**UNIVERSITY OF PARMA**
Department of Life Sciences
Ph.D. in Biotechnologies
XXVIII Course

# Isolation, characterization and discovery of new alleles of *sHsp* genes in durum wheat (*Triticum durum* Desf.)

**Coordinator:**
Prof. Nelson Marmiroli

**Supervisor:**
Prof. Nelson Marmiroli
**Tutor:**
Dr. Michela Janni

**Ph.D. Candidate:**
Alessia Comastri

# Index

Abstract ..... 5
Riassunto ..... 6
List of abbreviations ..... 8
List of figures ..... 9
List of tables ..... 10
Chapter 1. General introduction ..... 12
1.1 The Wheat ..... 12
1.1.1 Importance of wheat ..... 12
1.1.2 The today's challenges of the wheat production ..... 12
1.1.3 Domestication evolution in wheat ..... 15
1.1.4 Wheat genetics and genomics ..... 17
1.1.5 Generalities and developmental stages ..... 19
1.2 Abiotic stress in plants ..... 21
1.2.1 Plant responses to abiotic stresses ..... 21
1.2.2 The Heat Stress Response (HSR) ..... 23
1.2.3 Transcription factors and stress responsive elements in abiotic stress tolerance ..... 25
1.2.4 Effect of heat stress on chloroplast and photosynthesis ..... 27
1.2.5 The Heat Shock Proteins (HSPs) ..... 28
1.2.6 The small Heat Shock Proteins (sHSPs) ..... 34
1.3 Functional genomics in crop improvement ..... 39
1.3.1 TILLING in functional genomics ..... 39
1.3.1.1 Creation of the mutagenized population ..... 41
1.3.1.2 Mutation detection methods ..... 43
1.3.2 TILLING *by sequencing* and *online* TILLING ..... 46
1.3.3 EcoTILLING ..... 48
1.3.4 SNP-based functional markers ..... 49
1.3.4.1 KASP genotyping applied in TILLING ..... 53
Objectives ..... 55
Chapter 2. Identification and characterization of the *T. durum* chloroplastic *sHsps* genes (*TdHsp26*) and promoters ..... 56
2.1 Introduction ..... 56
2.1.1 The role of sHSP26s in thermotolerance ..... 56
2.1.2 Known *sHsp26* in wheat ..... 56
2.1.3 Wheat *sHsp26* promoter ..... 57

2.2 Material and methods ..... 59
2.2.1 Vegetal material and growth conditions ..... 59
2.2.2 Database research for wheat *sHsp26* gene sequence identification ..... 59
2.2.3 *TdHsp26* genes isolation and chromosome localization ..... 59
2.2.3.1 Primers design ..... 59
2.2.3.2 Polymerase Chain Reaction (PCR) ..... 60
2.2.4 Nucleic acid extraction and analysis ..... 60
2.2.5 Bioinformatics ..... 61
2.3 Results and discussion ..... 62
2.3.1 Identification of wheat *Hsp26* genes ..... 62
2.3.2 Isolation and sequence analysis of *TdHsp26* genes ..... 63
2.3.3 *TdHsp26* chromosome localization ..... 67
2.3.4 Analysis of the predicted TdHSP26 proteins ..... 68
2.3.5 Promoter isolation and analysis ..... 73
2.4 Conclusion ..... 77
Chapter 3. Identification of novel *TdHsp26* alleles ..... 78
3.1 Material and methods ..... 78
3.1.1 Vegetal material and growth conditions ..... 78
3.1.2 Online TILLING analysis ..... 78
3.1.3 Conventional screening of TILLING population by High Resolution Melting (HRM) ..... 79
3.1.3.1 Allele-specific primer design ..... 79
3.1.3.2 HRM screening of the Cham1 TILLING population ..... 79
3.1.3.3 Mutation identification ..... 82
3.1.4 Bioinformatics ..... 83
3.1.5 SNPs assay by KASP molecular markers ..... 83
3.1.5.1 DNA extraction ..... 83
3.1.5.2 KASP markers development ..... 83
3.1.5.3 SNP assay ..... 86
3.1.6 Phenotypical and physiological evaluation of *TdHsp26s* mutants ..... 87
3.1.6.1 Phenotypical evaluation ..... 87
3.1.6.2 Physiological evaluation: Chlorophyll fluorescence ..... 87
3.2 Results and discussion ..... 88
3.2.1 Online SNPs detection ..... 88
3.2.2 Mutation detection by High Resolution Melting (HRM) ..... 94
3.2.3 SNPs validation using KASP markers ..... 98
3.2.4 Preliminary phenotypical and physiological evaluation of the mutant and wild type lines ..... 103
3.3 Conclusion ..... 107

Chapter 4. Work in progress: backcross of the most interesting *TdHsp26* mutant lines, double mutants production and $F_1$ genotyping by KASP markers ..... 108
4.1 Material and methods ..... 108
4.1.1 Crossing ..... 108
4.1.2 SNP assay by KASP markers ..... 109
4.1.2.1 Vegetal material and DNA extraction ..... 109
4.1.2.2 SNP assay ..... 109
4.2 Results and discussion ..... 110
4.2.1 Genotyping of the backcrossed and double mutant lines ..... 110
4.3 Conclusion ..... 114
Chapter 5. General discussion and conclusions ..... 115
Web Sites ..... 116
References ..... 117
Acknowledgments ..... 136
List of publications ..... 138

# Abstract

Plants show high adaptability to the environment also in stress conditions; study the molecular basis of adaptability in a crop plants as wheat is of particular interest due to its important role as staple food. Moreover climate changes are predicted to have severe negative effects on wheat production over the next decades.

Wheat is between the top five cultivated crops in the world and durum wheat (*Triticum durum* Desf.), in particular, is cultivated in more than 40 countries with a production of about 37 Mtons.

The exposure to heat stress during wheat developmental stages negatively affect yield and quality, therefore the current challenge of basic and applied research is to decipher the molecular mechanisms that take place during the adaptation to heat stress to produce new genetic variability to employ in wheat breeding programs.

Small heat shock proteins (sHSPs) are chaperone-like proteins that play an important role in the prevention of irreversible aggregation of misfolded protein. The chloroplast sHSPs are known to have an important role in the thermotolerance acquisition protecting the Photosystem II and thylakoid membranes during heat stress.

Three complete durum wheat gene sequence coding for chloroplast-localized sHSPs named *TdHsp26-A1, TdHsp26-A2, TdHsp26-B1* and a putative pseudogene sequence named *TdHsp26-A3,* members of the chloroplast sHSP sub-family, have been isolated and characterized. Novel allelic variations in the target genes were detected combining a conventional TILLING approach (HRM detection) with a new online TILLING database search. The search of the TILLING database, based on the exome captured sequencing of the durum wheat cv. Kronos TILLING population, leads to the detection of 36 mutant lines carrying missense mutations spanning the entire coding sequence of *TdHsp26-A1, TdHsp26-A2* and *TdHsp26-B1* genes. Moreover the conventional HRM screening of 960 genotypes of the durum wheat cv. Cham1 TILLING population leads to the detection of 3 missense mutations within a functional interest region of the *TdHsp26-B1*. The more interesting SNP mutations inducing non-synonymous substitutions in functional domains of the *TdHSP26-A1* and *TdHSP26-B1,* were confirmed *in vivo* through the development of specific codominant KASP molecular markers.

During this thesis new genetic variability in genes coding for sHSPs has been identified; this represent a valuable tool for elucidate gene function and to produce new durum wheat varieties with increased adaptation capabilities to heat stress.

# Riassunto

Gli organismi vegetali mostrano una notevole capacità di adattamento alle condizioni di stress e lo studio delle componenti molecolari alla base dell'adattamento in colture cerealicole di interesse alimentare, come il frumento, è di particolare interesse per lo studio di varietà che consentano una buona produzione con basso input anche in condizioni ambientali non ottimali. L'esposizione delle colture cerealicole a stress termico durante determinate fasi del ciclo vitale influisce negativamente sulla resa e sulla qualità, a questo fine è necessario chiarire le basi genetiche e molecolari della termotolleranza per identificare geni e alleli vantaggiosi da impiegare in programmi di incrocio volti al miglioramento genetico. Numerosi studi dimostrano il coinvolgimento delle sHSP a localizzazione cloroplastica (in frumento sHSP26) nel meccanismo di acquisizione della termotolleranza e la loro interazione con diverse componenti del fotosistema II (PSII) che determinerebbe un'azione protettiva in condizioni di stress termico e altri tipi di stress.

Lo scopo del progetto è quello di caratterizzare in frumento duro nuove varianti alleliche correlate alla tolleranza a stress termico mediate l'utilizzo del TILLING (Target Induced Local Lesion In Genome), un approccio di genetica inversa che prevede la mutagenesi e l'identificazione delle mutazioni indotte in siti di interesse.

Durante la tesi sono state isolate e caratterizzate 3 sequenze geniche complete per smallHsp26 denominate *TdHsp26-A1*; *TdHsp26-A2*; *TdHsp26-B1* e un putativo pseudogene denominato *TdHsp26-A3*. I geni isolati sono stati usati come target in analisi di TILLING in due popolazioni di frumento duro mutagenizzate con EMS (EtilMetanoSulfonato). Nel nostro studio sono stati impiegati due differenti approcci di TILLING: un approccio di TILLING classico mediante screening con High Resolution Melting (HRM) e un approccio innovativo che sfrutta un database di TILLING recentemente sviluppato.

La popolazione di mutanti cv. Kronos è stata analizzata per la presenza di mutazioni in tutti e tre i geni individuati mediante ricerca online nel database di TILLING, il quale sfrutta la tecnica dell'exome capture sulla popolazione di TILLING seguito da sequenziamento ad alta processività. Attraverso questa tecnica sono state individuate, nella popolazione mutagenizzata di frumento duro cv. Kronos, 36 linee recanti mutazioni missenso. Contemporaneamente lo screening con HRM, effettuato su 960 genotipi della libreria di TILLING di frumento duro cv. Cham1 ha consentito di individuare mutazioni in una regione di 211bp di interesse funzionale del gene *TdHsp26-B1*, tra le quali 3 linee mutanti recanti mutazioni missenso in omozigosi.

Alcune mutazioni missenso individuate sui due geni *TdHsp26-A1* e *TdHsp26-B1* sono state confermate *in vivo* nelle piante delle rispettive linee mutanti generando marcatori codominanti

KASP (Kompetitive Allele Specific PCR) con cui è stato possibile verificare anche il grado di zigosità di tali mutazioni.

Al fine di ridurre il numero di mutazioni non desiderate nelle linee risultate più interessanti, è stato eseguito il re-incrocio dei mutanti con i relativi parentali wild type ed inoltre sono stati generati alcuni doppi mutanti che consentiranno di comprendere meglio i meccanismi molecolari presieduti da questa classe genica. Gli individui F1 degli incroci sono stati poi genotipizzati con i medesimi marcatori KASP specifici per la mutazione di interesse per verificare la buona riuscita dell'incrocio. Questo approccio ha permesso di individuare ed implementare risorse genetiche utili ad intraprendere studi funzionali relativi al ruolo di smallHSP plastidiche implicate nella acquisizione di termotolleranza in frumento duro e di generare marcatori potenzialmente utili in futuri programmi di breeding.

# List of abbreviations

ACD= Alpha-Crystalline Domain

BLAST= Basic Local Alignment Search Tool

CODDLE= Codons Optimized to Discover Deleterious Lesions

CR= Consensus Region

CSS= Chinese Spring Survey Sequence

FAO= Food and Agriculture Organization

Fm= Maximum Fluorescence

FRET= Fluorescence Resonance Energy Transfer

Fv = Variable Fluorescence

HSF= Heat Shock Factor

HSE= Heat Shock Element

HSP= Heat Shock Protein

HRM= High Resolution Melting

KASP= Kompetitive Allele-Specific PCR

IAEA= International Atomic Energy Agency

IGC= International Grain Council

IPCC= International Panel of Climate Change

IWGSC= International Wheat Genome Sequencing Consortium

MrD= Methionin-rich Domain

MVD= Mutant Variety Database

NSPP= Number of Spike per Plant

NSPS= Number of Spikelet per Spike

PARSESNP= Project Aligned Related Sequences and Evaluate SNPs

PCR= Polymerase Chain Reaction

PEA= Plant Efficiency Analyzer)

PH= Plant Height

PSSM= Position Specific Scoring Matrix

sHSP= small Heat Shock Protein

SIFT= Sorting Intolerant from Tolerant

SNP= Single Nucleotide Polymorphisms

TILLING= Targeting Induced Local Lesion IN Genoms

*TdHsp*= *Triticum durum* HSP gene

TdHSP= *Triticum durum* HSP protein

URGI= Unité de Recherche Génomique

UTR= UnTranslated Region

WGS= Whole Genome Sequencing

# List of figures

Figure 1.1 Projection of change in world average surface temperature and precipitations.

Figure 1.2 Summary of projected changes in crop yields due to climate change over the 21st century.

Figure 1.3 The evolution of wheat from the prehistoric grasses to modern durum and bread wheat.

Figure 1.4 Spikes of wild and modern wheat.

Figure 1.5 Schematic diagram of the relationships between wheat genomes with polyploidization history.

Figure 1.6 The complexity of the plant response to abiotic stress.

Figure 1.7 Overview of signaling pathways and factors involved in thermotolerance.

Figure 1.8 Role and mechanism of HSPs and HSFs in the heat stress and acquired thermotolerance in plants.

Figure 1.9 Small heat shock protein domains.

Figure 1.10 Structural organization of a typical sHSP.

Figure 1.11 Model for sHSP chaperone activity.

Figure 1.12 The general strategy used to create a mutant population in self-fertilizing crops.

Figure 1.13 Mutation detection techniques for TILLING screening.

Figure 1.14 Production of EMS-mutagenized rice samples and exome capture analysis.

Figure 1.15 *In silico* TILLING.

Figure 1.16 KASP mechanism of action.

Figure 2.1 Alignment of the complete genomic sequences of *TdHsp*26-A1, *TdHsp*26-A2, *TdHsp*26-A3 and *TdHsp*26-B1.

Figure 2.2 Gene structure of *TdHsp26-A1, TdHsp26-A2, TdHsp26-A3, TdHsp26-B1*.

Figure 2.3 Chromosomal localization of *TdHsp26-A1, TdHsp26-A2, TdHsp26-A3, TdHsp26-B1*.

Figure 2.4 Alignment of the isolated sequence of Cham1 TdHSP26-A1, TdHSP26-A2 and TdHSP26-B1 and homology matrix.

Figure 2.5 Alignment of the sHSP26 of *T. aestivum* and *T. durum* and other plant species.

Figure 2.6 Phylogenetic tree and homology matrix of wheat HSP26 proteins.

Figure 2.7 Alignment of the *TdHsp26-A1* and *TdHsp26-B1* promoter sequences of Cham1 with the regulatory elements predicted by PLACE and PlantCARE software.

Figure 3.1 Experimental design used for the HRM screening of the Cham1 TILLING library.

Figure 3.2 Plots of the melt curves in the HRM analysis.
Figure 3.3 Example of KASP data plotted using KBiosciences Klustercaller software with ROX normalization.
Figure 3.4 Map of the missense mutations of the Kronos lines on the TdHSP26-A1 protein sequence.
Figure 3.5 Map of the missense mutations of the Kronos lines on the TdHSP26-A2 protein sequence.
Figure 3.6 Map of the missense mutations of the Kronos lines on the TdHSP26-B1 protein sequence.
Figure 3.7 Position of the primers used in the HRM screening on *TdHSP26-B1* gene and single peak obtained for the wild-type after the melt cycle.
Figure 3.8 HRM analysis results for detected mutations.
Figure 3.9 Map of the missense mutations detected by HRM screening of the Cham1 TILLING library on the TdHSP26-B1 protein sequence.
Figure 3.10 KASP markers performed on the DNA of four Kronos mutant lines detected by searching the TILLING database.
Figure 3.11 KASP markers performed on the DNA of Cham1 W4-1771, W4-0844 and W4-0181 mutant lines detected by HRM.
Figure 3.12 Histogram of the KASP assay results.
Figure 3.13 Measure of plant height (PH), number of spike per plant (NSPP) and number of spikelets per spike (NSPS) of the Kronos and Cham1 mutant and wild type plants.
Figure 3.14 Picture taken during the phenotype observation of the mutant and wild type plants.
Figure 3.15 Maximum photochemical efficiency measured as Fv/Fm ratio for the Kronos and Cham1 mutant lines and wild type.
Figure 4.1 General scheme of the backcrossing and of the crossing between two different mutant lines carrying the mutation on the A or B copy of the *TdHsp26* gene.
Figure 4.2 KASP genotyping for the $F_1$generated mutant lines.

# List of tables

Table 1.1 Crop growth stages for cereals. Adapted Zadok Decimal Code.
Table 1.2 sHSP subfamilies in angiosperm and correspondent *A.thaliana* sHSP.
Table 2.1 Survey of *Triticum sHsp26* cDNA and protein sequence.
Table 2.2 Description of some *cis*-acting elements found in the promoter of *TaHsp26* through PLACE database.

Table 2.3 Primer pairs used for the *TdHsp26* genes isolation.

Table 2.4 *In silico* research on EnsemblPlants database and assembly strategy to isolate *Hsp26* genes in durum wheat.

Table 2.5 *cis*-elements identified by PLACE and PlantCARE on *Prom_TdHsp26-A1* and *Prom_TdHsp26-B1*.

Table 3.1 Primer pairs used for the HRM screening of the Cham1 TILLING library.

Table 3.2 KASP assay primers designed for each mutant line.

Table 3.3 Kronos lines carrying mutation on the *TdHsp26-A1* gene retrieved in the wheat TILLING database

Table 3.4 Kronos lines carrying mutation on the *TdHsp26-A2* gene retrieved in the wheat TILLING database

Table 3.5 Kronos lines carrying mutation on the *TdHsp26-B1* gene retrieved in the wheat TILLING database

Table 3.6 List of the mutations identified by HRM in the Cham1 TILLING library.

Table 3.7 Mutant lines selected for *in vivo* validation.

Table 3.8 Results of the SNP assay performed by KASP markers on the mutant plants.

Table 4.1 Full list of the $F_1$ mutant lines generated through crossing for *TdHsp26-A1* and *TdHsp26-B1* genes and genotyping results.

# Chapter 1. General introduction

## 1.1 The Wheat

### 1.1.1 Importance of wheat

Wheat is one of the more important source of carbohydrates for humans (Curtis *et al.* 2002) feeding about 40% of the world population and providing, on average, one-fifth of the total food calories and protein in the human nutrition. Wheat consists mainly of two types: the hexaploid bread wheat (*Triticum aestivum* L.) accounting for about 95% of world wheat production, and the tetraploid durum wheat (*Triticum durum* Desf.) accounting for the other 5%. Of the 720 Mton of total annual wheat production (FAOSTAT website) approximately 35 Mton are represented by the durum wheat production (IGC report 2014). The latter is more adapted to the dry Mediterranean climate than bread wheat ad is often called pasta wheat to reflect its major end-use. The wheat directly consumed by humans is estimate about one-third and a large part is used to feed livestock. The current dominance of wheat in food and feed production lies in its adaptability and high yields together with the unique proprieties of dough that can be processed into a range of breads, other baked products, pasta and other processed food.

