chory, J. (2011). Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature*, *474*(7352), 467-471.

Hou, Q., Saima, S., Ren, H., Ali, K., Bai, C., Wu, G., & Li, G. (2019). Less conserved LRRs is important for BRI1 folding. *Frontiers in Plant Science*, *10*, 634.

Houston, K., Burton, R. A., Sznajder, B., Rafalski, A. J., Dhugga, K. S., Mather, D. E., & Fincher, G. B. (2015). A genome-wide association study for culm cellulose content in barley reveals candidate genes co-expressed with members of the CELLULOSE SYNTHASE A gene family. *PloS one*, *10*(7), e0130890.

Huang, X., Tian, H., Park, J., Oh, D. H., Hu, J., Zentella, R., & Sun, T. P. (2023). The master growth regulator DELLA binding to histone H2A is essential for DELLA-mediated global transcription regulation. *Nature plants*, *9*(8), 1291-1305.

Hyles, J., Vautrin, S., Pettolino, F., MacMillan, C., Stachurski, Z., Breen, J., & Spielmeyer, W. (2017). Repeat-length variation in a wheat cellulose synthase-like gene is associated with altered tiller number and stem cell wall composition. *Journal of experimental botany*, *68*(7), 1519-1529.

Ji, H., Han, C. D., Lee, G. S., Jung, K. H., Kang, D. Y., Oh, J., & Kim, K. H. (2019). Mutations in the microRNA172 binding site of Super Numerary Bract (SNB) suppress internode elongation in rice. *Rice, 12*(1), 1-14.

Jobson, E. M., Martin, J. M., Sharrock, R., Hogg, A. C., & Giroux, M. J. (2021). Identification and molecular characterization of novel Rht-1 alleles in hard red spring wheat. *Crop Science, 61*(2), 1030-1037.

Junier, T., & Pagni, M. (2000). Dotlet: diagonal plots in a web browser. *Bioinformatics, 16*(2), 178-179.

Kashiwagi, T., & Ishimaru, K. (2004). Identification and functional analysis of a locus for improvement of lodging resistance in rice. *Plant physiology, 134*(2), 676-683.

Kaur, S., Dhugga, K. S., Gill, K., & Singh, J. (2016). Novel structural and functional motifs in cellulose synthase (CesA) genes of bread wheat (*Triticum aestivum*, L.). *PLoS One, 11*(1), e0147046.

Kedisso, E. G. (2012). *Manipulation of gibberellin biosynthesis for the control of plant height in Eragrostis tef for lodging resistance* (Doctoral dissertation, University of Pretoria).

Ketema, S. (1997). *Tef-Eragrostis tef (Zucc.)* (Vol. 12). Bioversity International.

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., & Hayashizaki, Y. (2003). Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *science, 301*(5631), 376-379.

Kinoshita, T., Caño-Delgado, A., Seto, H., Hiranuma, S., Fujioka, S., Yoshida, S., & Chory, J. (2005). Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature, 433*(7022), 167-171.

Kwezi, L., Meier, S., Mungur, L., Ruzvidzo, O., Irving, H., & Gehring, C. (2007). The *Arabidopsis thaliana* brassinosteroid receptor (AtBRI1) contains a domain that functions as a guanylyl cyclase in vitro. *PLoS one, 2*(5), e449.

Li, D., Wang, L., Wang, M., Xu, Y. Y., Luo, W., Liu, Y. J., & Chong, K. (2009). Engineering OsBAK1 gene as a molecular tool to improve rice architecture for high yield. *Plant biotechnology journal*, *7*(8), 791-806.

Li, J., & Chory, J. (1997). A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell*, *90*(5), 929-938.

Li, Q., Fu, C., Liang, C., Ni, X., Zhao, X., Chen, M., & Ou, L. (2022). Crop lodging and the roles of lignin, cellulose, and hemicellulose in lodging resistance. *Agronomy*, *12*(8), 1795.

Liang, S., Xu, S., Qu, D., Yang, L., Wang, J., Liu, H., & Zheng, H. (2022). Identification and functional analysis of the caffeic acid O-methyltransferase (COMT) gene family in rice (Oryza sativa l.). *International Journal of Molecular Sciences*, *23*(15), 8491.

Liu, Q., & Fu, X. (2023). Can heterotrimeric G proteins improve sustainable crop production and promote a more sustainable Green Revolution?. *The Innovation Life*, *1*(2), 100024-1.