With the 60-70% content of starch and the 8-15% protein content wheat is an important resource for both human and livestock nutrition, it is also a significant source of minerals, mainly iron and zinc, and vitamins. The wholegrain wheat products contains a range of phytochemicals components with health benefits which are concentrated in the bran, as phenolic acids, tocols and sterols with antioxidant activities. Wheat cultivars contain also wide variation in soluble and insoluble fiber which has high healthy and dietary value. The genetically determined variation in wheat phytochemical and fiber contents is commonly exploited in molecular breeding for the production of varieties with increased nutritional benefits (Shewry 2009).

The economic importance of wheat has also triggered intense cytogenetic and genetic studies in the past decades aimed to develop wheat cultivars with enhanced biotic and abiotic stress tolerance together with increased yield and quality (Carver 2009).

### 1.1.2 The today's challenges of the wheat production

As assessed by the Intergovernmental Panel of Climate Change (IPCC) Fifth Assessment Report (IPCC 2014) all aspects of food security are potentially affected by climate change, including food production, access, use and price stability.

The RPCs (Representative Concentration Pathways (RCPs) based projection for the change in average surface temperature and average precipitation for 2081–2100 relative to 1986–2005 in the

best (RCP2.6) and worst (RCP8.5) greenhouse gas emission scenario are reported in Figure 1.1. Across all RCPs that are consistent with a wide range of possible changes in future anthropogenic greenhouse gas emissions, global mean temperature is projected to rise by 0.3 to 4.8 °C by the late-21st century (Figure 1.1a).

Changes in precipitation will not be uniform. The high latitudes and the equatorial Pacific are likely to experience an increase in annual mean precipitation under the RCP8.5 scenario, whereas in many mid-latitude and subtropical dry regions mean precipitation will likely decrease, with a 2-4% decreasing in the Mediterranean areas (Figure 1.1b).

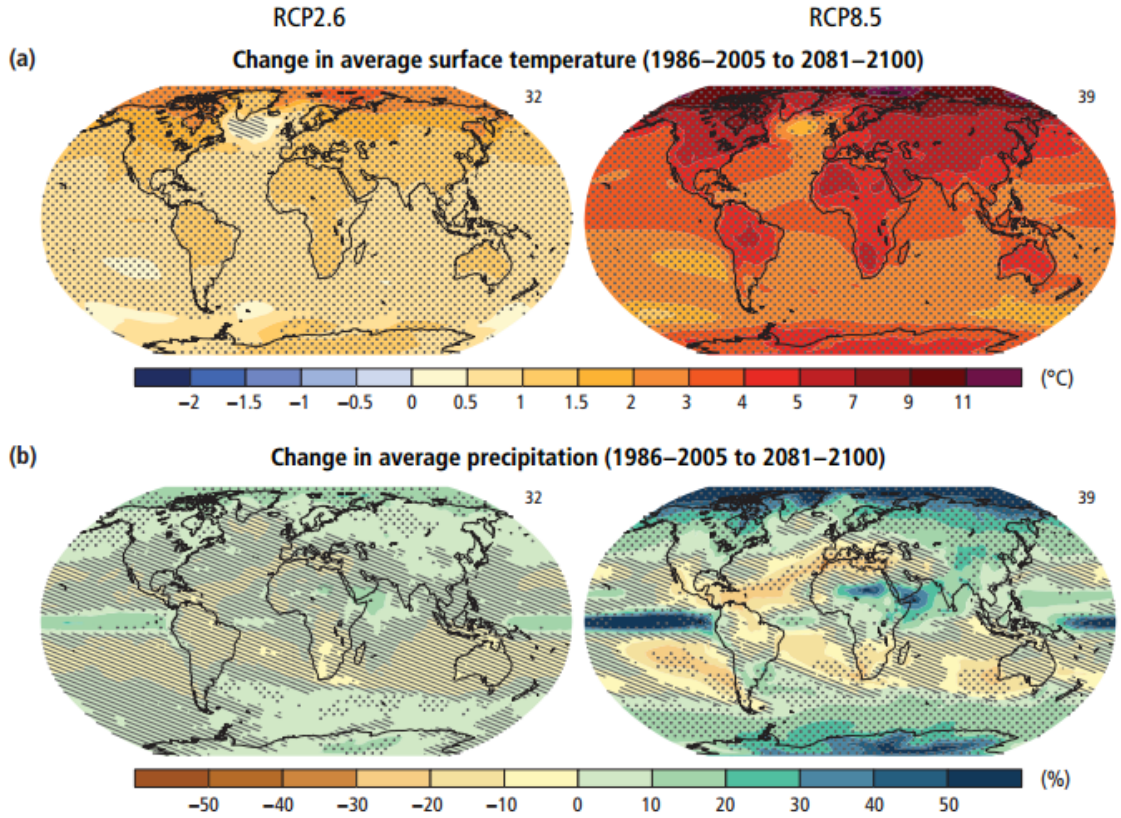



**Figure 1.1** Projection of change in world average surface temperature (a) and in average precipitation (b) based on multi-model mean projections for 2081–2100 relative to 1986–2005 under the RCP2.6 (left) and RCP8.5 (right) scenarios. The number of models used to calculate the multi-model mean is indicated in the upper right corner of each panel (From IPCC report 2014).

For wheat, rice and maize in tropical and temperate regions, climate change without adaptation is projected to negatively impact production. Durum wheat is traditionally grown under rain-fed conditions and in semiarid tropics, here, drought and high temperature stresses at the end of the growing season can reduce crop yield potential of about 50%, as they coincide with the grain filling period (Hasanuzzaman *et al.* 2013, Altenbach 2012, Dupont *et al.* 2006).

In the actual climate changes scenario wheat production in Europe, representing 29% of global wheat production will be affect by the frequency of days with high temperature together with the occurrence of drought, late spring frosts and severe winter frosts associated with inadequate snow cover. In addition, overly wet and cool weather enhances disease occurrence, contribute to lodging and complicates crop management increasing yield variability (Trnka *et al.* 2014).

Projected climate change impacts vary across crops and regions and adaptation scenarios, with about 10% of projections for the 2030–2049 period showing yield gains of more than 10%, and about 10% of projections showing yield losses of more than 25%, compared with the late 20th century (Figure 1.2).

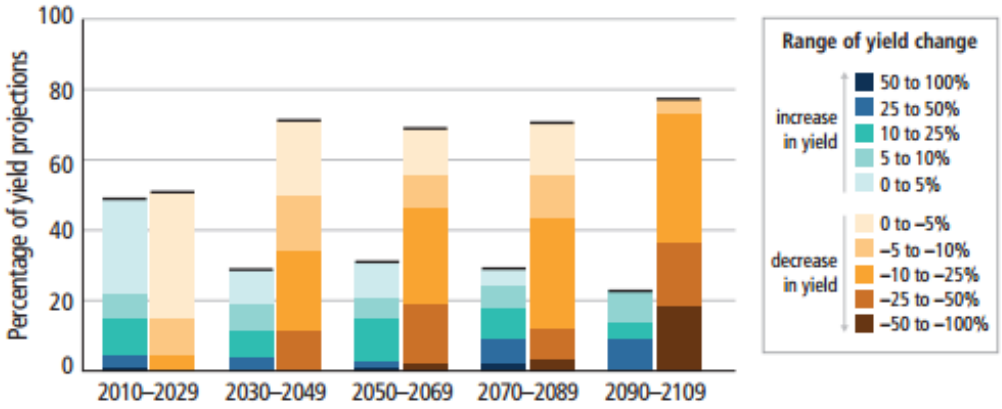



**Figure 1.2** Summary of projected changes in crop yields (mostly wheat, maize, rice and soy) due to climate change over the 21st century. The figure combines 1090 data points from crop model projections, covering different emission scenarios, tropical and temperate regions and adaptation and no-adaptation cases. The projections are sorted into the 20-year periods (horizontal axis). Changes in crop yields are relative to late 20th century levels and data for each time period sum to 100%. (From IPCC report 2014).

Moreover, to meet growing human needs, wheat grain production must increase of 2% each year (Foley *et al.* 2011), but no additional land will become available for this crop. To take on this challenge different strategies may be adopted: whereas for some regions it is important to breed cultivars that are capable of coping with an increased frequency and magnitude of heat stress around flowering, in some regions it will be equally important to maintain tolerance of low temperatures. In other regions, research should focus on water logging, lodging or field accessibility. (Trnka *et al.* 2014).

An urgent challenge of plant research is, therefore, to acquire new knowledge on the mechanisms guiding heat stress response and thermotolerance in crop plants.

## 1.1.3 Domestication evolution in wheat

The earliest cultivated forms of wheat were essentially landraces, selected by farmers from wild populations presumably because of their superior yield and other characteristics. Domestication is the results of a selection process that leads to an increased adaptation of plants to cultivation and use by humans (Brown 2010). Modern wheat cultivars usually refer to two species: hexaploid bread wheat, *T. aestivum* (2n = 6x = 42, $A^uA^uBBDD$), and tetraploid durum wheat *T. durum* (2n = 4x = 28, $A^uA^uBB$). Bread wheat is known to have three closely related subgenomes named A, B and D considered as homoeologous. The wild diploid *Triticum* species *Triticum urartu* (*Triticum urartu* Tumanian; 2n = 2x = 14, genome $A^uA^u$) has played an essential role in wheat evolution and contributed the $A^u$ genome to all tetraploid and hexaploid wheat.

*T. urartu* hybridized with the closest relative of goat grass *Aegilops speltoides* (2n = 2x = 14, genome SS), the B genome ancestor, 300,000–500,000 years before present (BP) to produce wild emmer wheat *Triticum dicoccoides* (*Triticum dicoccoides (Körn. ex Asch. & Graebner) Schweinf*; 2n = 4x = 28, genome $A^uA^uBB$) which was discovered in Israel. About 10,000 BP, hunter-gatherers began to cultivate wild emmer creating the domesticated form, cultivated emmer *Triticum dicoccum* (*Triticum dicoccum* Schrank ex Schübler; 2n = 4x = 28, genome $A^uA^uBB$). *T.dicoccum* spontaneously hybridized with another goat grass *Ae. Tauschii* (2n = 2x = 14, genome DD), the diploid D genome donor, around 9,000 BP to produce *T. spelta*, 2n = 6x = 42 (genome $A^uA^uBBDD$). About 8,500 BP, natural mutation changed the ears of both emmer and spelt to a more easily threshed type that later evolved into the free-threshing ears of durum wheat (*T. durum*) and bread wheat (*T. aestivum*) (Peng *et al.* 2011) (Figure 1.3). Whereas emmer clearly developed from the domestication of natural populations, bread wheat has only existed in cultivation, having arisen by hybridization of cultivated emmer with the unrelated wild grass *Ae. Tauschii* (Shewry 2009).

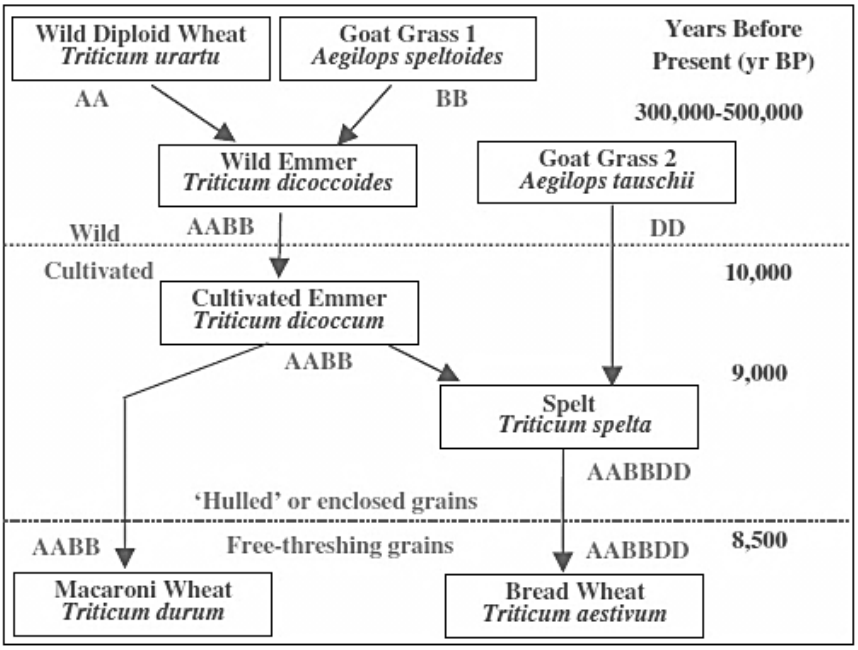



**Figure 1.3** The evolution of wheat from the prehistoric grasses to modern durum and bread wheat (From Peng *et al.* 2011, reproduced from http://www.newhallmill.org.uk/wht-evol.htm).

The domestication was associated with the selection of genetic traits that separated the cultivated wheat from their wild relatives and makes all the cultivars capable of surviving only under cultivation in human agricultural niches.

Two traits in particular were important modifications for the success of wheat domestication. The first is the loss of shattering of the spike at maturity which is clear important for ensuring seed dispersal in natural population, the non-shattering trait is determined by mutations at the Br (brittle rachis) locus (Nalam *et al.* 2006). The second is the change from hulled forms, in which the glumes adhere tightly to the grain, to free-threshing naked forms determined by a dominant mutation at the Q locus which modified the effects of recessive mutations at the Tg (tenacious glume) locus (Dubcovsky and Dvorak 2007, Simons *et al.* 2006, Jantasuriyarat *et al.* 2004).

All the cultivated forms of diploid, tetraploid, and hexaploid wheat, except from the spelt form of bread wheat have a tough rachis. The early domesticated forms of emmer and spelt are all hulled; whereas modern forms of tetraploid and hexaploid wheat are free-threshing (Figure 1.4) (Peng *et al.* 2011).

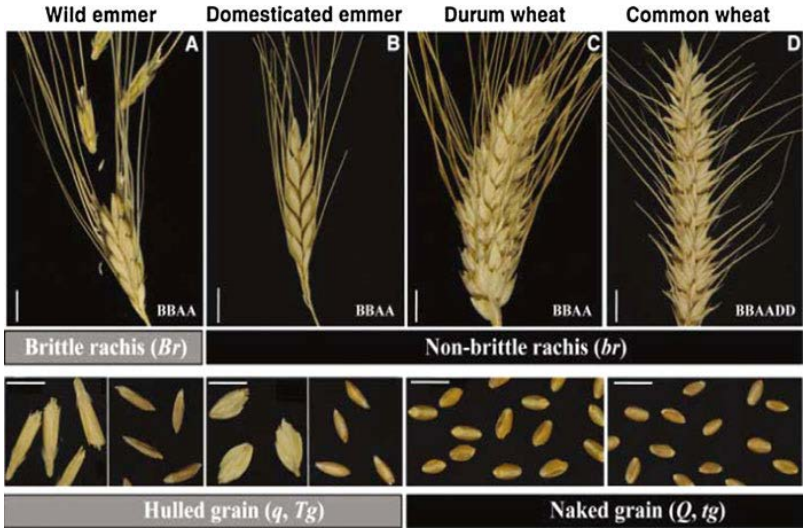



**Figure 1.4** Spikes of wild and modern wheat showing: brittle rachis (A), non-brittle rachis (B-D), hulled grain (A, B) and naked grain (C, D). (A) Wild emmer wheat (*T. dicoccoides*), (B) domesticated emmer (*T. dicoccum*), (C) durum (*T. durum*), (D) common wheat (*T. aestivum*). Gene symbols: Br= Brittle rachis, Tg= Tenacious glumes and Q= square head. (From Peng *et al.* 2011, reproduced from Dubcovsky and Dvorak 2007).

## 1.1.4 Wheat genetics and genomics

Bread wheat has a very complex genome composed of 21 pairs of chromosome that can be further divided into three homoelogous sets of seven chromosomes in each of the A, B and D sub-genomes (Figure 1.5, IWGSC 2014). The allopolyploid structure of the wheat genome is considered one of the key factors in the success of wheat as a global food crop. In polyploidy, the presence of two or three homoelogous gene copies may confers more plasticity that allows adaptation to changing environmental conditions (IWGSC, 2014). Each of the sub-genomes is about 5.5 Giga base pairs (Gb); the bread wheat genome has a size of 17 Gb (which is about five times the human genome size) with a hexaploid AABBDD genome structure and *T. durum* genome has a size of 12 Gb with a tetraploids AABB genome structure. Beside the big size of the genome, more than 90% of it is composed by repetitive DNA and is estimated that protein-coding sequence represents less than 2% of the genome (IWGSC 2014). Due to the big size, the high sequence similarity between homoeologous and the presence of repetitive sequence, establish a reference genome sequence for bread wheat is a big challenge.

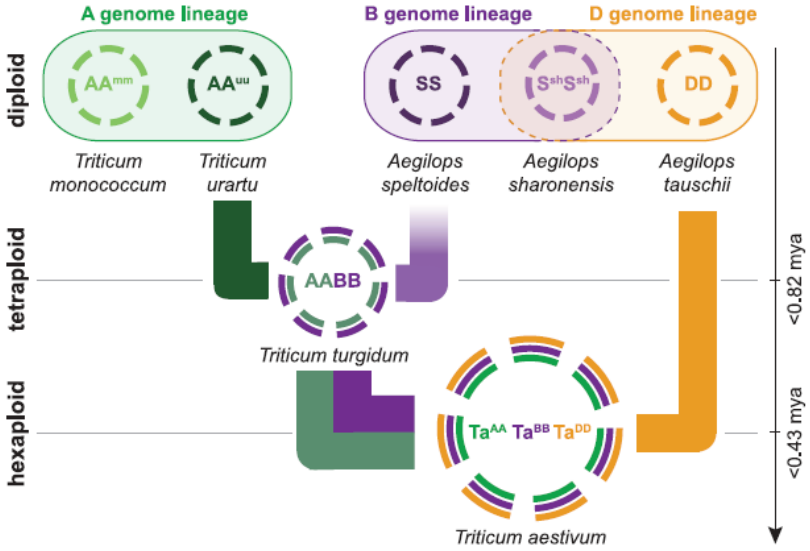



**Figure 1.5** Schematic diagram of the relationships between wheat genomes with polyploidization history. The circles provide a schematic representation of the chromosomal complement for each species. The allohexaploid genome structure of *Triticum aestivum* arose as a result of two polyploidization events (From IWGSC 2014).

The modern technologies of Next Generation Sequencing (NGS) through a combination of Whole Genome Shotgun (WGS) and chromosome-based approaches leads to the generation of a 10.2 Gb assembly of Chinese Spring named Chinese Spring Survey Sequence (CSS) (IWGSC 2014). This assembly allows to assign contigs (contiguous nucleotide sequence (DNA or RNA) obtained through direct sequencing, Borrill *et al.* 2015) as to specific chromosome arms giving the possibility to characterize homoeologous genes (Borrill *et al.* 2015). In parallel, separated assemblies of each of the three homoelogous genomes was conducted through a WGS approach in synthetic hexaploid wheat 'Synthetic W7984' (Chapman *et al.* 2015), using large-insert sequencing libraries. Both CSS and W7984 assemblies have been anchored to a high density genetic map using population sequencing (POPSEQ, Mascher *et al.* 2013). The genetic map was generated by sequencing with low coverage several individuals from a recombinant double haploid population (Synthetic W7984X Opata M85, Sorrells *et al.* 2011) and *in silico* SNPs-based mapping of the sequenced contigs. POPSEQ allows to anchors between 4.5 Gb and 7.1 Gb of the CSS and W7984 shotgun assemblies of the wheat genome to a genetic position (Borrill *et al.* 2015). An exception is the largest chromosome 3B that was successfully flow sorted as a single chromosome and has been fully sequenced (Choulet *et al.* 2014).