Ma, J., Jiang, Q. T., Zhao, Q. Z., Zhao, S., Lan, X. J., Dai, S. F., ... & Zheng, Y. L. (2013). Characterization and expression analysis of waxy alleles in barley accessions. *Genetica*, *141*, 227-238.

Ma, X., Li, C., Huang, R., Zhang, K., Wang, Q., Fu, C., & Deng, X. (2021). Rice Brittle Culm19 encoding cellulose synthase subunit CESA4 causes dominant brittle phenotype but has no distinct influence on growth and grain yield. *Rice*, *14*, 1-12.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, *47*(W1), W636-W641.

Matsushima, N., & Miyashita, H. (2012). Leucine-rich repeat (LRR) domains containing intervening motifs in plants. *Biomolecules*, *2*(2), 288-311.

Merga M. (2018). Progress, Achievements and Challenges of Tef Breeding in Ethiopia. *J Agri Sci Food Res*, 9(1): 204.

Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., Maehara, Y., & Minobe, Y. (2002). Positional cloning of rice semi-dwarfing gene, SD-1: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA research*, *9*(1), 11-17.

Naaz, H., Pandey, V. P., Singh, S., & Dwivedi, U. N. (2013). Structure–function analyses and molecular modeling of caffeic acid-O-methyltransferase and caffeoyl-C o A-O-methyltransferase: Revisiting the basis of alternate methylation pathways during monolignol biosynthesis. *Biotechnology and Applied Biochemistry*, *60*(2), 170-189.

Nadolska-Orczyk, A., Rajchel, I. K., Orczyk, W., & Gasparis, S. (2017). Major genes determining yield-related traits in wheat and barley. *Theoretical and Applied Genetics*, *130*, 1081-1098.

Nakamura, A., Fujioka, S., Sunohara, H., Kamiya, N., Hong, Z., Inukai, Y., & Matsuoka, M. (2006). The role of OsBRI1 and its homologous genes, OsBRL1 and OsBRL3, in rice. *Plant physiology*, *140*(2), 580-590.

Nam, K. H., & Li, J. (2002). BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell*, *110*(2), 203-212.

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*, *31*(13), 3812-3814.

Niu, L., Feng, S., Ding, W., & Li, G. (2016). Influence of speed and rainfall on large-scale wheat lodging from 2007 to 2014 in China. *PLoS One*, *11*(7), e0157677.

Niu, Y., Chen, T., Zhao, C., & Zhou, M. (2021). Improving crop lodging resistance by adjusting plant height and stem strength. *Agronomy*, *11*(12), 2421.

Noguchi, T., Fujioka, S., Choe, S., Takatsuto, S., Yoshida, S., Yuan, H., & Tax, F. E. (1999). Brassinosteroid-insensitive dwarf mutants of Arabidopsis accumulate brassinosteroids. *Plant physiology*, *121*(3), 743-752.

Okuno, A., Hirano, K., Asano, K., Takase, W., Masuda, R., Morinaka, Y., & Matsuoka, M. (2014). New approach to increasing rice lodging resistance and biomass yield through the use of high gibberellin producing varieties. *PloS one*, *9*(2), e86870.

Paff, K., & Asseng, S. (2018). A review of tef physiology for developing a tef crop model. *European journal of agronomy*, *94*, 54-66.

Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A., & Xiong, G. (2013). Hemicellulose biosynthesis. *Planta*, *238*, 627-642.

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*, *42*(1), 1-3.

Pękala, P., Szymańska-Chargot, M., & Zdunek, A. (2023). Interactions between non-cellulosic plant cell wall polysaccharides and cellulose emerging from adsorption studies. *Cellulose*, *30*(15), 9221-9239.

Peng, D., Chen, X., Yin, Y., Lu, K., Yang, W., Tang, Y., & Wang, Z. (2014). Lodging resistance of winter wheat (Triticum aestivum L.): Lignin accumulation and its related enzymes activities due to the application of paclobutrazol or gibberellin acid. *Field Crops Research*, *157*, 1-7.

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., & Harberd, N. P. (1999). 'Green revolution'genes encode mutant gibberellin response modulators. *Nature*, *400*(6741), 256-261.

Phokas, A., & Coates, J. C. (2021). Evolution of DELLA function and signaling in land plants. *Evolution & Development*, *23*(3), 137-154.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572-1574.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), 406-425.