The CSS assembly contigs, ordered into chromosomal pseudomolecules through anchoring POPSEQ data, and combined with the reference sequence of chromosome 3B, have been incorporated in EnsemblPlants resource (Kersey *et al.* 2015).

## 1.1.5 Generalities and developmental stages

Wheat belongs to the Poaceae (Gramineae) family which includes major crop plants such as wheat (*Triticum* spp. L.), barley (*Hordeum vulgare* L.), oat (*Avena sativa* L.), rye (*Secale cereale* L.), maize (*Zea mays* L.) and rice (*Oryza sativa* L.). The Triticeae tribe, within the Pooideae subfamily, contains more than 15 genera and 300 species (Soreng *et al.* 2015).

Wheat is classified into spring or winter wheat traditionally referred to the season during which the crop is grown. Spring wheat is sown in the spring and matures in late summer, it is induced to flower by increasing day length only and it grows quickly so that mature grains can be harvested within 20-25 weeks. Winter wheat development, instead, is promoted by exposure of the seedlings to a period of cold (3-8 °C range); it is usually planted in the autumn to germinate and develop into young plants that remain in the vegetative phase during the winter and resume growth in early spring. Winter wheat has been selected to take advantage of a longer leafy stage when plant development is checked by the cooler temperatures and shorter day length of winter. During this time the extra leaf area can build up greater reserves which support higher grain yields when the crop matures (WheatBP website).

The growth cycle of wheat is divided in different phases: germination, seedling establishment and leaf production, tillering and head differentiation, stem and head growth, head emergence and flowering, and grain filling and maturity. Detailed background informations on all aspects of wheat are provided on the WheatBP (Wheat, the big picture) site from the Bristol Wheat Genomics site of the School of Biological Sciences, University of Bristol (WheatBP website).

Several systems have been developed to provide numerical designations for growth and developmental stages. Among these, the Feekes, Zadoks, and Haun scales are used the most frequently. The Zadoks Decimal Code (Zadoks *et al.* 1974) is internationally used to describe growth stages of cereals and is the most widely accepted system for describing wheat development because its stages are easy to identify in the field and it is more detailed than other systems allowing for precise staging. This system is based on two-digit code: the first digit refers to the principal stage of development beginning with germination (stage 0) and ending with kernel ripening (stage 9) and the second digit, instead, between 0 and 9 subdivides each principal growth stage. An adapted Zadoks Decimal Code is listed in Table 1.1 (Stapper 2007).

**Table 1.1** Crop growth stages for cereals. Adapted Zadok Decimal Code (From Stapper 2007). The precise stage of a wheat plant can be designed with the Zadoks decimal code as Zxx (also DCxx, Dxx or GSxx).

| **0 Emergence** |
|---|
| 00 Dry seed sown |
| 01 Seed absorbs water |
| 03 Germination, seed swollen |
| 05 Radicle emerged from seed |
| 07 Coleoptile emerged from seed |
| 09 Leaf at coleoptile tip |
| 10 First leaf through coleoptile and tip visible |
| **1 Seedling growth** |
| 11 1st leaf more then half visible |
| 12 2nd leaf more then half visible |
| 13 3rd leaf more then half visible |
| 14 4th leaf more then half visible |
| 19 6th leaf more then half visible |
| 17 7th leaf more then half visible |
| 18 8th leaf more then half visible |
| 19 9 or more leaves visible and stem not elongating |
| **2 Tillering** |
| 20 Main shoot only |
| 21 Main shoot and 1 tiller |
| 22 Main shoot and 2 tillers |
| 23 Main shoot and 3 tillers |
| 24 Main shoot and 4 tillers |
| 25 Main shoot and 5 tillers |
| 26 Main shoot and 6 tillers |
| 27 Main shoot and 7 tillers |
| 28 Main shoot and 8 tillers |
| 29 Main shoot and 9 or more tillers |
| **3 Stem elongation** |
| 30 stem starts to elongate, 'spike at 1cm' |
| 31 swelling 1st node detectable |
| 32 swelling 2nd node detectable |
| 33 swelling 3rd node detectable |
| 34 swelling 4th node detectable |
| 35 swelling 5th node detectable |
| 36 swelling 6th node detectable |
| **4 Flag leaf to Booting** |
| 37 Flag leaf tip visible |
| 38 Flag leaf half visible |
| 39 Flag leaf ligule just visible |
| 41 Early boot, flag sheath extending |
| 43 Mid-boot, boot opposite ligule of 2nd last leaf |
| 45 Full-boot, boot above ligule of 2nd last leaf |
| 47 Flag leaf sheath opening |
| 49 First awns visible |

| **5 Heading** |
|---|
| 51 10% of spikes visible (ear peep) |
| 52 |
| 53 30% of spikes visible |
| 54 |
| 55 50% of spikes visible |
| 56 |
| 57 70% of spikes visible |
| 58 |
| 59 90% of spikes visible |
| 60 Whole spike visible, no yellow anthers |
| **6 Flowering (anthesis)** |
| 61 Early– 20% of spikes with anthers |
| 62 |
| 63 30% of spikes with yellow anthers |
| 64 |
| 65 Mid– half of spikes with anthers |
| 66 |
| 67 70% of spikes with anthers |
| 68 |
| 69 Late– 90% of spikes with anthers |
| **7 Kernel and Milk development** |
| 70.2 Kernels middle spike extended 20% |
| 70.5 Kernels middle spike half formed |
| 70.8 Kernels middle spike extended 80% |
| 71 Watery ripe, clear liquid |
| 73 Early milk, liquid off-white |
| 75 Medium milk, contents milky liquid |
| 77 Late milk, more solids in milk |
| 79 Very-late milk, half solids in milk |
| **8 Dough development** |
| 81-85 spikes turn colour from light-green to yellow-green to yellow |
| 81 Very early dough, more solids and slides when crushed |
| 83 Early dough, soft, elastic and almost dry, shiny |
| 85 Soft dough, firm, crumbles but fingernail impression not held |
| 87 Hard dough, fingernail impression held, spike yellow-brown |
| 89 Late hard-dough, difficult to dent |
| **9 Ripening** |
| 91 Kernels hard, difficult to divide by thumb-nail |
| 92 Harvest ripe, kernels can no longer be divided by thumb-nail and straw still firm |
| 93 Kernels loosening in daytime |
| 94 Over-ripe, straw brittle |

# 1.2 Abiotic stress in plants

## 1.2.1 Plant responses to abiotic stresses

Abiotic stress is the primary cause of crop loss worldwide, reducing average yields for most major crop plants by more than 50% leading to a series of morphological, physiological, biochemical and molecular changes that adversely affect plant growth and productivity (Wang *et al.* 2003). Primary stresses, such as drought, salinity, cold, heat and chemical pollution are often interconnected, and cause cellular damage and secondary stresses, such as osmotic and oxidative stress. Oxidative stress frequently accompanies high temperature, salinity, or drought stress, and may cause denaturation of functional and structural proteins through the activation of similar cell signalling pathways and cellular responses, such as the production of stress proteins, up-regulation of anti-oxidants and accumulation of compatible solutes (Vierling and Kimpel 1992, Wang *et al.* 2003).

Plant response to abiotic stresses involves changes at different levels: from whole-plant organs and tissues to the cellular, physiological and molecular levels leading the plant to survive under unfavourable environmental conditions. The physiological and biochemical modifications include: leaf wilting, leaf area reduction, leaf abscission, root growth stimulation, alterations in relative water content, electrolytic leakage, production of reactive oxygen species (ROS) and accumulation of free radicals which disturb cellular homeostasis ensuing lipid peroxidation, membrane damage, and inactivation of enzymes thus influencing cell viability. Other than these, abscisic acid (ABA), a plant stress hormone, induces leaf stomata closure reducing transpiration, water loss and photosynthetic rate. Molecular responses to abiotic stresses include perception, signal transduction, gene expression and metabolic changes in the plant providing stress tolerance. Several genes, named stress-related genes, are activated in response to abiotic stresses at the transcriptional level, and their products provide stress tolerance by the production of vital metabolic proteins and also in regulating the downstream genes (Lata *et al.* 2011).

The stress-related genes can be classified in three major categories (Wang *et al.* 2003):

1) The genes encoding for product that function directly in the protection of macromolecules and membranes such as gene encoding for: heat shock proteins (HSPs) and other chaperones, late embryogenesis abundant (LEA) proteins, enzymes for biosynthesis of osmoprotectants as proline, betaine and sugars; detoxification enzymes and free-radical scavengers such as glutathione S-transferase, catalase, superoxide dismutase, ascorbate peroxidase; fatty acid metabolism enzymes, proteinase inhibitors, ferritin and lipid-transfer proteins.
2) The genes related to the passive and active transport system through the membrane, water and ion uptake.

3) The genes encoding for proteins that further regulate the stress signal transduction and alter gene expression such as: transcription factors (TFs), including Heat Shock Factors (HSFs), C-repeat binding factors (CBF)/dehydration-responsive transcription factors (DREB) and ABA-responsive element-binding protein (AREB)/ ABA-binding factor (ABF), MYC/MYB, protein kinases (such as mitogen activated protein kinase (MAPK), calcium dependent protein kinase CDPK, receptor protein kinase), protein phosphatases, proteinases and phospholipases.

The complex mechanism of plant response to abiotic stresses is schematically represented in Figure 1.6.

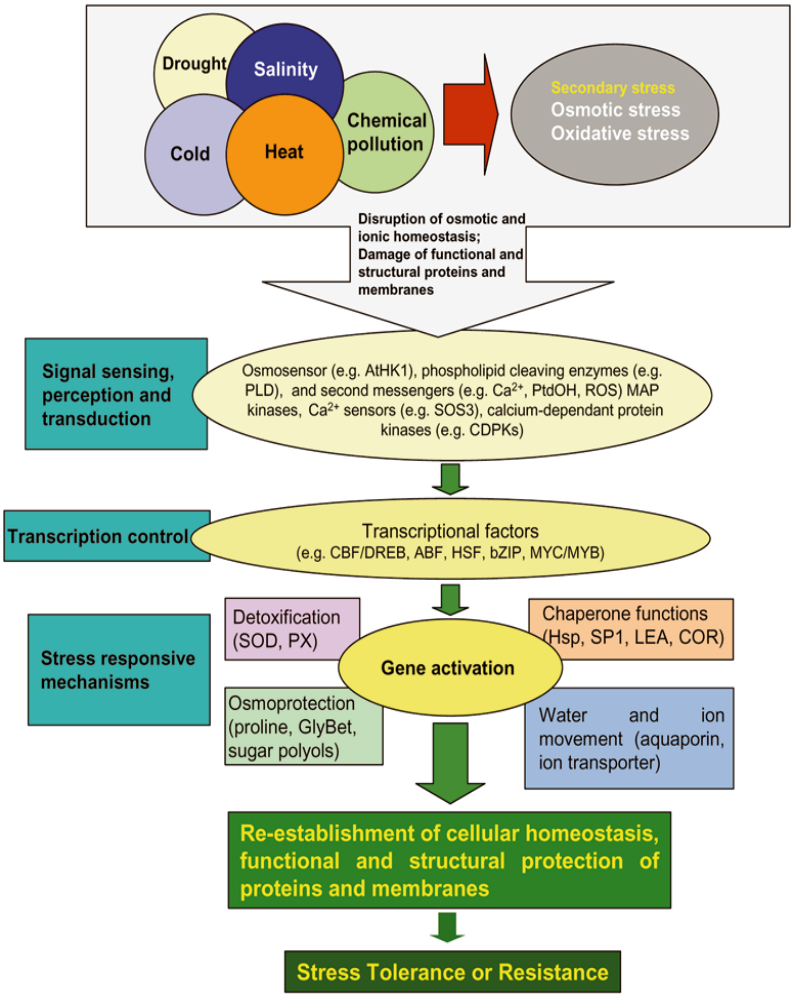



**Figure 1.6** The complexity of the plant response to abiotic stresses. Primary stresses, such as drought, salinity, cold, heat and chemical pollution are often interconnected, and cause cellular damage and secondary stresses, such as osmotic and oxidative stress. The initial stress signals (e.g. osmotic and ionic effects, or temperature, membrane fluidity changes) trigger the downstream signaling process and transcription controls which activate stress-responsive mechanisms to re-establish homeostasis and protect and repair damaged proteins and membranes. Inadequate response at one or several steps in the signaling and gene activation may ultimately result in irreversible changes of cellular homeostasis and in the destruction of functional and structural proteins and membranes, leading to cell death (From Wang *et al.* 2003).

The strategy used to enhance stress tolerance in plants is mainly based on the manipulation of genes that protect and maintain the function and structure of cellular components. Engineering strategies rely on the transfer of one or several genes involved in signalling and regulatory pathways, mostly TFs. For example different CBF/DREB in *A. thaliana* (Haake *et al.* 2002, Gilmour *et al.* 2000, Kasuga *et al.* 1999) and *Brassica nabus* (Jaglo *et al.* 2001), ABF3 or ABF4 in *A.thaliana* (Kang *et al.* 2002), AtMYC2 and AtMYB2 (Abe *et al.* 2003), HsfA1 in *Solanum lycopersicum* (Mishra *et al.* 2002); genes encoding stress-tolerance-conferring proteins, such as *Hsp21* in *A. thaliana* (Härndahl *et al.* 1999), or genes encoding enzymes osmolytes and antioxidants, for example GST and GPX in *N. tabacum* (Roxas *et al.* 1997). However, the nature of the genetically complex mechanisms of abiotic stress tolerance, and the potential detrimental side effects, make the improvement through gene transformation extremely difficult (Wang *et al.* 2003).

## 1.2.2 The Heat Stress Response (HSR)

Plants have an optimal temperature range for growth and development; in general they are able to tolerate temperature 5-10 °C above the optimum without being stressed, but at temperature 12-15 °C higher than the optimum plants suffer from heat stress and growth and development may be seriously affected (Leone *et al.* 2003). Heat stress (HS) is a complex function of intensity (temperature in degrees), duration and rate of temperature increasing (Wahid *et al.* 2007, Leone *et al.* 2003). The extent of damage caused by heat stress depends on the stage at which the plant is exposed to it; in wheat the most critical stage for heat stress has been observed to be pollination and milky dough kernel stages (Z73-Z89, Table 1.1) (Kumar and Rai 2014).

Heat shock events provoke irreversible damage during both vegetative and reproductive stages, causing low photosynthetic activity, poor floral development, pollen sterility, which affect seed and fruit set and quality. High temperature impairs many physiological activities associated with seedling growth and vigour, root growth, nutrient uptake, water relation, solute transport, photosynthesis, respiration, general metabolism, fertilization; moreover it induces ultrastuctural changes of nucleus, endoplasmic reticulum, mitochondria and plastid. The rate of photosynthesis in most species declines at about 35 °C which is ascribed to protein denaturation, loss of membrane integrity, photoinhibition and ion imbalance. High temperature affects chloroplast biogenesis and senescence, causes disintegration of chloroplast grana, disruption of membrane proteins, influences protein-lipid interactions, affects electron transport activity and decreases the activity of Rubisco (Ribulose-1,5-bisphosphate carboxylase/oxygenase). The combination of these cellular changes induced by heat and the inability to restore normal cellular function leads to lethality. To alleviate these effects, plants activate a series of physiological and biochemical changes necessary to re-

establish a new cellular homeostasis compatible with the increase in temperature. The adaptive response of plants to high temperature stress is a typical polygenic trait, controlled by a network of genes (Leone *et al.* 2003).

Heat tolerance is generally defined as the ability of the plant to grow and produce economic yield under high temperatures (Wahid *et al*. 2007). Tolerance to high temperature has two components: the basal or constitutive thermotolerance, due to evolutionary change in the plant for thermal adaptation under the changing habitat; and the acquired thermotolerance due to acclimation, that is the ability to survive lethal temperatures following the exposure to a mild heat stress (Maestri *et al.* 2002). The acquisition of thermotolerance is an autonomous cellular phenomenon resulting from a prior exposure to a conditioning pre-treatment and protects cells and organisms from a subsequent lethal stress.

The heat stress response (HSR) is defined as a transient reprogramming of gene expression and is a conserved biological reaction of cells and organisms to elevated temperature. The accumulation of HSPs under the control of HSFs is assumed to play a central role in the HSR and in acquired thermotolerance in plants (Kotak *et al.* 2007).

Multiple signalling pathways are implicated in the HSR, some of which control HSPs whereas others control the production or activation of diverse effector components (Figure 1.7). Emerging evidence highlights the cross-talk between heat and oxidative stress signalling: the increase of reported to occur after very short high temperature exposure have been correlated with the induction of heat stress genes through the direct sensing of $H_2O_2$ by HSFs (Volkov *et al.* 2006).

Phytohormones, such as abscisic acid (ABA), salicylic acid (SA) and ethylene, have also been linked to HS signalling in different plant species (Gururani *et al.* 2015). The initial demonstration that HSP and ABA, ROS and SA pathways are involved in the development and maintenance of acquired thermotolerance comes mainly from studies on *Arabidopsis* mutants (Larkindale *et al.* 2005).

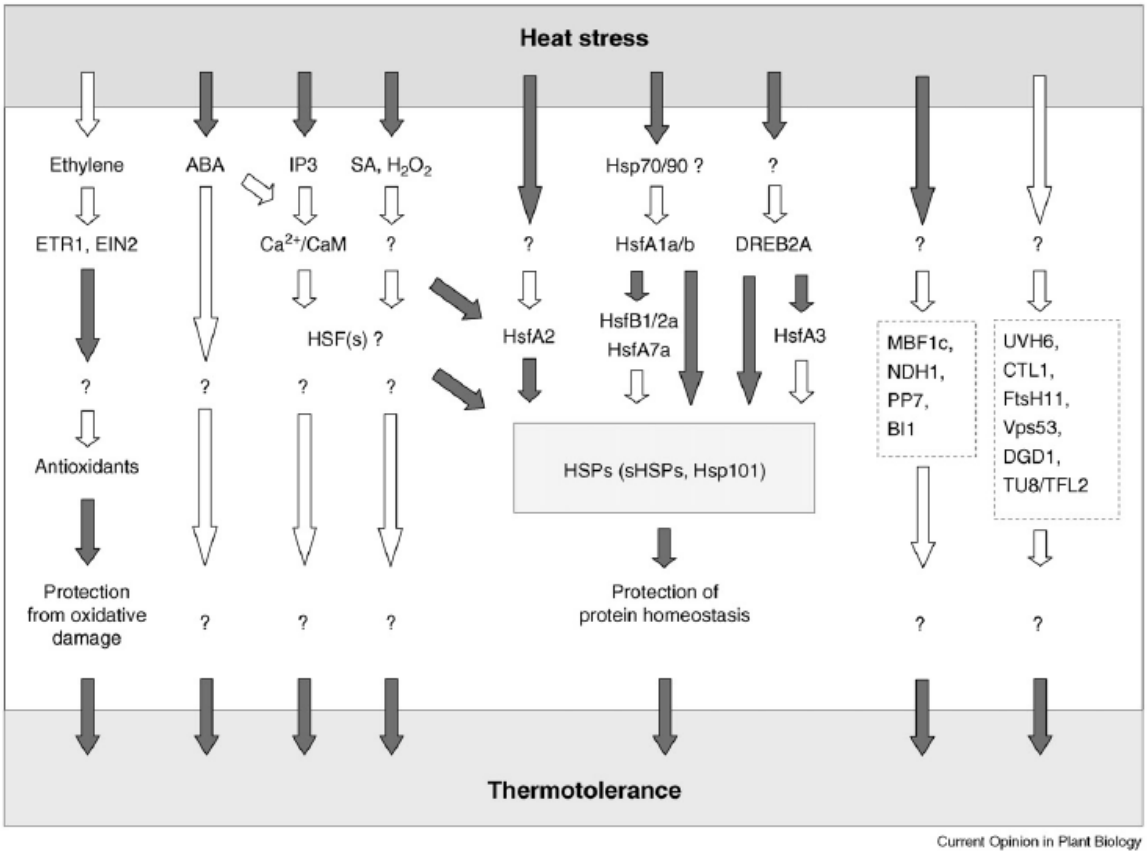



**Figure 1.7** Overview of signalling pathways and factors involved in thermotolerance. Signalling components involved in the heat stress response and protective factors leading to thermotolerance are shown. The yet unidentified factors in the different signal transduction pathways are indicate as question mark. Boxes that have dotted lines represent a collection of gene products that are known to affect thermotolerance, but whose particular functions in the network are still unknown. The most-characterized part of the network contains heat stress transcription factors (HSFs) that regulate genes encoding heat stress proteins (HSPs), which act as molecular chaperones. Dark arrows indicate connections with experimental evidence, white arrows mark hypothesized connections (From Kotak *et al.* 2007).