Sashankar, P., Hegde, S. N., & Sathyanarayana, N. (2021). Gene Identification and Structure Annotation. *Bioinformatics in Rice Research: Theories and Techniques*, 163-177.

Sattler, S. E., Palmer, N. A., Saballos, A., Greene, A. M., Xin, Z., Sarath, G., & Pedersen, J. F. (2012). Identification and characterization of four missense mutations in brown midrib 12 (Bmr12), the caffeic O-methyltranferase (COMT) of sorghum. *BioEnergy Research*, *5*, 855-865.

Sattler, S. E., Saathoff, A. J., H aas, E. J., Palmer, N. A., Funnell-Harris, D. L., Sarath, G., & Pedersen, J. F. (2009). A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the sorghum brown midrib6 phenotype. *Plant physiology*, 1*50*(2), 584-595.

Scheible, W. R., Eshed, R., Richmond, T., Delmer, D., & Somerville, C. (2001). Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in Arabidopsis Ixr1 mutants. *Proceedings of the National Academy of Sciences*, *98*(18), 10079-10084.

Scheller, H. V., & Ulvskov, P. (2010). Hemicelluloses. *Annu al review of plant biology*, *61*, 263-289.

Shah, L., Yahya, M., Shah, S. M. A., Nadeem, M., Ali, A., Ali, A., & Ma, C. (2019). Improving lodging resistance: using wheat and rice as classical examples. *International journal of molecular sciences*, *20*(17), 4211.

Shen, W., & Li, Y. (2016). A novel algorithm for detecting multiple covariance and clustering of biological sequences. *Scientific reports*, *6*(1), 30425.

Silverstone, A. L., Jung, H. S., Dill, A., Kawaide, H., Kamiya, Y., & Sun, T. P. (2001). Repressing a repressor: gibberellin-induced rapid reduction of the RGA protein in Arabidopsis. *The Plant Cell*, *13*(7), 1555-1566.

Singh, A., Breja, P., Khurana, J. P., & Khurana, P. (2016). Wheat Brassinosteroid-Insensitive1 (TaBRI1) interacts with members of TaSERK gene family and cause early flowering and seed yield enhancement in Arabidopsis. *PLoS One*, *11*(6), e0153273.

Slabaugh, E., Scavuzzo-Duggan, T., Chaves, A., Wilson, L., Wilson, C., Davis, J. K., & Haigler, C. H. (2016). The valine and lysine residues in the conserved FxVTxK motif are important for the function of phylogenetically distant plant cellulose synthases. *Glycobiology*, *26*(5), 509-519.

Solovyev, V., Kosarev, P., Seledsov, I., & Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome biology*, *7*(1), 1-12.

Speicher, T. L., Li, P. Z., & Wallace, I. S. (2018). Phosphoregulation of the plant cellulose synthase complex and cellulose synthase-like proteins. *Plants*, *7*(3), 52.

Spielmeyer, W., Ellis, M. H., & Chandler, P. M. (2002). Semi-dwarf (SD-1),"green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proceedings of the National Academy of Sciences*, *99*(13), 9043-9048.

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics-Oxford*, *19*(2), 215-225.

Sun, C., Yan, K., Han, J. T., Tao, L., Lv, M. H., Shi, T., & Li, J. (2017). Scanning for new BRI1 mutations via tilling analysis. *Plant Physiology*, *174*(3), 1881-1896.

Tadele, Z., Ferede Haile, B., Abreha, E., Assefa, K., Chanyalew, S., & Mekbib, F. (2018). Morpho-Physiologic. *Genotype X Environment Interaction and In Vitro Evaluation for Drought Tolerance in Tef Eragrostis tef (Zucc.) Trotter, Ethiopia*.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, *38*(7), 3022-3027.

Taylor, N. G. (2008). Cellulose biosynthesis and deposition in higher plants. *New phytologist*, *178*(2), 239-252.

Tesema, A. (2013). Genetic resources of tef in Ethiopia. *Tef Improvement*, *15*.

Till, B. J., Cooper, J., Tai, T. H., Colowit, P., Greene, E. A., Henikoff, S., & Comai, L. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC plant biology*, *7*, 1-12.

Tobias, C. M., & Chow, E. K. (2005). Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification. *Planta*, *220*, 678-688.

Tong, H., Xiao, Y., Liu, D., Gao, S., Liu, L., Yin, Y., & Chu, C. (2014). Brassinosteroid regulates cell elongation by modulating gibberellin metabolism in rice. *The Plant Cell*, *26*(11), 4376-4393.