## 1.2.3 Transcription factors and stress responsive elements in abiotic stress tolerance

Transcription factors (TFs) account for for approximately 7% of the plant genome and act together with other transcriptional regulators, including chromatin remodelling/modifying proteins, to recruit or obstruct RNA polymerases on the promoters for the regulation of gene transcription (Agarwal and Jha 2010). TFs have a classical structure constituted by a DNA-binding domain (DBD) and an activation domain; through the DBD the TFs interact with *cis*-elements in the promoter regions of several stress-related and thus up-regulate the expression of many downstream genes leading to abiotic stress tolerance (Udvardi *et al.* 2007). Most of the TFs belong to a few large multigene families, e.g. MYB (myeloblastosis oncogene), AP2 (APETALA2)/EREBP (ethylene-responsive element binding proteins), bZIP (Basic Leucine Zipper) and WRKY (containing the typical

"WRKY signature" in the DBD). Individual members of the same family often respond differently to various stress stimuli; on the other hand, some stress responsive genes may share the same TFs, as indicated by the significant overlap of the gene-expression profiles that are induced in response to different stresses (Wang *et al.* 2003).

Stress tolerance or susceptibility are controlled at the transcriptional level by an extremely intricate gene regulatory network involving several pathways that independently respond to environmental stresses in ABA-dependent and ABA-independent manner (Lata *et al.* 2011).

ABA signalling plays a vital role in plant stress responses, in fact many of the drought-inducible genes studied to date are also induced by ABA. Two TF families bZIP and MYB, are involved in ABA signalling and its gene activation (Lata *et al.* 2011). Many ABA-inducible genes share the presence of a *cis*-acting ABA-responsive element (ABRE) in their promoter regions. ABA-dependent signalling systems, therefore, mediate the stress adaptation by inducing genes controlled by the AREB/ABF (ABA-responsive element-binding protein/ ABA-binding factor) TFs and the MYC (Myelocytomatosis-related oncogene)/MYB (Myeloblastosis-related oncogene) TFs. Constitutive expression in *Arabidopsis* plants of the ABRE binding factors ABF3 or ABF4 enhanced drought tolerance and leads to other ABA-associated phenotypes caused by an altered expression of ABA/stress-responsive genes (Kang *et al.* 2002). Moreover, *Arabisopsis* transgenic plants 35S:AtMYC2/AtMYB2 showed improved osmotic-stress tolerance as indicated by the electrolyte-leakage test (Abe *et al.* 2003). Between the most known TF, WRKY have been demonstrated to be key regulators of many processes in plants including the responses to biotic and abiotic stresses, senescence, seed dormancy and seed germination and some developmental processes. The WRKY transcription factors bind the pathogen related W-box element and play a crucial role in plant pathogen interaction (Rushton *et al.* 2010).

Similarly to other stresses, the HSR is primarily regulated at the transcriptional level and is mainly attributed to conserved *cis*-regulatory promoter elements HSEs (Heat Shock Elements) that are the binding sites for the *trans*-active heat-shock factors (HSFs; Schöffl *et al.* 1998) and are present in the promoters of *Hsp* genes (Morimoto 1998). HSF serves as the terminal component of signal transduction and mediates the expression of *Hsp* through the binding on the HSEs. The HSEs are composed of at least three contiguous repeats of nGAAn units and share a common consensus sequence nGAAnnTTCnnGAAn. The HSEs are divided into three types: perfect-type (nGAAnnTTCnnGAAn), gap-type (nGAAnnTTCn(5bp)nGAAn and step-type (nTTCn(5bp)nGAAn(5bp)nTTCn) (Chauhan *et al.* 2012). The overexpression of HSF leads to the expression of several *Hsp* genes conferring thermotolerance in *Arabidopsis* transgenic plants (Prändl *et al.* 1998, Lee *et al.* 1995). A total of 19 and 21 HSFs members have been cloned in rice and *Arabidopsis* respectively. On the basis of their structural domains plant HSFs proteins comprise

three conserved evolutionary classes named A, B and C. *Hsf*A genes has been relatively well studied, HSFA1a is known to be a master regulator and HSFA2 is a major heat stress factor (Qu *et al.* 2013, Kotak *et al.* 2007). In tomato plants, the overexpression of *HsfA1* resulted in heat-stress tolerance, while as expected *HsfA1* antisense plants were extremely sensitive to elevated temperatures (Mishra *et al.* 2002). HSE have been confirmed to be important also for the heat shock induction of the L-ascorbate peroxidase 1 (*apx1*) gene (Storozhenko *et al.* 1998). *In vitro* analysis of the interaction between tomato HSFs and the *apx1* promoter confirmed that HSE represents a functional HSF binding site, indicating that HSE is responsible for *apx1* heat shock induction and also contributing partially to oxidative stress induction (Khurana *et al.* 2013).

In wheat 56 *TaHsf* members, classified into A, B, and C classes were identified (Xue *et al.* 2014). any *TaHsfs* were constitutively expressed. Subclass A6 members were predominantly expressed in the endosperm under non-stress conditions. Upon heat stress, the transcript levels of A2 and A6 members became the dominant Hsfs, suggesting an important regulatory role during heat stress. Many *TaHsfA* members as well as B1, C1, and C2 members were also up-regulated during drought and salt stresses (Xu *et al.* 2014).

Another regulatory element important in the abiotic stress response is the CCAAT box which is one of the most common regulatory elements present in 30% of eukaryotic promoters and is conserved in promoters of the heat-shock genes (Bucher 1990). The CCAAT box has been found to act co-operatively with the HSEs to enhance HSP promoter activity during heat stress through maintains the promoter in open chromatin configuration so that HSF could rapidly activate after thermal stress (Landsberger and Wolffe 1995).

## 1.2.4 Effect of heat stress on chloroplast and photosynthesis

The photosynthetic activity of chloroplasts is considered among the most heat sensitive cell functions (Leone *et al.* 2003). The thylakoids, composing the chloroplast grana, are the structural units of photosynthesis. Photochemical reactions in thylakoid lamellae and carbon metabolism in the stroma of chloroplast have been suggested as the primary sites of injury at high temperatures (Wise *et al.* 2004). The negative effects on the photosynthetic apparatus are associated with the production of reactive oxygen species (ROS) and the PhotosystemII (PSII) located on the thylakoid membranes is known to be the most heat sensitive component of photosynthesis. *In vivo*, the extent of damage under any type of stress depends on the balance between damage and repair processes during the stress; this is particularly true for PSII and the various components involved in the PSII protections provides the basis for acclimation and photosynthetic recovery processes (Rezaei *et al.* 2015).

The main target processes of the high temperature induced damage are mainly the $CO_2$ fixation system, photophosphorylation, the electron transport chain and the oxygen evolving complex (OEC). The suppression of carbon fixation by (Rubisco) it's due to the inactivation of the Rubisco activase enzyme. The reduction in carbon fixation and oxygen evolution activities is followed by disruption of the linear electron flow, which induces the generation of ROS. ROS inhibits the repair of damaged PSII by suppressing the *de novo* synthesis of proteins of the photosynthetic machinery (Allakhverdiev *et al.* 2008). Strong heat stress may leads to the dissociation of the OEC in PSII resulting in an imbalance between the electron flow from OEC toward the acceptor side of PSII in the direction of the PSI reaction center; whereas moderate heat stress, which does not directly inactivate the OEC, stimulates photoinhibition of PSII. The photoinibition is due to an imbalance between the rate of photodamage to PSII and the rate of the repair of damaged PSII. Murata and co-workers proposed that photodamage occurs via a two-step process where the first step is the light-dependent destruction of the Mn cluster of the oxygen-evolving complex and the second step is the inactivation of the photochemical reaction center of PSII by light that has been absorbed by chlorophyll (Murata *et al.* 2007).

Alterations in various parameters associated with photosynthesis under heat stress are good indicators of the plant thermotolerance. Parameters like chlorophyll fluorescence, base fluorescence, photon flux density, transpiration rate, stomata aperture have been used to characterize different wheat cultivars for their thermotolerance capacity. Contents of photosynthetic pigments as chlorophylls and carotenoids and the ratios of Chl a/b are also good indicators of stress detection and tolerance in plants (Kumar and Rai 2014, Wahid *et al.* 2007).

## 1.2.5 The Heat Shock Proteins (HSPs)

Under heat stress plants accelerate the synthesis and the accumulation of heat-related proteins and in particular of heat shock protein (HSPs) which act as molecular chaperons and play an essential role in maintaining cellular homeostasis assisting the correct folding of nascent and stress accumulated misfolded or denatured proteins (Tyedmers *et al.* 2010, Wang *et al.* 2004).

Synthesis and accumulation of HSPs during heat stress has been reported for several plant species, including wheat, and has been linked to thermotolerance (Pandey *et al.* 2015, Grigorova *et al.* 2011, Sarkar *et al.* 2009, Rampino *et al.* 2009a, Rampino *et al.* 2009b, Gullì *et al.* 2007, Maestri *et al.* 2002, Marmiroli *et al.* 1994, Marmiroli *et al.* 1989a, Marmiroli *et al.* 1989b). HSPs comprise five conserved protein families HSP100s, HSP90s, HSP70s, HSP60s and the small HSPs (sHSPs); each HSP family has a unique mechanism of action with chaperone activity (Al-Whaibi 2011). In stress condition HSPs are essential for the proteins and membranes stabilization but are also responsible

for protein folding, assembly, translocation and degradation in normal cellular processes indicating their essential role for normal growth and development in plants. HSPs are ubiquitous in the cell; they are located in cytoplasm and organelles such as nucleus, mitochondria, chloroplasts and endoplasmic reticulum (ER). Different class of HSPs binds a specific target substrate in a non-covalent manner and do not form part of the final product. Despite their close working relationship in the cell, the individual HSP families are evolutionary unrelated to each other (Waters 2013).

HSFs and HSPs play a central role in the heat stress and in the acquisition of thermotolerance in plants, the general mechanisms is reported in Figure 1.8. HSPs primarily assist the folding, intracellular distribution, assembly and degradation of proteins through the stabilization of partially unfolded proteins. These proteins do not contain specific information for correct folding but collaborate with other components as protein disulphide isomerase that act as direct folding catalysts (Qu *et al.* 2013). The activation of HS genes transcription rely on the accumulation of denaturated proteins occurring under heat stress and is regulated through a feedback control involving the HSP70 and the HSFs as shown in Figure 1.8. In no stress conditions the interaction between HSP70 and HSFs prevents the trimerization and binding of HSF to HSE thereby blocking the transcriptional activation of heat-shock genes by their HSFs. When the content of the inactive proteins exceeds the threshold value during heat stress HSFA1a triggers the heat stress response through the induction of *HsfA1b* and *HsfA2* expression, which also forms co-activators with both proteins. HSFA2 induces the expression of various *Hsp* genes. HSP70, HSP101 and sHSP participate in the repair of damaged proteins. When the heat stress response diminishes, HSBP1 (Heat Shock Binding Protein1) and HSP70 participate in the negative regulation converting the active HSFA2 homotrimers into inactive monomers that participate in the recycling of HSFs. Regulative mechanisms of the heat stress response have been purposed also for the HSP90 which participate through the interaction of HSP70 in the negative regulation of the HSF in absence of stress (Saidi *et al.* 2011).

The protective effects of HSPs can be attributed to the network of the chaperone machinery, in which many chaperones act in concert (Wang *et al.* 2004). Protein denaturation is not the exclusive mechanisms to accumulate HSPs, even if the complete mechanisms of the heat stress regulation are not completely understood, many studies are focused on the role played by the membranes as primary sensor of heat shock (Saidi *et al.* 2011). Recent knowledge about heat perception and signalling mechanisms in plants have been reviewed by Mittler and collaborators in 2012.

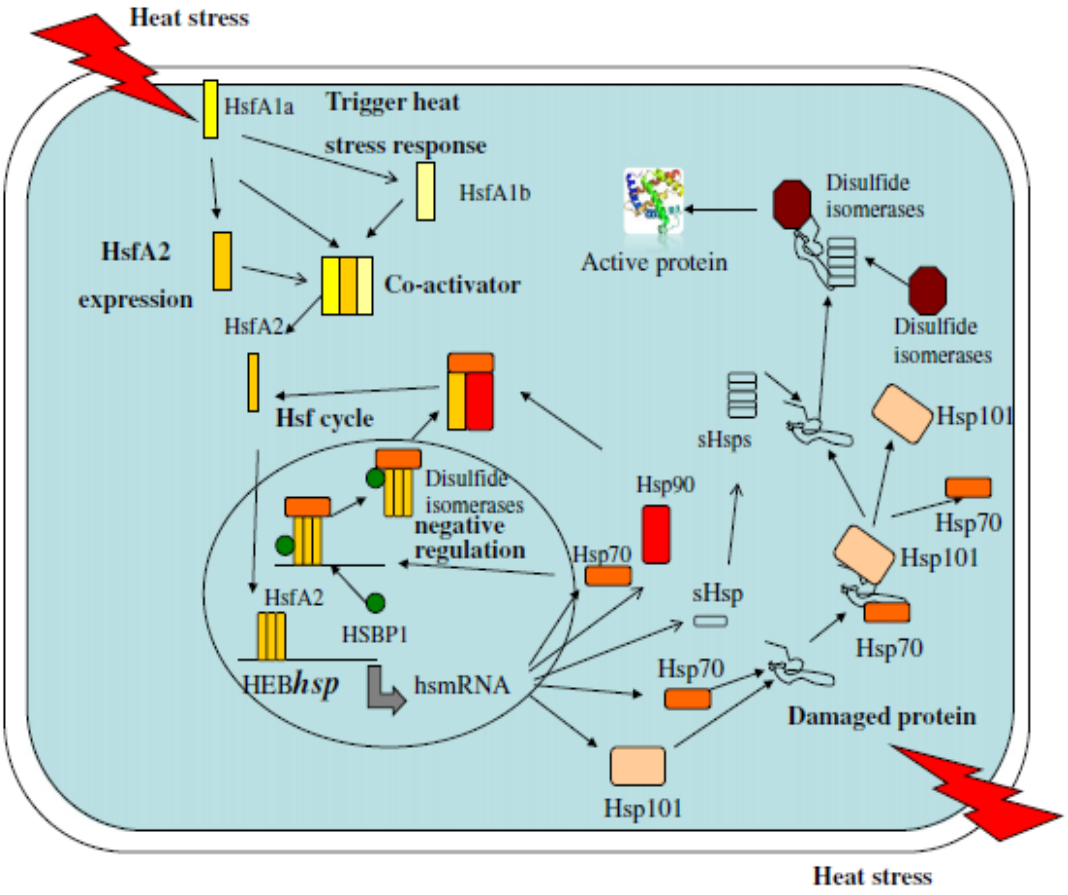



**Figure 1.8** Role and mechanism of HSPs and HSFs in the heat stress and acquired thermotolerance in plants (From Qu *et al.* 2013).

The five major multigenic families of *Hsp*s genes can be distinguished according to the protein molecular weights, amino acid sequence homologies and functions in the five protein families HSP100s, HSP90s, HSP70s, HSP60s and the small HSPs (sHSPs) (Waters 2013).

**HSP100s**

The HSP100 family chaperones are members of the large AAA ATPase superfamily. Rather than the regular chaperone function of preventing protein aggregation and misfolding, one particular function of this class is the reactivation of aggregated proteins by the resolubilization of non-functional protein aggregates and help the degradation of the irreversibly damaged proteins. Members of the HSP100 family were first described as components of the two-subunit bacterial Clp protease system (therefore they are named also Clp family) which consists of regulatory ATPase/chaperones (such as ClpA and ClpX) and proteolytic (ClpP) subunits. The family is further divided into two major classes and eight distinct subfamilies within these classes. Members of the first class contain two nucleotide-binding domains (ATP-binding domains), whereas those in the second class have only one ATP-binding domains. HSP100/Clp proteins are typically hexameric

rings (Schirmer *et al.* 1996). The mechanism for rescuing proteins from aggregation also involves the cooperation of another ATP-dependent chaperone system, the HSP70. The HSP100/Clp family solubilizes the aggregated protein and releases it in a state that can be refolded with the assistant of the HSP70 system. HSP100/Clp proteins have been reported in many plant species, such as *Arabidopsis*, soybean, tobacco, rice, maize and wheat. Like many other HSPs/chaperones, HSP100/Clp family chaperones are often constitutively expressed in plants, but their expression is developmentally regulated and is induced by different environmental stresses, such as heat, cold, dehydration, high salt or dark-induced etiolation (Agarwal *et al.* 2001). Genetic evidence indicating a role for this family of proteins in the acquisition of thermotolerance through functional complementation of the temperature-sensitive yeast HSP104 mutant cells was shown using *Arabidopsis AtHsp*101 cDNAs (Lee *et al.* 1995).

The cytoplasmic HSP101 members in *Arabidopsis* are necessary for the plant thermotolerance (Queitsch *et al.* 2000). In durum wheat different HSP101 stress-induced members have been identified that might have different roles during acquisition of thermotolerance (Gullì *et al.* 2007).

**HSP90s**

HSP90 family has a distinct feature from all the other molecular chaperones. In fact the known substrates binds by HSP90 are signal transduction proteins such as steroid hormone receptors and signalling kinases. The major role of HSP90s is to manage protein folding but it also plays a key role in signal-transduction networks, cell-cycle control, protein degradation and protein trafficking. In addition, it might also play a role in morphological evolution and stress adaptation in *Drosophila* and *Arabidopsis* (Krishna and Gloor 2001). HSP90s interact with the 26S proteasome and plays a central role in its assembly and maintenance. HSP90 is one of the major species of molecular chaperones that requires ATP for its functions. It is among the most abundant proteins in cells: 1–2% of total cellular protein (Frydman 2001). To implement its cellular role, HSP90 acts as part of a multichaperone machinery together with HSP70 and co-operates with a cohort of co-chaperones (among these the most important are: Hip (HSP70 interacting protein), Hop (HSP70/HSP90 organizing protein), p23 and HSP40). In plants, cytosol-, ER- and plastid-localized *Hsp90* genes have been isolated from several plant species, sharing 63–71% amino acid identities with HSP90 of yeast origin. In the *Arabidopsis* genome, the HSP90 family includes seven members: AtHSP90-1 to AtHSP90-4 constitute the cytoplasmic subfamily; AtHSP90-5, AtHSP90-6 and AtHSP90-7 are predicted to be localized to the plastid, mitochondria and ER, respectively. Although HSP90s chaperones are constitutively expressed in most organisms, their expression increases in response to stress in both prokaryotes and eukaryotes (Krishna and Gloor 2001). Expression of HSP90 in *Arabidopsis* is developmentally regulated and responds to heat, cold, salt stress, heavy metals,

phytohormones and light/dark transitions (Queitsch *et al.* 2002). Cytoplasmic HSP90 is responsible for pathogen resistance by reacting with resistance protein (R-protein) which is the signal receptor for pathogen. Moreover, in *Arabidopsis* has been investigated the role of HSP90 in the regulation of heat stress response (as indicated also in Figure 1.8). In absence of heat stress the cytoplasmic HSP90 negatively regulate HSF, whereas under heat stress HSP90 acitivity do not take place, thanks probably to the amount of denatured proteins in the cells that attract HSP90 and other chaperones leaving HSF to induce heat shock gene expression (Richter *et al.* 2010).