Tong, J. P., Liu, X. J., Zhang, S. Y., Li, S. Q., Peng, X. J., Yang, J., & Zhu, Y. G. (2007). Identification, genetic characterization, GA response and molecular mapping of Sdt97: a dominant mutant gene conferring semi-dwarfism in rice (Oryza sativa L.). *Genetics Research*, *89*(4), 221-230.

Torvalds, L. (1969). Linus Torvalds. *Free and Open Source Software*, 82. Demo.wegov.it

UniProt Consortium. (2007). The universal protein resource (UniProt). *Nucleic acids research*, *36*(suppl_1), D190-D195.

VanBuren, R., Man Wai, C., Wang, X., Pardo, J., Yocca, A. E., Wang, H., & Michael, T. P. (2020). Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal tef. *Nature communications*, *11*(1), 884.

Vandavasi, V. G., Putnam, D. K., Zhang, Q., Petridis, L., Heller, W. T., Nixon, B. T., & O'Neill, H. (2016). A structural study of CESA1 catalytic domain of Arabidopsis cellulose synthesis complex: evidence for CESA trimers. *Plant physiology*, *170*(1), 123-135.

Vandercasteelen, J., Dereje, M., Minten, B., & Taffesse, A. S. (2018). Perceptions, impacts, and rewards of row planting. *The Economics of Tef: Exploring Ethiopia's Biggest Cash Crop*. Amazon.com.

Vavilov, N. I. (1951). The origin, variation, immunity and breeding of cultivated plants. *72*(6), 482. LWW.

Verma, V., Worland, A. J., Savers, E. J., Fish, L., Caligari, P. D. S., & Snape, J. W. (2005). Identification and characterization of quantitative trait loci related to lodging resistance and associated traits in bread wheat. *Plant Breeding*, *124*(3), 234-241.

Vignols, F., Rigau, J., Torres, M. A., Capellades, M., & Puigdomènech, P. (1995). The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. *The Plant Cell*, *7*(4), 407-416.

Wang, M., Zhu, X., Wang, K. E., Lu, C., Luo, M., Shan, T., & Zhang, Z. (2018). A wheat caffeic acid 3-O-methyltransferase TaCOMT-3D positively contributes to both resistance to sharp eyespot disease and stem mechanical strength. *Scientific Reports*, *8*(1), 6543.

Wang, P., & Oliphant, T. (2012). Founding of Anaconda. Retrieved from https://www.anaconda.com/about-us.

Watson, L., & Dallwitz, M. J. (1992). *The grass genera of the world*. CAB international.

Wu, W., & Ma, B. L. (2018). Assessment of canola crop lodging under elevated temperatures for adaptation to climate change. *Agricultural and Forest Meteorology*, *248*, 329-338.

Xiong, W., Li, Y., Wu, Z., Ma, L., Liu, Y., Qin, L., & Fu, C. (2020). Characterization of two new brown midrib1 mutations from an EMS-mutagenic maize population for lignocellulosic biomass utilization. *Frontiers in Plant Science*, *11*, 594-798.

Xue, H., Gao, X., He, P., & Xiao, G. (2022). Origin, evolution, and molecular function of DELLA proteins in plants. *The Crop Journal*, *10*(2), 287-299.

Yamaguchi, S. (2008). Gibberellin metabolism and its regulation. *Annu. Rev. Plant Biol.*, *59*, 225-251.

Yamamuro, C., Ihara, Y., Wu, X., Noguchi, T., Fujioka, S., Takatsuto, S., & Matsuoka, M. (2000). Loss of function of a rice brassinosteroid insensitive1 homolog prevents internode elongation and bending of the lamina joint. *The Plant Cell*, *12*(9), 1591-1605.

Zhang, B., Deng, L., Qian, Q., Xiong, G., Zeng, D., Li, R., & Zhou, Y. (2009). A missense mutation in the transmembrane domain of CESA4 affects protein abundance in the plasma membrane and results in abnormal cell wall biosynthesis in rice. *Plant molecular biology*, *71*, 509-524.

Zheng, M., Chen, J., Shi, Y., Li, Y., Yin, Y., Yang, D., & Li, Y. (2017). Manipulation of lignin metabolism by plant densities and its relationship with lodging resistance in wheat. *Scientific Reports*, *7*(1), 41805.