**HSP70s**

HSP70 consists of a highly conserved N-terminal ATPase domain of 44 kDa and a C-terminal peptide-binding domain of 25 kDa. HSP70 family chaperones are considered to be the most highly conserved HSPs, with 50% identical residues between the *Escherichia coli* homolog DnaK and the eukaryotic HSP70. HSP70s have essential functions in preventing aggregation and in assisting refolding of non-native proteins under both normal and stress conditions. They are also involved in protein import and translocation processes, and in facilitating the proteolytic degradation of unstable proteins by targeting the proteins to lysosomes or proteasomes (Hartl 1996). Some family members of HSP70 are constitutively expressed and are involved in assisting the folding of *de novo* synthesized polypeptides and the import/translocation of precursor proteins, while other members are expressed only in environmental stressing conditions and are involved in facilitating refolding and proteolytic degradation of non-native proteins (Frydman 2001).

The mechanism of action consists of successive cycles of substrate binding and release are coupled to the intrinsic ATPase activity of HSP70, which requires the participation of other interacting components (co-chaperones) such as DnaJ/Hsp40 (Boston *et al.* 1996). The specific role of individual HSP70 proteins is determined by their location in different subcellular compartments, by their differential expression at different stages of development or by their interaction with specific sets of HSP70-associated proteins. The overexpression of *Hsp70* genes positively correlates with the acquisition of thermotolerance resulting in enhanced tolerance to salt, water and high-temperature stress in plants; however the cellular mechanisms of HSP70s function under stress conditions are not fully understood (Wang *et al.* 2004). Various members of the HSP70 molecular chaperones have also been reported to be involved in protein import and translocation into chloroplasts and mitochondria, as well as in the cell-to-cell movement of proteins and viruses through the plasmodesmata (Aoki *et al.* 2002). In addition to its general chaperone functions HSP70 also plays a regulatory role in other stress-associated gene expression. The interaction between HSP70 and HSF has been suggested as a negative regulatory mechanism for HSF-mediated

transcriptional activation in the heat shock response as reported in Figure 1.8 (Qu *et al.* 2013, Morimoto 1998); moreover HSP70s are also involved in the modulation of signal transducers such as protein kinase A, protein kinase C and protein phosphatase (Ding *et al.* 1998). In this regard, the HSP70 chaperones might play a role in the modulation of the expression of many downstream genes during stress and under normal growth conditions. This last aspect has not been studied in plants yet (Wang *et al.* 2004).

**HSP60s**

The HSP60s family is known also as Chaperonins, as they are evolutionarily homologous to prokaryotic GroEL molecular chaperones. Chaperonins are further classified into two subfamilies: the GroE chaperonins (Group I) are found in bacteria, mitochondria and chloroplasts (e.g. GroE and chCpn60) and the CCT chaperonins (chaperonins containing t-complex polypeptide 1 (TCP1); Group II) found in Archaea and in the cytosol of eukaryotes. Chaperonins play a crucial role by assisting a wide range of newly synthesized and newly translocated proteins to achieve their native forms. The structure and function of chaperonins, especially of the Group I chaperonins, have been extensively studied. In prokaryotes, Group I chaperonins (e.g. GroEL of E. coli) consists of two distinct family members, chaperonin 60 (Cpn60) and chaperonin 10 (Cpn10) as co-chaperone, which function together in an ATP-dependent manner. They are double-ring assemblies composed of back-to-back stacked rings of identical or closely related rotationally symmetrical subunits (Saibil and Ranson 2002). Owing to the high conservation of the primary sequence among Group I chaperonins, it is generally accepted that organellar chaperonins function similarly to the bacterial ones. However, recent studies indicate that plastid chaperonins possess unique structural and functional properties that distinguish them from their bacterial homologs. In contrast to Group I, Group II chaperonins form eight or nine-member rings, each member is encoded by related but distinct genes, and they are independent of a general co-chaperone (although a protein cofactor, prefoldin, has been identified) (Wang *et al.* 2004). It has been shown that CCT chaperonins assist in the folding of tubulin and actin. Functional characterization of plant chaperonins is limited. It is generally agreed that they are important in assisting plastid proteins such as Rubisco (Boston *et al.* 1996). It has been reported that a mutated species of *Arabidopsis* chloroplast Cpn60 exhibits defects in chloroplast development and, subsequently, in the proper development of the plant embryo and seedling (Wang *et al.* 2004).

## 1.2.6 The small Heat Shock Proteins (sHSPs)

In plants sHSPs are among the most abundant proteins accumulated in plants upon heat stress (Waters 2013). The sHSPs are low-molecular-mass HSPs (12–40 kDa) and, usually, the different sHSPs are named on the basis of their molecular weight. Small HSPs are synthesized ubiquitously in prokaryotic and eukaryotic cells in response to heat and other stresses, and some sHSPs are expressed during developmental stages (Waters *et al.* 1996). The sHSPs have a high capacity to bind non-native proteins, probably through an hydrophobic interaction; they are not able to refold non-native proteins themselves but they stabilize and prevent non-native aggregation, thereby facilitating their subsequent refolding by ATP-dependent chaperones such as the HSP70 or HSP100. Recent findings showed that the sHSP18.1 identified in *Pisum sativum*, as well as HSP16.6 identified in *Synechocystis* under in vitro conditions, bind the unfolded proteins and allows further refolding by HSP70/HSP100 complexes (Mogk *et al.* 2003). In plants, conversely to other eukaryotes, sHSPs are numerous and have been distincted in different subfamilies on the basis of their sequence conservation and functional localization in the cells. The high diversification of plant sHSPs probably reflects a molecular adaptation to stress conditions that are unique to plants. The sHSPs are encoded by nuclear genes and different gene subfamily encodes for proteins found in a distinct cellular compartment; many of them are highly expressed under different environmental stresses, including heat, cold, drought, salinity and oxidative stress (Wang *et al.* 2003). An increasing number of data suggest a strong correlation existing between sHSPs accumulation and plant tolerance to stress (Sun *et al.* 2002, Wang *et al.* 2002, Hamilton and Heckathorn 2001, Heckathorn *et al.* 1998). Maize mitochondrial sHSPs were shown to improved mitochondrial electron transport during salt stress, mainly protecting the NADH:ubiquinone oxidoreductase activity (Complex I) (Hamilton and Heckathorn 2001). The importance of the chloroplast sHSPs was reported in *Agrostis stolonifera* grass, from which sHSP26.2 was isolated in a heat-tolerant variant. An identical sHSP, the sHSP26.2m, characterized by the presence of one non-sense point mutation, was isolated from the heat-sensitive variant and was not accumulated upon heat stress (Wang and Luthe 2003). The abundance of sHSPs in plants and their functional characteristics of binding and stabilizing denatured proteins suggest that sHSPs play an important role in plant-acquired stress tolerance (Wang *et al.* 2004, Wang *et al.* 2003, Waters *et al.* 1996). The numbers of sHSPs identified in plants varies between species; for example 19 sHSPs have been detected in *Arabidopsis* (Scharf *et al.* 2001, Visioli *et al.* 1997), while rice (*Oryza sativa*) has 23 sHSPs (Sarkar *et al.* 2009), and poplar (*Populus tricocarpa*) has 36 (Waters *et al.* 2008).

In Angiosperms, 11 diverse multigene subfamilies have been identified (Waters 2013, Table 1.2). These include six cytoplasmic/nuclear localized (CI–CVI) sHSP subfamilies and five organelles

localized subfamilies. The organelle subfamilies include one subfamily that localizes to the endoplasmic reticulum (ER), one to the peroxisome (PX), one to the chloroplast (CP) and two to the mitochondria (MTI and MTII) (Waters 2013). Within the same subfamily, different members of sHSPs share high sequence homology at the amino acid level (Sun *et al.* 2002). Sequence alignment of amino acids of sHSPs showed that members of different subfamilies do not share high sequence similarity, yet the secondary structure is conserved across subfamilies (Sarkar *et al.* 2009). Almost all of the sHSPs are induced upon heat stress, but some also in specific developmental stages like embryogenesis, seed maturation and pollen development (Table 1.2).

**Table 1.2** sHSP subfamilies in angiosperm and correspondent *A.thaliana* sHSP. The expression pattern and cellular locations of each sHSP members is based on published data (From Waters 2013).

| | Subfamily | Abbreviation | *A. thaliana sHSP* | Stress-induced expression | Developmental expression |
|---|---|---|---|---|---|
| 1 | Cytoplasmic/nuclear I | CI | Hsp17.4CI | H, Os, S, D, Ox, UV-B, W | Pollen, seeds |
| | | | Hsp17.6ACI | H, Os, S, D, Ox, UV-B, W | Pollen, seeds |
| | | | Hsp17.6BCI | H, Os, S, D, Ox, UV-B, W | Pollen, seeds |
| | | | Hsp17.6CI | H, Os, S, D, Ox, UV-B, W | Pollen, seeds |
| 2 | Cytoplasmic/nuclear II | CII | Hsp17.6CII | H, Os, S, Ox, UV-B, W | |
| | | | Hsp17.7CII | H, Ox, S, D, Ox, UV-B | None |
| 3 | Cytoplasmic/nuclear III | CIII | Hsp17.4CIII | Constitutive, H | Apex, flowers, pollen, siliques, seeds |
| 4 | Cytoplasmic/nuclear IV | CIV | Hsp15.4CIV | Constitutive, H, W | Apex, flowers, siliques |
| 5 | Cytoplasmic/nuclear V | CV | Hsp21.7CV | Constitutive | |
| 6 | Cytoplasmic/nuclear VI | CVI | Hsp18.5CVI | Constitutive, H | Flowers, pollen, siliques, seeds |
| 7 | Chloroplast | CP | Hsp21CP | H, Os, S | None |
| 8 | Mitochondrial I | MTI | Hsp23.5MTI | H, Ox, UV-B | |
| | | | Hsp23.6MTI | H, Ox, UV-B | |
| 9 | Mitochondrial II | MTII | Hsp23.5MTII | H, Os, S | Seeds |
| 10 | Endoplasmic reticulum | ER | Hsp22ER | H, Os | |
| 11 | Peroxisome | PX | Hsp15.7PX | H, Os, S, Ox, B | Seeds |

The characteristic feature of the sHSPs is the presence of an evolutionarily-conserved sequence of 80–100 amino acids called α-crystalline domain (ACD), related to a domain from the vertebrate α-crystallin proteins of the eye lens.

The sHSP primary structures consists of a conserved α-crystalline domain (ACD) flanked by an N- and a C- terminal region: three distinct regions that are believed to contribute differently to sHSPs function (Bondino *et al.* 2012, Figure 1.9). The ACD is the core structure of the protein, it is formed by two consensus regions, CRI and CRII that form a sandwich of two pleated β-sheets and has been shown to be essential for the construction of dimers and higher-order assemblies (Poulain *et al.* 2010). The N-terminus participates in substrate binding and binds denatured proteins (Jaya *et al.* 2009, Basha *et al.* 2006) while the C-terminus is involved in the homo-oligomerization and in the formation of heat stress granule (Kirschner *et al.* 2000). The N-terminal region preceding the ACD is variable in length and in the amino acid sequence. Despite the lack of conservation in the N-terminal region across sHSP subfamily, it is quite conserved within the subfamilies and therefore

would be interesting to understand the functional difference among the subfamilies. The N-terminal region of the protein carried also a targeting sequence (or leader sequence) needed for the localization of the organelle-targeted sHSPs into the proper cellular compartments (Waters 2013, Figure 1.9).

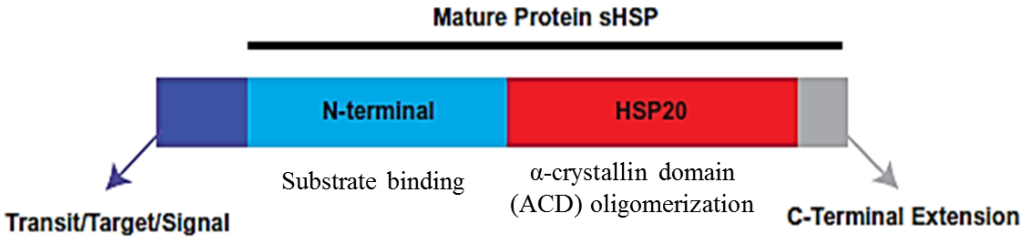



**Figure 1.9** Small heat shock protein domains. The sHSP monomer contains a variable N-terminal domain (light blue) involved in substrate binding, a conserved α-crystalline domain (ACD) or HSP20 domain involved in oligomerization, and a variable C-terminal extension (grey). The sHSPs that localize to the organelles (chloroplast, mitochondrion, ER, peroxisome) have a transit, targeting or signal sequences (dark blue) that are needed for the entry into the organelles and are not part of the mature sHSP. (Modified from Waters 2013).

The presence of subfamily-specific N-terminal motif strengthens the evidence that N-terminal region is functionally important and crucial for substrate binding. The chloroplast sHSP have a methionine-rich amphipathic α-helix that is highly conserved across angiosperms; this region has been demonstrated to be important in substrate binding and is sensitive to the chloroplast redox state (Sundby *et al.* 2005, Gustavsson *et al.* 2001).

The high-resolution crystal structure has been determined for the sHSP16.9 from wheat *(T.aestivum)* (van Montfort *et al.* 2002; Figure 1.10a, b, c). The HSP16.9 forms a dodecamer with three tetramers that forms two six-membered ring and therefore a hexameric double-disk (Figure 1.10c). The structure is a β-sandwich of two antiparallel sheets and the building blocks of the oligomer are dimers (Figure 1.10a).

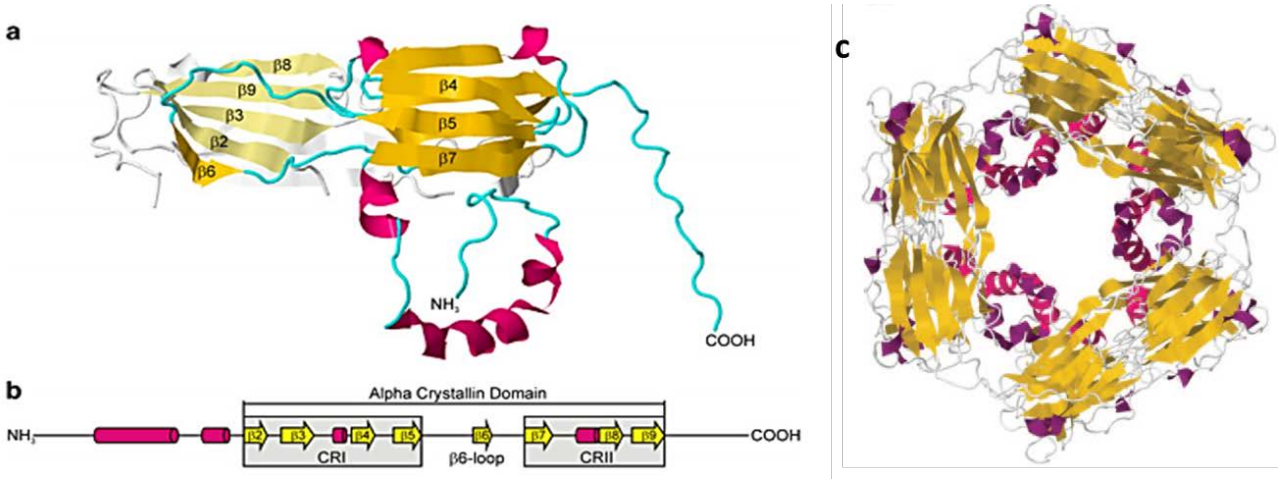



**Figure 1.10** Structural organization of a typical sHSP. A) Cartoon representation of the X-ray structure of a wheat sHSP dimer (PDB ID: 1GME). B) Primary structure of the protein. The ACD consists of the conserved regions I (CRI) and II (CRII) and the β6-loop. The conserved regions form a sandwich of two β-pleated sheets. The β6-loop bulges out of the monomer and forms a β-sheet with the β2-strand from another monomer, thereby establishing a dimer. **C)** The oligomeric structure of Hsp16.9 that is a dodecamer that form a β-sandwich. The strands are in yellow, the helixes are in red and the loops are in grey. The loops of a single monomer are in light blue. (Modified from Bondino *et al.* 2012 and Waters 2013).

Small HSPs are shown to be ATP-independent molecular chaperones. The sHSP are considered the first line of defence in the cell when protein begin to misfolded but the details of the specific interactions between the sHSP and their protein substrates are still unclear. The current model for the function of the sHSPs chaperones is reported in Figure 1.11. According to this model the sHSPs are dynamic oligomeric structures existing in two states for the binding of non-native proteins: an inactive, low-affinity state and an active, high-affinity state. The equilibrium between these two modes is controlled by the temperature and upon heat stress the equilibrium is shifted toward the high-affinity state where the active oligomer is competent to associate with substrate proteins to form a stabile sHSP–substrate complex. This allows efficient prevention of irreversible substrate aggregation. The release of active substrate proteins from these complexes requires cooperation with proteins of the ATP-dependent family such as HSP70/HSP40 or HSP100. Although HSP70/HSP40 and HSP100 can act directly on protein aggregates, the presence of sHSPs increases the process efficiency (Haslbeck *et al.* 2005).

The most important evidence for the functional role of sHSP in plant thermotolerance are related to HSP101 (mentioned above) and sHSPs (Rampino *et al.* 2012, Kotak *et al.* 2007 and Figure 1.7). Between the sHSP, the chloroplast one, known as HSP21 in *Arabidopsis* and HSP26 in wheat are implicated in the protection of the photosynthetic machinery and in particular of the PSII under heat stress (Chauhan *et al.* 2012, Heckathorn *et al.* 1998).

Evidence of the positive relation of the chloroplast HSPs with the thermotolerance acquisition have been found in tomato species (Heckathorn *et al.* 1998) and also in *Chenopodium album* where the ecotypic variation in chloroplast sHSPs have been related to different levels of thermotolerance (Shakeel *et al.* 2011). In wheat the massive induction of sHSP26 has been linked to the thermotolerance acquisition (Rampino *et al.* 2012, Rampino *et al.* 2009b) and different levels of thermotolerance have been related to genotypic variation.

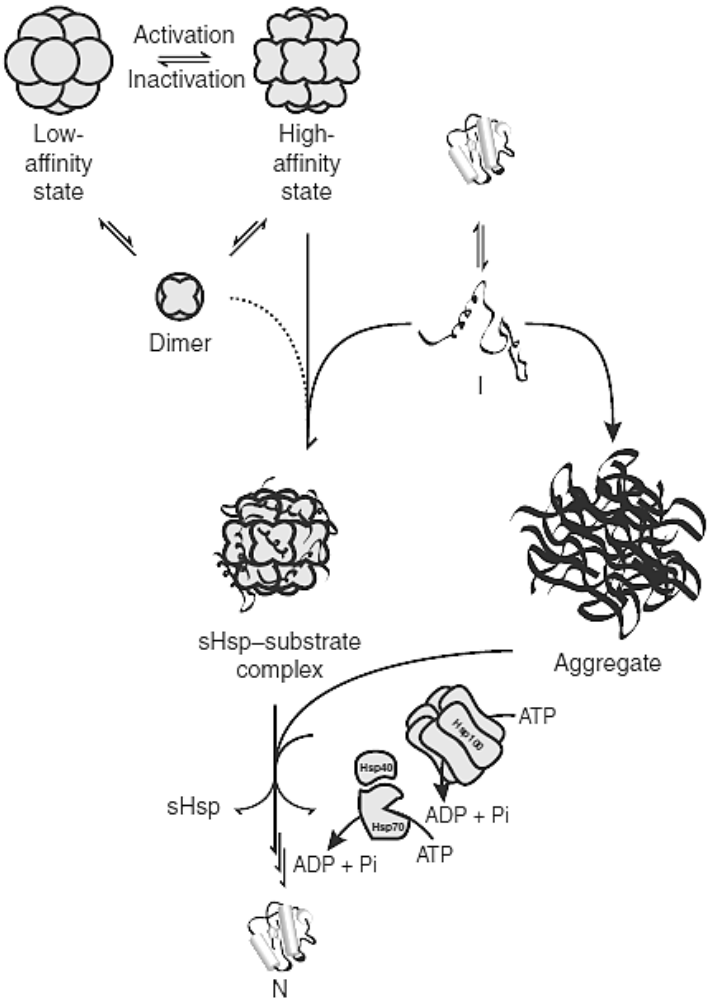



**Figure 1.11** Model for sHSP chaperone activity. The oligomeric structures of sHSP in active high-affinity states binds non-native proteins in heat stress conditions to form a stable sHSP-substrate complex. This allows efficient prevention of irreversible substrate aggregation and increase the efficiency of the refolding operated with the collaboration of ATP-dependent chaperone such as HSP70 and HSP100 (From Haslbeck *et al.* 2005).

# 1.3 Functional genomics in crop improvement

## 1.3.1 TILLING in functional genomics

Functional genomics aim to acquire knowledge about gene function through linking the mutation in a specific gene to the phenotype change in the mutated organism (Wang *et al.* 2012). Two approaches can be distinguished. A forward genetics approach (meant "from phenotype to the gene") through which mutated populations have been created and screened for alterations in the traits or biological process of interest, and a reverse genetics approach (meant "from gene sequence to phenotype") based on the alteration of a gene structure or activity followed by the analysis of associated phenotype changes. Artificial mutagenesis is at the basis of both these approaches as it allows to induce mutations in genome at higher frequency than typically observable in nature. After the development of large mutant collections the former approach was commonly used to identify a phenotype of interest and isolate the responsible gene sequence by using a map-based cloning approach. Thanks to the recent improvements of sequencing technology and the development of annotation database that generate collections of genes with no understood biological functions, an emerging approach was the disruption of specific genes to test their function *in vivo.* The reverse genetics strategy opened new possibilities for the application of mutation techniques in both basic studies and crop improvement (Kurowska *et al.* 2011). To date, many successful varieties have been developed through induced mutations including released mutant cultivars improved for biotic and abiotic stress resistance, improved quality, yield and plant architecture that are the common targets of improvement through breeding (Oladosu *et al*. 2015). The worldwide officially released mutant crop varieties are listed in a specific database named Mutant Variety Database (MVD) developed by the International Atomic Energy Agency (IAEA) together with the Food and Agriculture Organization (FAO).

TILLING (Targeting Induced Local Lesions IN Genomes) is a reverse genetics technique that usually combines mutagenesis of plants by chemical mutagens with high-throughput screening for single nucleotide mutations. It allows the isolation of mutations in genes of interest and can be used to investigate gene function and to hunt for new alleles with functional characteristics useful in breeding (Wang *et al.* 2012, Wang *et al.* 2010, Henikoff *et al.* 2004).

The main advantages of TILLING in comparison to other reverse genetics techniques as insertional mutagenesis is being a not-transgenic approach and hence most appropriate for crop species, allowing the easy introgression of a desired alleles into different genetic background or the use of mutagenesis in advanced genetic material avoiding the necessity of an efficient transformation system and the problems related to the GMO use.

In polyploids organisms, a reverse genetics approach is widely used because the presence of multiple homoeologous copies limits the use of forward genetics phenotypic screens; in fact the effect of single-gene knockout in polyploids is frequently masked by the functional redundancy. For the same reason, polyploids organisms can tolerate a high mutation loads leading to a high probability to identify mutation in the gene of interest in relatively small populations (Uauy *et al.* 2009). Nevertheless, the high mutation frequency is an important drawback for the mutation-phenotype correlation due to the long time required to clean the background mutations through backcross prior to gene characterization. The improvements of techniques to reveal multiple deleterious mutations helped to partially overcome this issue leading to the possibility to score most phenotype after only one or two generation. This can be done by looking for the segregation after the self impollination of heterozygous plant or by single cross between homozygous alleles, where the maintenance of the phenotype in the $F_1$ could guide to the individuation of the correct phenotype (Henikoff *et al.* 2004). Moreover, the use of homozygous sibling lines segregating for wild type and mutant alleles provides a powerful tool to reduce interference from background mutations and correlate targeted mutation with function (Parry *et al.* 2009). For traits involving several genes that contribute to the phenotype, a high density chemical mutagenesis comparable to those used in TILLING mutant lines continue to be a basic learning tool for genetics (Henikoff *et al.* 2004).

TILLING is also an invaluable resource for structure-function studies since it generates allelic series including null alleles at low frequency. This last characteristic is important to study the protein functions through evaluating the effect of different missense mutations.

In the recent years TILLING have been widely used to discover gene function and to mine new alleles in complex genomes like wheat (Hazard *et al.* 2012, Botticella *et al.* 2011, Sestili *et al.* 2010, Dong *et al.* 2009, Slade *et al.* 2005) and it's emerging as important plant breeding tool (Sestili *et al.* 2015, Wang *et al.* 2012, Uauy *et al.* 2009). Mutant lines with allelic variation on *Glutenin* genes were identified in a durum wheat TILLING population leading to the identification of two lines with high protein contents potentially useful as genetic resource for breeding programs to improve technological quality parameters on dough (Meryem *et al.* 2014). In spite of the long time needed to lead a final new improved variety, TILLING have been successfully applied to develop a high-amylose durum and bread wheat varieties through combining null mutation in each of the three *SbeIIa* homoeologous (Slade *et al.* 2012, Slade *et al.* 2005).

Different mutant populations have been created for tetraploid and exaploid wheat for TILLING to determine gene function (Uauy *et al.* 2009). A tetraploid TILLING population could be used to generate mutants for basic research project due to the relatively easy possibility to generate null mutants through a single generation of crosses between A and B genome mutations, followed by

selection of homozygous double mutants in the $F_2$ populations. Hexaploid population instead is most relevant for breeding application in bread wheat but require an additional cross to combine the three mutant alleles within a single plant (Borrill *et al.* 2015, Uauy *et al.* 2009). The TILLING method is composed of different steps starting with the creation of the mutagenized population followed by the sample pooling and the mutation detection.

#### 1.3.1.1 Creation of the mutagenized population

The creation of a mutagenized populations is the basis of the functional genetic approach in both forward and reverse genetics. Many different mutagen agents can be chosen on the basis of the wanted effect and in any case a compromise between highest mutation density and fecundity of the plants should be reached (Lee *et al.* 2014). Chemical mutagenesis is widely used as it introduces mostly single nucleotide mutations that are generally of the most interest for crop improvement through breeding. One of the most used chemical mutagen agent is ethyl methane sulphonate (EMS) which alkylate guanine bases creating mainly G–A and C-T transitions on the DNA and inducing often novel stop codons. The well-characterized and highly predictable effect of EMS allowed to the development of computational tools to define the putative regions of a genes with the highest probability of accumulating deleterious mutations (McCallum *et al.* 2000).

EMS-mutagenized populations are generated by exposing seeds (M0) to the mutagen and allowing the resultant M1 plants to self-fertilize and originate M2 seed where mutations will be stably inherited. The seeds must be exposed to sufficient mutagen to ensure a high level of mutations but without affecting viability and fertility. To ensure the greatest number of unique novel mutations it is recommended that only one seed is taken from each M1 plant. Leaf material is taken from the resultant M2 plants for the isolation of genomic DNA that is used as the resource for mutation detection. The M2 plants are allowed to mature and the M3 seeds are archived and stored so that gene function can be studied in any plants in which mutations are identified. The M3 population is still segregating and not all the M3 plants will carry the mutations identified in the M2 parent. The mutagenized population may be taken through further generations by single seed descent (SSD) to generate near-homozygous material (with estimate 97% of the mutations being homozygous at M6), although up to half of all mutations present in the M1 are lost in the process (Figure 1.12) (Parry *et al.* 2009).

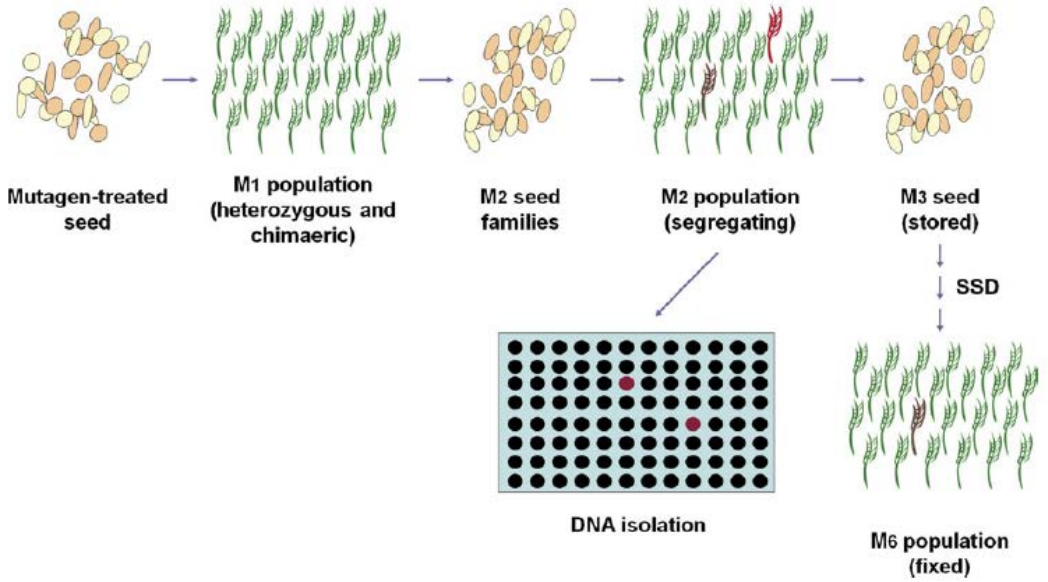



**Figure 1.12** The general strategy used to create a mutant population in self-fertilizing crops (From Parry *et al.* 2009).

To ensure that any gene of interest carries sufficient significant mutations, the populations of induced mutations may need to be very large, however polyploidy species have a high tolerance of mutations and thus populations saturated with mutations can be smaller (Wang *et al.* 2012).

This aspect make polyploidy species highly suitable for TILLING. In published TILLING populations the average mutation densities from diploid species are 1/380 kb, whereas it increase to 1/49 kb in tetraploids (as durum wheat and tobacco) and 1/32 kb in hexaploids (as bread wheat and oat) (Wang *et al.* 2012). This high mutation frequency observed in polyploidy TILLING populations facilitates the identification of large allelic series in target genes screening a relatively small number of individuals and also increase the probability of identify at least one non-sense mutation per target genes (Wang *et al.* 2012). In according to the data of published TILLING population Wang and co-workers (2012) state the probability of identifying at least one truncation mutation by screening 2000 individuals at 98%, 91% in hexaploid and tetraploid population respectively, whereas only 27% in diploids.

Numerous TILLING resources have been developed in wheat, including tetraploid and hexaploid libraries in several varieties (Dong *et al.* 2009, Parry *et al.* 2009, Uauy *et al.* 2009) providing different background to create variability. Recently a TILLING resource have been developed in the élite cultivar Svevo, one of the most important durum Italian varieties, providing variation in a high-quality genetic background, particularly suitable for breeders and researchers (Bovina *et al.* 2014).

### 1.3.1.2 Mutation detection methods

The most used techniques to screen a TILLING population are based on the PCR amplification of a target region from pooled DNA of mutagenized plants followed by eteroduplex detection. The amplified products are denatured by heating and then cooled slowly allowing random re-annealing. During the re-annealing between mutated and non-mutated fragments, heteroduplex molecules created by the mismatch between wildtype and mutant nucleotide will be detected. Pooling the DNA samples from several lines is important to ensure the presence of the heteroduplex and to increase the throughput of the technique. Once a positive pool is identified, the positive sample can be detected and confirmed through sequencing. The most common detection strategies for TILLING screening includes the enzymatic digestion with the CEL I endonuclease and the high-resolution melt analysis (HRM) (Parry *et al.* 2009).

**CEL I endonuclease digestion**

A common detection system rely on the use of a mismatch-specific nuclease from celery, CEL I (Figure 1.13A). In this technique the pooled DNA is amplified using fluorescently labelled gene-specific primers (with different fluorophores on the 3' and 5' tags) (Comai and Henikoff 2006). Heteroduplex molecules are formed, after denaturation and renaturation, when a pool includes at least one plant with a mutation in the amplified fragment. The re-annealed products are digested with CEL I nuclease which cleaves the heteroduplex mismatch site. Cleaved fragments are resolved on gels (polyacrylamide denaturing gel such as a LI-COR gel system, or agarose gel), that allow the identification of positive pools and therefore the individual samples carrying the mutation within the amplicon. The position of the mutation can be estimated looking the size of the fragments carrying 3' and 5' fluorescent tags (Sharp and Dong 2014, Colbert *et al.* 2001). The use of fluorescent labelling, with different fluorophores on the forward and reverse primers, allow to easily detect the products of the heteroduplex cleavage in separate channels of the DNA analyser. The sum of the fragments size should equal that of the original amplicon, while the sizes of the respective fragments is an indication of the approximate position of the mutation within the amplicon, facilitating subsequent identification by sequencing (Parry *et al.* 2009).

**High Resolution Melting (HRM)**

High-resolution melting (HRM) analysis is a simple, closed tube, post-PCR method used to identify genetic variation. In respect to the CEL I digestion HRM is more rapid and automatable, therefore considered high throughput. HRM analysis depends on the loss of fluorescence from intercalating dyes bound to double-stranded DNA during thermal denaturation (Ririe *et al.* 1997). HRM experiments involve three main steps: (1) PCR amplification of a target region to high copy

numbers in the presence of a fluorescent double-stranded (ds) DNA binding dye, (2) melting of the fluorescently labelled product, and (3) analysis of melt curves (Tucker and Lam Huynh 2014). The method is highly sensitive and can discriminate DNA sequence variants on the basis of length (such as insertions or deletions), composition (Single Nucleotide Polymorphisms, SNPs) or strand complementarity (such as heterozygous or homozygous material). Using the same experimental principles and techniques, HRM can be used for discovery of genetic variants in "gene scanning" experiments as in TILLING application, or for genotyping of known sequence variants in optimized assays (Tucker and Lam Huynh 2014).

The success of HRM as a screening technique has largely been due to the development of third generation dsDNA binding dyes and high precision melting instruments that are capable of acquiring large amounts of fluorescent data over incremental temperature changes. The third generation dsDNA binding dyes, such as SYTO®9(Life Technologies), EvaGreen™ dye (Biotium, Inc.), or LC Green® (BioFire Diagnostics, Inc.) can be used at higher concentrations to completely saturate dsDNA samples without inhibiting PCR (Tucker and Lam Huynh 2014).

The dsDNA binding dye intercalates with the double-stranded PCR product emitting fluorescence. Following PCR, the fluorescently labelled dsDNA sample is melted through a range of temperatures on a specialized melting instrument that controls the temperature while continuous monitoring the fluorescence. As the DNA progressively dissociates into single-stranded DNA, fluorescence of the sample decreases until the sample is completely single-stranded and no fluorescence is detected. Analysis of the fluorescent data in instrument-specific software generates melt profiles that are characteristic and reflect the DNA composition of a given sample. Data can be visualized as raw melt curves, as normalized melt curves or as difference plots (Figure 1.13B). The critical aspect for a successful HRM experiment is the primer design; primers should amplify a specific product 100–300 bp in length, have a melting temperature (*Tm*) of 60 °C, and produce no secondary structures (Tucker and Lam Huynh 2014). HRM optimization strategies often involve testing multiple primer sets to identify the most robust pair and optimize the PCR cycling conditions. Examples have been reported in diverse species including diploids (Lochlainn *et al.* 2011) and polyploids (Botticella *et al.* 2011, Dong *et al.* 2009).

Mutation scanning by HRM in polyploid wheat requires a two-step amplification process, using firstly homoelogue-specific primers to amplify a larger amplicon containing several coding regions, followed by HRM analysis using primers specific for each target region (Figure 1.13B).

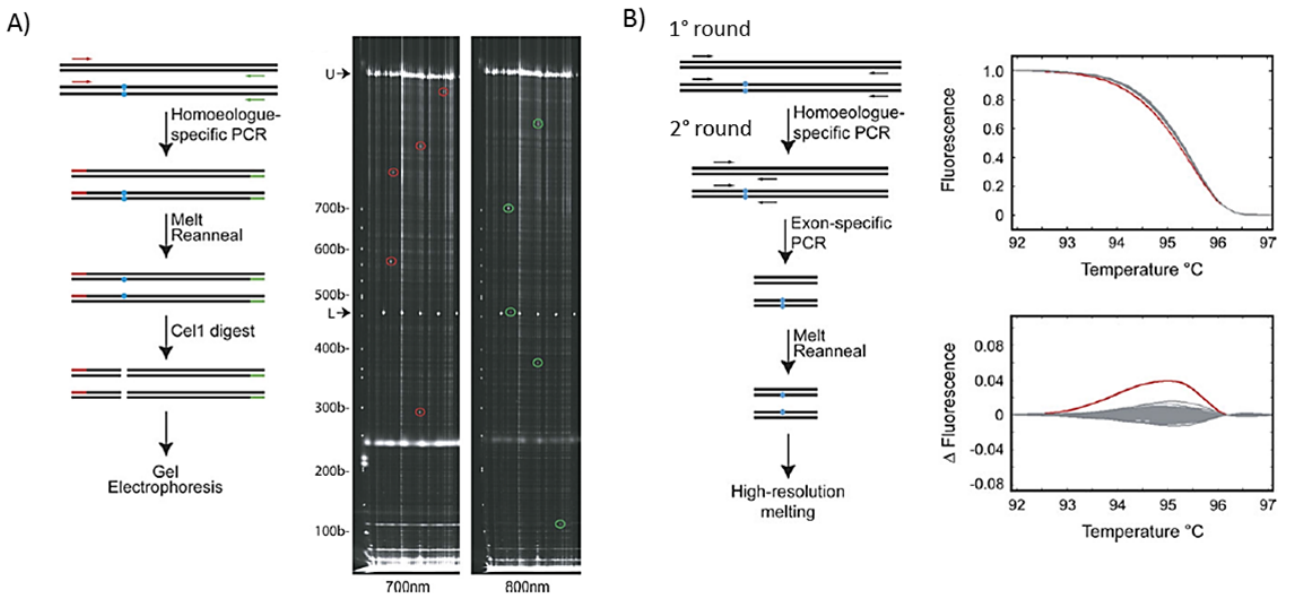



**Figure 1.13** Mutation detection techniques for TILLING screening. A) Mutation detection in wheat by CEL I-based TILLING with a schematic of PCR and CEL I digestion on the left and typical gel running on the right. B) Mutation detection in wheat using high-resolution melt analysis (HRM) with a scheme of PCR and heteroduplex production on the left and an example of melt curves on the right. Temperature-normalized fluorescence curves (left up) and difference fluorescence plotted against temperature from melt analysis of the second round PCR product (left down) are reported. Wild-type samples are in grey, the single putative mutant pool identified through its different melting profile is in red. (Modified from Parry *et al.* 2009).