Zhou, J. M., Lee, E., Kanapathy-Sinnaiaha, F., Park, Y., Kornblatt, J. A., Lim, Y., & Ibrahim, R. K. (2010). Structure-function relationships of wheat flavone O-methyltransferase: Homology modeling and site-directed mutagenesis. *BMC Plant Biology*, *10*, 1-12.

Zhou, J. M., Seo, Y. W., & Ibrahim, R. K. (2009). Biochemical characterization of a putative wheat caffeic acid O-methyltransferase. *Plant Physiology and Biochemistry*, *47*(4), 322-326.

Zhu, Q., Smith, S. M., Ayele, M., Yang, L., Jogi, A., Chaluvadi, S. R., & Bennetzen, J. L. (2012). High-throughput discovery of mutations in tef semi-dwarfing genes by next-generation sequencing analysis. *Genetics*, *192*(3), 819-829.

Annex I: Reverse complement of *in silico* probes [python code]

```
def dna_sequence_complement(dna):
        complements = {'A': 'T','T': 'A','G': 'C','C': 'G'}
        return ''.join([complements[c] for c in dna])
```

Annex II: *In silico* probe sub-sequences [on the Linux terminal]

```
# echo –e
     ">seq\nGGCACGGACCAGGTCATGTCCGAGGTGTACCTCGGCCGGCAGATCTGCAA
     CGT" | seqkit sliding -s 1 -W 15 > Rhtf1.txt
# echo -e
     ">seq\nACGTTGCAGATCTGCCGGCCGAGGTACACCTCGGACATGACCTGGTCCGT
     GCC" | seqkit sliding -s 1 -W 15 > Rhtr1.txt
# echo -e ">seq\nGAAATGGAGACCATTGGCAAGATCAAACACCG" | seqkit
     sliding -s 1 -W 15 > BRIf.txt
# echo -e ">seq\nCGGTGTTTGATCTTGCCAATGGTCTCCATTTC" | seqkit
     sliding -s 1 -W 15 > BRIr.txt
# echo -e ">seq\nGACGGCGGCATCCCGTTCAACAAGGCGTACGGGATG" |
     seqkit sliding -s 1 -W 15 > COMT1f.txt
# echo -e ">seq\nCATCCCGTACGCCTTGTTGAACGGGATGCCGCCGTC" |
     seqkit sliding -s 1 -W 15 > COMTr1.txt
# echo -e ">seq\nCAGATCGAGGTCGTCAAGATGGACTACGTCAACCAGGC" |
     seqkit sliding -s 1 -W 15 > CAD8Cf.txt
# echo -e ">seq\nGCCTGGTTGACGTAGTCCATCTTGACGACCTCGATCTG" |
     seqkit sliding -s 1 -W 15 > CAD8Cr.txt
# echo -e ">seq\nAATGAACAGTTCTGGGTCATTGG" | seqkit sliding -s
     1 -W 15 > CESA1f.txt
# echo -e ">seq\nCCAATGACCCAGAACTGTTCATT" | seqkit sliding -s
     1 -W 15 > CESAr1.txt
# echo -e ">seq\nGTGCTCTGGTCTGTCCTGCTCGCCTCC" | seqkit sliding
     -s 1 -W 15 > CESAf4.txt
# echo -e ">seq\nGGAGGCGAGCAGGACAGACCAGAGCAC" | seqkit sliding
     -s 1 -W 15 > CESA4r.txt
```

Annex III: Retrieval of tef subgenome sequences [on the Linux terminal]

```
Wget    https://www.ebi.ac.uk/ena/browser/api/fasta/Tef    sub-
     genome sequence accession numbers?download=true
```

[*Tef subgenome sequence accession numbers**: CM044754.1, CM044756.1, CM0447658.1, CM044760.1, CM044762.1, CM044764.1 CM044766.1 CM044768.1, CM044770.1, CM044772.1, CM044755.1, CM044757.1, CM044759.1, CM044761.1, CM044763.1, CM044765.1, CM044767.1, CM044769.1, CM044771.1, CM044773.1]

Tef subgenome accession numbers presented in the order 1A-10A and 1B-10B.

Annex IV: To extract the estimated entire tef homologous genes

`grep "both the forward and reverse complement in silico probe sub-squences" 1A-10A and 1B-10B --before-context=number --after-context=number >name of each gene.txt`

[where 'number' provides the length 5' or 3' of the *in silico* probes]