Several bioinformatics tools are needed in all TILLING steps. The target gene sequence must be known or retrieved by searching the most common databases (NCBI, Unigene, etc.), the availability of the entire genomic sequence is crucial for the determination of the gene region to target by TILLING. This is easily performed with the CODDLE (Codons Optimized to Discover Deleterious Lesions) software which performs a BLAST alignment to identify conserved region with the most potential to generate deleterious changes, that are usually coding regions.

PARSESNP (Project Aligned Related Sequences and Evaluate SNPs; Taylor and Greene 2003) is another important tool for the analysis of polymorphisms in genes; by using a reference DNA sequence, an exon/intron position model and a list of polymorphisms, it determines the effects of these polymorphisms on the expressed gene product, as well as the changes in restriction enzyme recognition sites (Taylor and Greene 2003). PARSESNP returns a PSSM (Position Specific Scoring Matrix) and a SIFT (Sorting Intolerant from Tolerant) value that help in the prediction of the severity of missense changes. Large positive PSSM (>10) means that missense change in the analysed sequence of the DNA can be dramatic for protein function. SIFT allows to predict the severity of change on the basis of the position and the physical proprieties of the amino acid substitution. A SIFT score lower than 0.05 could have the same effect as a PSSM higher than 10 and the substitution is predicted as not tolerate (Botticella *et al.* 2011, Kurowska *et al.* 2011).

## 1.3.2 TILLING *by sequencing* and *online* TILLING

The classical TILLING approach rely on the generation of PCR-based target amplicons and the detection of variation on these targets. Currently the use of TILLING for functional genomics analysis in wheat is mainly limited by the unavailability of genomic sequence and the difficulties to obtain genome-specific amplicon of the target gene.

Thank to the recent improvement of sequencing technologies, mutations and polymorphisms can be discovered through high-throughput sequencing followed by comparison of a putative new allele with a reference sequence. The new high-throughput sequencing technologies, called Next-Generation Sequencing (NGS) include different commercially available platform as mainly 454 genome sequencer from Roche, Solexa genome analyzer from Illumina, SOLiD from Applied Biosystem and HeliScope from Helicos (Metzker 2010). The NGS approach have been applied to TILLING leading to increase the throughput and accurate recovery of induced mutations extending the analysis to multiple amplicons within one screen (Borrill *et al.* 2015, Tsai *et al.* 2011). The validity of TILLING by sequencing had been demonstrated in an EMS-mutagenized tomato population using the Roche 454 technology combined with a multidimensional pooling strategy leading to the identification of two mutations in the *eIF4E* gene among more than 3,000 M2 families in a single sequencing run, and six haplotypes of the *eIF4E* gene were discovered by re-sequencing three amplicons in a subset of 92 tomato lines (Rigola *et al.* 2009). A method based on Illumina sequencing of target genes amplified using bar-coded primers for sample tracking from multidimensional pool of more than 700 individuals per experiment was successfully performed for rice and wheat leading to the discovery of novel rare mutants (Tsai *et al.* 2011). The NGS opened the possibility of a fast genome-wide discovery of mutations from a population of individuals, however the whole genome sequencing costs are still prohibitive limiting the possibility for whole-genome re-sequencing of mutant populations. A reduced representation methodology such as exome-capture, which allow to restrict sequencing to the coding part of the genome, has been demonstrated to be effective in animal and plant genomes and could constitute a powerful tool for mutation discovery when applied to mutagenized populations (Bolon *et al.* 2011, Ng *et al.* 2010, Ng *et al.* 2009). Recently, exome capture has been demonstrate to be a suitable method to describe the types and frequency of mutations present in EMS-mutagenized rice and wheat plants (Henry *et al.* 2014). A scheme of the exome capture approach performed by Henry and co-workers (2014) on EMS-mutagenized rice population is reported in Figure 1.14. The target exon region were arrayed into overlapping capture probes and were used for mutation discovery in an EMS-mutagenized rice population (*O.sativa* cv. Nipponbare). The M2 generation of the rice TILLING population was selected; genomic DNA was extracted from leaves tissues and genomic sequencing libraries were

produced from each individual. Pooled DNA was subjected to exome capture using custom rice exome probes. The amplified post-capture DNA was sequenced on Illumina apparatus. The sequence data were processed and mutations were identified using a custom mutation discovery pipeline called MAPS (Comai Lab, UC Davis Genome Center; http://comailab.genomecenter.ucdavis.edu/index.php/MAPS) that allows also to differentiate homozygous or heterozygous mutations.

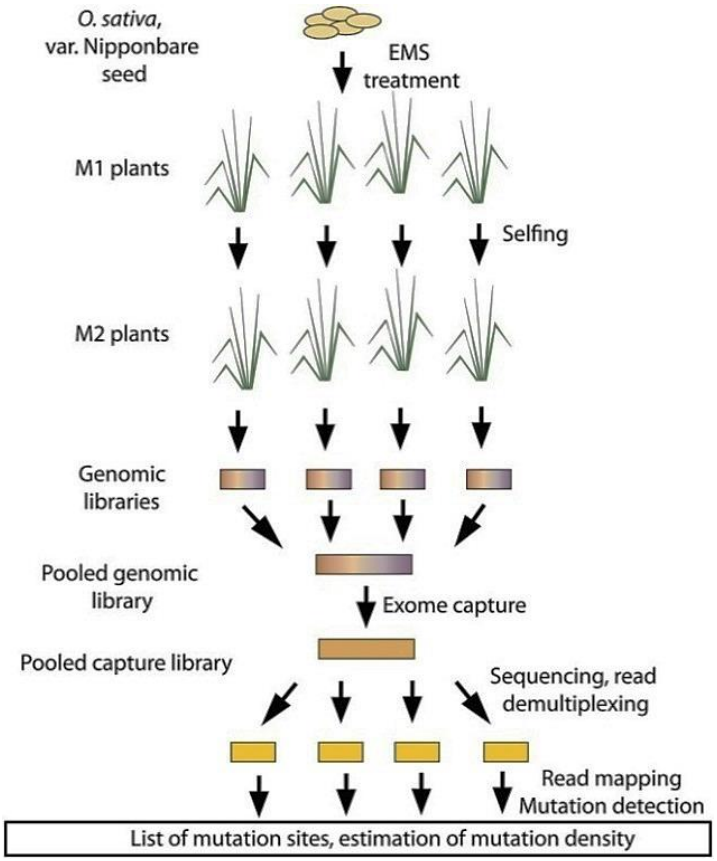



**Figure 1.14** Production of EMS-mutagenized rice samples and exome capture analysis. Independent M2 mutant individuals were produced by EMS treatment of seeds followed by selfing of the M1 individuals. Indexed genomic libraries were produced independently from each M2 plant and pooled prior to sequence capture. Captured sequences were submitted for Illumina sequencing. Sequencing reads were assigned to specific M2 individual based on their index sequence. Mutation detection and estimation of mutation density were performed for each M2 individual (From Henry *et al.* 2014).

The wheat exome is estimated to account for less than 2% of the complete genome and exome capture in wheat allows to reduce the wheat genome from 17 Gb to <150 Mbp suitable for NGS (Figure 1.15a). An exome capture platform have been developed for wheat including gene selected from previous published transcriptome (Krasileva *et al.* 2013) and have been used to undertake the sequencing of EMS-mutants of the durum wheat cv. Kronos. The wheat online TILLING database is currently under construction and will be publicly available after publication (Uauy C. personal

communication, Dubcovsky *et al.* 2015). Exome capture and re-sequencing of TILLING populations will enrich the wheat online TILLING resource giving the possibility to identify mutations in the target gene with a predicted effect on the protein function leading TILLING in wheat toward an *in silico* screening activity (Figure1.15b, Borrill *et al.* 2015, Wang *et al.* 2012). The feasibility of using exon capture for genome re-sequencing as a method of mutation detection in polyploid wheat have been recently confirmed by King and co-workers (2015).

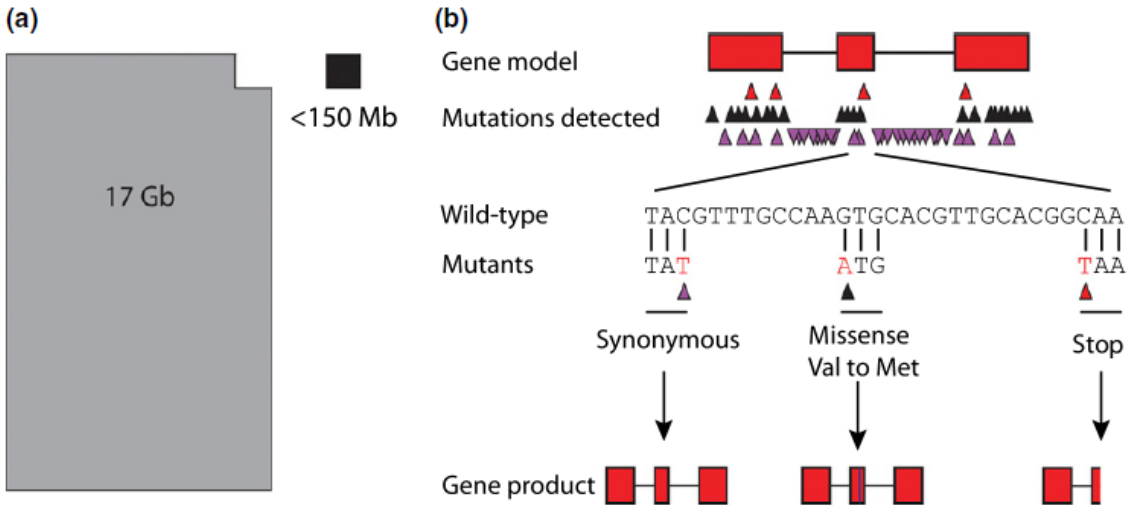



**Figure 1.15** *In silico* TILLING. (a) Exome capture reduces the wheat genome from 17 Gb to < 150 Mb for next-generation sequencing. (b) Variant calling identifies mutations within gene models. Amongst the mutations detected, the first row of triangles depicts truncation mutations (red), the second depicts missense mutations (black), and the third depicts silent mutations (purple). Upward-pointing triangles indicate an exonic change and downward-pointing triangles indicate an intronic change. The sequences of wild-type and mutant individuals within the second exon are shown, along with the effects on amino acid usage. The resulting gene products are indicated at the bottom. (From Borrill *et al.* 2015).

## 1.3.3 EcoTILLING

In a modified TILLING strategy called Eco-TILLING (Ecotype- Targeting Induced Local Lesions IN Genomes; Comai *et al.* 2004) the mutation detection could be used to discover polymorphisms in natural populations to exploit the variation existing in nature (Till *et al.* 2006). In this case, instead of using pools of DNA from mutagenized plants, the DNA of different ecotypes are pooled with a reference sample. The discover of natural genetic variations could have practical applications for plant breeding because the variation could be rapidly introgressed in traditional breeding programs and also for conserving the natural diversity (Chen *et al.* 2014). A recent paper shown that the natural alleles of *mlo* and *Mla* found by EcoTILLING in wild barley accessions can be combined to develop more durable resistance to powdery mildew disease (Mejlhede *et al.* 2006),

demonstrating that the natural alleles detected by EcoTILLING are valuable for molecular improvement of key agronomic traits. In another study Ecotilling approach have been used to investigate natural variations of *Pina* and *Pinb* genes in a Chinese wheat germoplasm collection exploiting useful resource for improving wheat kernel texture through molecular breeding (Wang *et al.* 2008). More recently EcoTILLING have been used to dissect allelic variation in barley *sHsp17.8* leading to the identification of SNPs related to different agronomic traits potentially contributing to drought tolerance (Xia *et al.* 2013).

### 1.3.4 SNP-based functional markers

Molecular markers, and especially DNA-based markers, are widely used for: characterize genetic diversity in germplasm collections, mapping and identification of genes associate with important traits, mining of superior alleles and selection of desirable genes or traits in breeding programmes (Kadirvel *et al.* 2015). DNA markers are the variations observed in a particular DNA portion among different individuals of a species; these variations may be due to different mutations (insertion, deletion, substitution or replication errors of tandem repeated DNA). The DNA markers can be considered an ideal genetic marker system as they meets the characteristics to be unlimited in number, insensitive to environmental conditions, highly reliable and easily automatable. Several DNA marker technologies have been developed and improved over the years from restriction based to PCR and to sequence based technologies; they differs on the basis of the polymorphism and detection methodology (Kadirvel *et al.* 2015). Markers can be classified on the basis of their ability to distinguish between heterozygotes and homozygotes. A codominant marker is able to distinguish the heterozygotes from the homozygote that carries that allele, whereas a dominant marker detects just the presence or absence of one particular allele.

Markers may belong to either transcribed or non-transcribed regions of the genome; those developed without prior knowledge of their function are called ‘anonymous’ or ‘random markers’ whereas those derived from the functional variants present in genic region, which causally govern the trait variation, are called Functional Markers (FMs). The FMs are diagnostics of desired traits associated alleles that make them markers of choice for targeted MAS in plant breeding programs can be used in the segregating population without the issue of recombination (Varshney *et al.* 2005). With the rapid progress in functional genomics together with genome and RNA sequencing, a lot of sequence data became available for the development of markers from functional sequence data in the form of bacterial artificial chromosome (BAC) clones, cDNA clones, Expressed Sequence Tags (EST) and gene from *in vitro* and *in silico* experiments available in public databases.

SNPs (Single Nucleotide Polymorphisms) are single nucleotide base difference between DNA sequences of individuals, which occur within coding/non-coding regions of genes or in the intergenic regions. They are numerous and occur at different frequencies in different regions of the genome. SNP becomes the most preferred marker system nowadays due to their abundance (two to three per kilobase), high level of polymorphism, high throughput capability and cost-effectiveness (Kadirvel *et al.* 2015). The use of SNPs as FM is convenient as they are present in large number and distributed throughout the genome (Kage *et al.* 2015, Jones *et al.* 2009). The FMs can be developed for these sequence polymorphisms and used in crop improvement (Bagge and Lübberstedt 2008, Andersen and Lübberstedt 2003). The polymorphisms involving non-synonymous SNPs, lead to phenotypic variations (Risch 2000); however, synonymous mutations (silent mutation) have also been reported to alter the phenotype (Sauna and Kimchi-Sarfaty 2011, Goymer 2007). The numerous FMs developed in wheat, rice, maize, sorghum, millets and other crops, and successfully used in breeding programs have been recently reviewed by Kage and co-workers (2015). In wheat, more than 30 genes were cloned and 97 FMs were developed to identify 93 alleles associated with food quality, agronomic and disease resistance traits (Liu *et al.* 2012). Thanks to the recent increment of sequencing data available for several crop species, SNP markers have been developed and have largely replaced other type of molecular markers as SSR (Simple Sequence Repeats). Because of their low assay cost, high genomic abundance, locus specificity, codominant inheritance, simple documentation, potential for high-throughput analysis, and relatively low genotyping error rates (Schlotterer 2004, Rafalski 2002), SNPs have emerged as powerful tools for many genetic applications and are often the preferred marker system in marker assisted breeding and MAS (marker-assisted selection) (Kage *et al.* 2015). The NGS technologies have drastically increased the speed at which the DNA sequence can be generated, while reducing the costs and allowed the development of genotyping assays for SNPs (Kage *et al.* 2015).

Numerous SNP genotyping platform are currently available combining a variety of chemistries or allele discrimination techniques (restriction enzyme digestion, sequencing, hybridization with allele-specific probes, single nucleotide primer extension, allele-specific amplification, oligonucleotide ligation), detection methods (colorimetry, spectrometry, fluorescence, fluorescence resonance energy transfer, fluorescent polarization, chemiluminescence), and reaction formats (solution-phase, solid-phase, and gel electrophoresis, next-generation sequencing) (Semagn *et al.* 2014). Moreover new approaches have been developed to use sequencing itself for genotyping for genetic studies with discovery and breeding applications. These approaches are called “genotyping by sequencing” (GBS; Poland and Rife 2012, Elshire *et al.* 2011). In this method, a reduced representation DNA library is developed from the genomic DNA of each sample and sequenced using any one of the NGS platform. Currently, multiplexed chip-based technology is the highest

throughput direct SNP genotyping platform generating till million SNPs per run (Semagn *et al.* 2014).

The most used techniques to detect SNPs rely basically on PCR through allele-specific hybridization, allele-specific nucleotide incorporation or the generation of length polymorphisms. The PCR-based methods to detect SNPs without relying on length polymorphisms include HRM, TaqMan® and KASP™ (Kompetitive Allele Specific PCR).

HRM have been already described as mutation detection method in the TILLING section (1.3.1.2), here is useful to underline that if HRM is used to distinguish between alleles that differ only by a SNP, primers are usually designed to amplify products smaller than about 150 bp, with the polymorphic site near the middle of the product (Dong *et al.* 2009). If the single SNP within the amplicon product is able to modify the melting curve sufficiently to differentiate the alternative alleles the resulting marker may be codominant (with three distinct curves for the heterozygote and the two homozygotes), otherwise the resulting marker may be dominant (with the heterozygote distinguishable from one homozygote but not the other) or it may distinguish the heterozygote from homozygotes but not the homozygotes from each other (Garcia and Mather 2014).

TaqMan® assays use PCR primers in combination with a dual-labelled allele-specific probe. The probe contains a fluorophore at its 5' end and a quencher at its 3' end. When the probe is intact, the quencher is close enough to the fluorophore to reduce the emission of fluorescence by FRET. During PCR, the probe binds specifically to the target site between the primers. During primer extension, the 5' exonuclease activity of the polymerase degrades the probe, releasing the fluorophore from the quencher. Fluorescence is emitted and can be detected with real-time PCR. Probes and primers for Taqman® assays can be designed using software provided by commercial suppliers of the probes. The probes are usually between 20 and 30 bp long, providing specificity, with the target polymorphism as close as possible to the center of the probe. The primers are designed to closely flank the probe target sequence, to amplify a product no longer than 300 bp (Garcia and Mather 2014).

KASP (Kompetitive Allele Specific PCR) is a homogeneous, fluorescence-based genotyping technology. The KASP SNP genotyping system (LGC Genomics, UK), initially developed by KBioscience, is based on allele-specific primer extension and fluorescence resonance energy transfer (FRET) for signal generation. Each assay involves three unlabelled primers: two allele-specific primers and a common primer. The two allele-specific primers are designed with their 3' ends complementary to each of the SNP alleles and with a non-complementary tail in the 5' end. Their tail sequences differ from each other and are complementary to two FRET quenching reporter oligonucleotides (each labelled with a different fluorophore) that are present in the KASP Mastermix. The KASP Master mix (LGC Genomics, UK) is provided in a ready-to-use 2× format

containing universal fluorescent reporting dyes FAM™ and HEX™ as well as Rhodamine X (ROX) dye as the passive reference; the mix is containing the two 5’ fluor-labelled primers with their complementary quencher-labelled primers. The KASP reaction consists of the KASP assay mix (assay specific, composed of the two allele-specific primers together with the common primer) and KASP Mastermix (universal; used with any assay mix) which are combined with the DNA sample to be analysed. In the initial stage of PCR, the appropriate allele-specific primer binds to its complementary region directly upstream of the SNP (with the 3’ end of the primer positioned at the SNP nucleotide) (Figure 1.16). The common reverse primer also binds and PCR proceeds, with the allele-specific primer becoming incorporated into the template. During this phase, the fluor-labelled primers remain bound to their quencher-labelled complementary primers, and no fluorescent signal is generated. As PCR proceeds further, one of the fluor-labelled primers, corresponding to the amplified allele, is also incorporated into the template, and is hence no longer bound to its quencher-labeled complement. As the fluorophore is no longer quenched, the resulting fluorescence can be detected using a plate reader or real-time PCR instrument. If the genotype at a given SNP is homozygous, only one or the other of the possible fluorescent signals will be generated. If the individual is heterozygous, the result will be a mixed fluorescent signal (Garcia and Mather 2014).

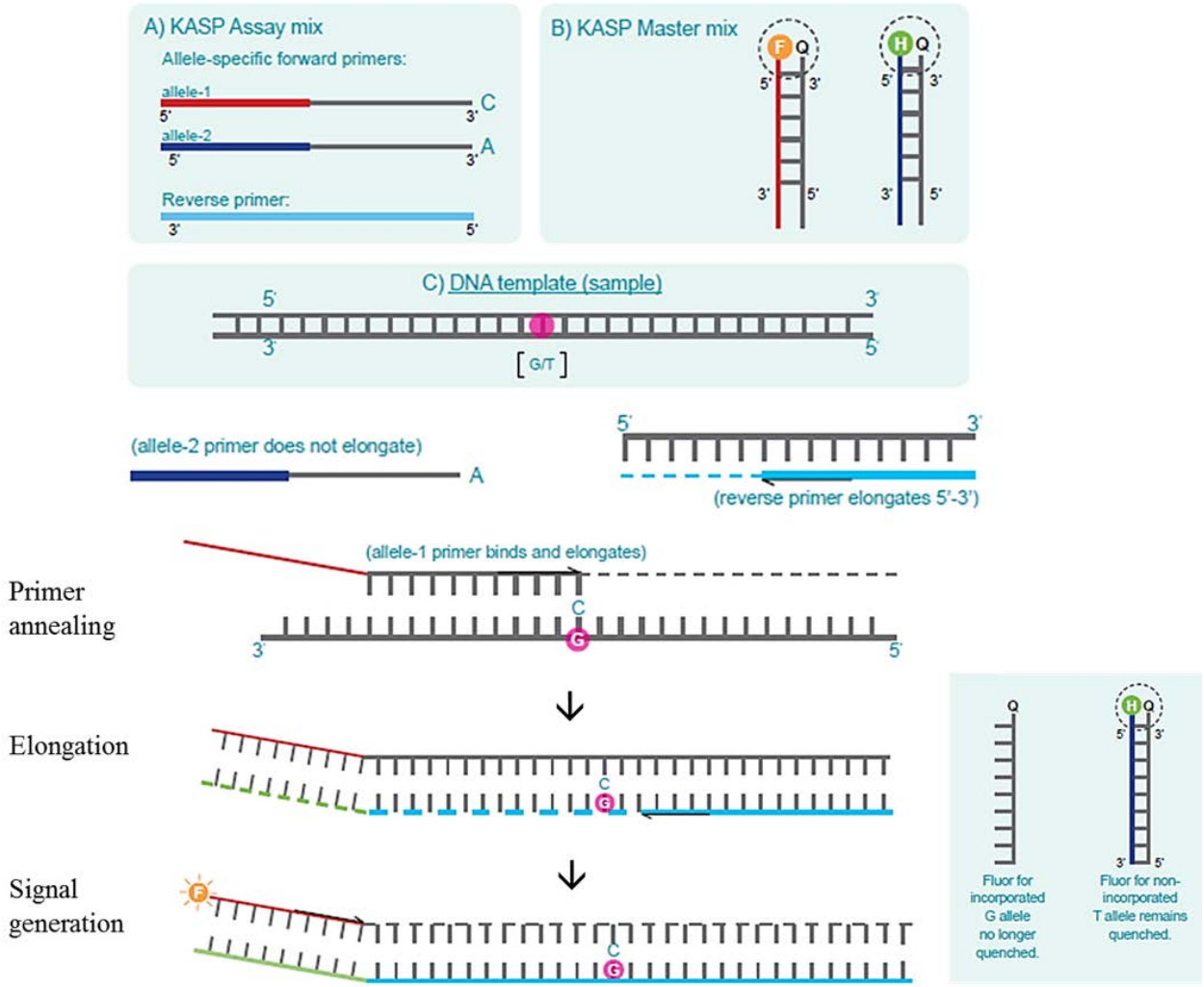



**Figure 1.16** KASP mechanism of action. The KASP assay components are reported in A, B and C and include the KASP Assay mix (A), the KASP Mastermix (B) and the DNA template (C). The KASP Assay mix contains the two allele-specific primers with FAM and HEX complementary tails (colored in red and blue respectively) together with the common primer. The KASP Mastermix contains the two FRET cassette composed of two 5' fluor-labelled primers with their complementary quencher–labelled primers indicated as Q). The orange and green circles indicate the FAM and HEX dye. The pink circle indicate the target SNP (Modified from KASP genotyping chemistry User guide and manual).

### 1.3.4.1 KASP genotyping applied in TILLING

KASP is a convenient technology for the use in single-plex SNP genotyping platform that are suitable for the application requiring a moderate number of SNPs detected for a large number of samples (King *et al.* 2015, Semagn *et al.* 2014). Recently a KASP assay specifically developed for the genotyping of leaf rust resistance locus Lr21 have been successfully used in differentiating resistant and susceptible genotypes over a panel of 384 US wheat lines (Neelam *et al.* 2013).

KASP SNP-based markers can be also used to follow the mutation detected by TILLING throught the back-crossing steps. To make the application of KASP marker easier, a bioinformatics pipeline called PolyMarker (Ramirez-Gonzalez *et al.* 2015, http://polymarker.tgac.ac.uk/) have been recently

implemented to automate the design of the three-primer needed for the KASP assay. PolyMarker generates a multiple alignment between the target SNP sequence and the CSS scaffolds (IWGSC, 2014) for each of the three wheat genomes allowing to develop genome specific primers taking advance of homoeologous SNPs with respect to the target genome. The common primer is selected to incorporate a homoeologue-specific or semi-specific base at the 3' end, whereas the competing diagnostic primers are designed to incorporate the alternative varietal bases at their 3' ends (Ramirez-Gonzalez *et al.* 2014). This tool allows to distinguish between heterozygote and homozygote individuals in polyploid species and is therefore a powerful tool for TILLING in wheat. The pipeline has been successfully applied to validate putative SNPs from RNAseq data and successfully converts these SNPs into codominant genome-specific KASP assays that allows to cope with the difficulties of the polyploid wheat genome and have been used for diagnostics across UK pedigree programmes (Ramirez-Gonzalez *et al.* 2015, Ramirez-Gonzalez *et al.* 2014).

# Objectives

Durum wheat is one of the most important human crops, since it accounts for 5% of the total wheat production with a world-cultivated area of 18 Mha. Wheat growth can be impaired by heat stress at any developmental stage, and modelling scenarios predict even warmer temperatures in the future.
Despite advances in our knowledge about the genes involved in heat adaptation in wheat, the genetic bases of thermotolerance are, to date, poorly understood. New insights in the mechanisms that take place during the heat stress in durum wheat are highly needed to meet the increasing of food demand in the actual climate change scenario.
The overall objective of this work was to identify new alleles of the chloroplast *sHsp26* coding genes to investigate the role of HSP26 in the wheat adaptation to heat stress and also to identify new resources for wheat breeding.

The targets of the present work can be summarized in three principal pillars:

- The isolation and the genomic characterization of the genes coding for the chloroplastic sHSP26 in *Triticum durum.*
- The identification of mutations in the genes of interest in two durum wheat TILLING populations using a conventional HRM mutation detection coupled with an online TILLING database search.
- The development of new genetic KASP markers to use for the *in vivo* confirmation of the identified mutation, for the genotyping during the crosses and the backcrosses and suitable also for marker assisted selection (MAS) in future breeding programs.

# Chapter 2. Identification and characterization of the *T. durum* chloroplastic *sHsps* genes (*TdHsp26*) and promoters

## 2.1 Introduction

### 2.1.1 The role of sHSP26s in thermotolerance

The involvement of the chloroplastic localized sHSP26s in the heat stress response has been shown in many plant species (Hu *et al.* 2015, Ul Haq *et al.* 2013, Zhang *et al.* 2014). The sHSP26s show a dynamical distribution, changing their partitions between the soluble and insoluble fractions of the chloroplast with an accumulation at the insoluble state after heat stress; they have been show to physically interactwith the Photosystem II (PSII) which is considered the most thermo-sensitive structure in photosynthesis (Xia *et al.* 2015, Rawland *et al.* 2010, Osteryoung and Vierling 1994). Recent studies of Chauhan and co-workers (2011) demonstrated that plastid sHSP26 is highly induced in wheat tissues under heat stress and that the constitutive expression of *TaHsp26* in *Arabidopsis* increases the PSII thermotolerance under sub-lethal continuous heat stress conditions. Transgenic *Arabidopsis* plants overexpressing the wheat gene showed a higher biomass and a higher active photochemical efficiency under heat stress than control plants (Chauhan *et al.* 2012, Khurana *et al.* 2013). The transcripts for plastidial *sHsp26s* were highly expressed, after sub-lethal temperature exposure (acclimation), in thermotolerant cultivars of wheat which were able to resist at subsequent extreme heat stress (Rampino *et al.* 2009b). These observations highlight the important role of sHSP26 in the thermotolerance acquisition.

### 2.1.2 Known *sHsp26* in wheat

To date, several cDNA and protein sequences for sHSPs have been isolated in *T. aestivum* (Nguyen *et al.* 1993, Weng *et al.* 1991, McElwain and Spiker 1989, Pandey *et al.* 2015, Basha *et al.* 1999, Rampino *et al.* 2009a). cDNA sequences of both *T. durum* and *T. aestivum sHsp26* (so far called only *Hsp26*) have been identified (Rampino *et al.* 2009b, Rampino *et al.* 2012, Nguyen *et al.* 1993, Weng *et al.* 1991, Table 2.1); however scarce information is available on the genomic structure of *Hsp26* genes in polyploid wheat (Chauhan *et al.* 2011, Campbell 1998). A survey of the *Triticum Hsp*26 cDNA and protein sequence retrieved in literature is presented in Table 2.1.

**Table 2.1.** Survey of *Triticum sHsp26* cDNA and protein sequence.
[1] Name, Nucleotide Accession Number, Protein Accession Number, Organism, Variety and Reference attributed to the nucleotide or protein sequence on NCBI.
[2] Mw of the protein calculated with DNAMAN software.

| Name[1] | Nucleotide Accession Number[1] | Protein[1] | Mw (Da)[2] | Variety[1] | Organism[1] | Reference[1] |
|---|---|---|---|---|---|---|
| *TdHsp26.5* | AJ971373 | CAI96515 | 26,489 | Creso | *T. turgidum* subsp. *durum* | Rampino *et al*, 2009 |
| *TaHsp26.6g* | AF097657 | AAC96315 | 26,844 | Mustang | *T. aestivum* | Campbell 1998 |
| *TaHsp26.6e* | AF097656 | AAC96314 | 26,469 | Mustang | *T. aestivum* | Campbell 1998 |
| *TaHsp26.6i* | AF097658 | AAC96316 | 26,583 | Mustang | *T. aestivum* | Campbell 1998 |
| *TaHsp26.6m* | AF097659 | AAC96317 | 26,560 | Mustang | *T. aestivum* | Campbell 1998 |
| *TaHsp26.6* | X58280 | CAA41219 | 26,566 | Mustang | *T. aestivum* | Weng *et al*, 1991 |
| *TasHsp* | HM802264 | ADN97108 | 26,844 | C306 | *T. aestivum* | Kumar *et al*, unpublished |
| *TaHsp26.6B* | X67328 | CAA47745 | 26,557 | Mustang | *T. aestivum* | Nguyen *et al*, 1993 |
| *TdiHsp26.4* | AJ971370 | CAI96512 | 26,417 | MG29896/212 | *T. turgidum* subsp. *dicoccoides* | Rampino *et al*, unpublished |
| *TdicHsp26.8* | AJ971372 | CAI96514 | 26,734 | MG5473/295 | *T. turgidum* subsp. *dicoccum* | Rampino *et al*, unpublished |
| *TdiHsp26.6* | AJ971371 | CAI96513 | 26,554 | MG29896/212 | *T. turgidum* subsp. *dicoccoides* | Rampino *et al*, unpublished |
| *TmHsp26.6* | AJ971374 | CAI96516 | 26,564 | ID362 | *T. monococcum* | Rampino *et al*, unpublished |
| *TmHsp26.5* | AJ971375 | CAI96517 | 26,547 | ID529 | *T. monococcum* | Rampino *et al*, unpublished |

## 2.1.3 Wheat *sHsp26* promoter

The promoter sequence (1,514bp) of *TasHsp26-g* allele (AF097657) was recently cloned by Chauhan and co-workers (2012) and analysed for the presence of *cis*-acting element using the PLACE database (Chauhan *et al.* 2012). The motifs found in the *TaHsp26* promoter are reported in Table 2.2. The motifs were mainly related to heat stress response: three Heat Shock Elements (HSEs) of which one perfect-type HSE (HSE-P), one gap-type HSE (HSE-G) and one step-type HSE (HSE-S); many other abiotic stress-related motifs as CBFHV (also known as Dehydration-Responsive Element, DRE) which is crucial for drought stress, STRE (Stress-Responsive Element), GT1 binding site, MYB binding site, MYC binding site (MYCATRD22 found also in *Arabidopsis* dehydration-responsive gene); some developmental stage specific elements as GCN4 motif, ABRELATERD1 and POLLEN1LELAT52. Other *cis*-acting elements found in the *TaHsp26* promoter were TATAbox, I-box and GATA motifs; the last two are conserved sequences present upstream of light regulated genes that are required for light regulated tissue-specific expression. Light regulation at the transcriptional level has been already demonstrated in chloroplast targeted proteins (Jung and Chory 2010) and it has also been proposed in *Arabidopsis* that HSPs induction is primarily not because of elevated temperatures but rather because of oxidizing environment of high light (Rossel *et al.* 2002, Chauhan *et al.* 2012).

**Table 2.2** Description of *cis*-acting elements found by Chauhan and co-workers in the promoter of *TaHsp26* using PLACE database (adjusted from Chauhan *et al.* 2012).

| Motif name | Sequence | Descriptipn |
|---|---|---|
| 2SSEEDPROTBANAPA | CAAACAC | Conserved in many storage-protein gene promoters (Stalberg *et al.* 1996) |
| ABRELATERD1 | ACGTG | ABRE-like sequence required for etiolation-induced expression of erd1 (early responsive to dehydration) in *Arabidopsis* (Nakashima *et al.* 2006) |
| CCAATBOX1 | CCAAT | Common sequence found in the 5′-non-coding regions of eukaryotic genes; 'CCAAT box' found in the promoter of heat shock protein genes; 'CCAAT box' act cooperatively with HSEs to increase the hs promoter activity; (Rieping & Schoffl 1992; Haralampidis *et al.* 2002) |
| CBFHV | RYCGAC | Binding site of barley CBF1, and also of barley CBF2; CBF = C-repeat (CRT) binding factors; CBFs are also known as dehydration-responsive element(DRE) binding proteins (DREBs) (Svensson *et al.* 2006) |
| GCN4OSGLUB1 | TGAGTCA | 'GCN4 motif' found in GluB-1 gene in rice; required for endosperm-specific expression (Onodera *et al.* 2001) |
| GT1GMSCAM4 | GAAAAA | 'GT-1 motif' found in the promoter of soybean CaM isoform, SCaM-4; Plays a role in pathogen- and salt-induced SCaM-4 gene expression (Park *et al.* 2004) |
| MYB2AT | TAACTG | ATMYB2 is involved in regulation of genes that are responsive to water stress in *Arabidopsis* (Urao *et al.* 1993) |
| MYCATRD22 | CACATG | MYC binding site in rd22 gene of *Arabidopsis*, ABA-induction (Abe *et al.* 1997) |
| POLLEN1LELAT52 | AGAAA | Responsible for pollen-specific activation (Bate & Twell 1998) |
| STRE | AGGGG | Stress-responsive elements (Martinez-Pastor *et al.* 1996) |
| HSE | nGAAnnTTCnnGAAn | Heat shock element found in the promoter of heat shock promoter genes (Nover *et al.* 2001; Yamamoto *et al.* 2005) |

The promoter activity was also characterized through β-glucoronidase (GUS) reporter system in transgenic rice and *Arabidopsis* plants where it has shown to be highly responsive to heat and other abiotic stresses (such as cold, simulated salinity and simulated drought). This can be due to the presence of all the three types of HSEs (HSE-P, HSE-G, HSE-S) and other heat responsive elements. Transcript accumulation pattern and promoter analysis showed also that HSP26 may have a role during seed germination and seed development in control and stressed plants; this can be attributed to the presence of multiple pollen specific *cis*-acting elements in the promoter of *TaHsp26* (Chauhan *et al.* 2012).

Khurana *et al.* (2013) further characterized the role of *cis*-acting elements in the *TaHsp26* promoter by deletion constructs and GUS induction in *Arabidopsis* in response to heat stress, concluding that HSEs alone are not sufficient for heat-shock induction and that a synergic effects of CCAAT box elements and HSE is required for the induced heat defence responses.

# 2.2 Material and methods

## 2.2.1 Vegetal material and growth conditions

The durum wheat cv. Cham1 seeds, kindly provided by Prof. Mario Pagnotta (University of Tuscia, Viterbo, IT), were germinated at room temperature in dark conditions and ten-days seedling leaves were sampled for DNA extraction.

The durum wheat cv. Kronos DNA was kindly provided by Dr. Cristobal Uauy (JIC, Norwich, UK).

The DNA of nulli tetrasomic lines N4AT4D, N4DT4B and of the ditelosomic lines DT4AS, DT4AL, DT4BS, DT4DS, DT4DL of the bread wheat cultivar Chinese Spring were kindly provided by Prof. Renato D'Ovidio (University of Tuscia, Viterbo, IT) and Prof. Antonio Blanco (University of Bari, Bari, IT).

## 2.2.2 Database research for wheat *sHsp26* gene sequence identification

The public NCBI database was used to identify the annotated *Triticum Hsp26* cDNA products. The *T. durum sHsp26.5* mRNA (AJ971373) was used as query for the BLASTN analysis on the IWGSC assembly performed through the EnsemblPlants IWGSC 1.0+popseq (release 29) for the *in silico* characterization of the putative *TaHsp26* genes. All the matches including both chromosome pseudomolecules and scaffold located in the A and B genome with a nucleotide identity greater than 89% and a threshold E-value of less than 2.5E-10 were considered.

The retrieved matches with the IWGSC contigs and the genomic CSS scaffolds carrying putative *TaHsp26* genes were analysed and used for primer design to isolate the orthologous *TdHsp26* genes.

## 2.2.3 *TdHsp26* genes isolation and chromosome localization

### 2.2.3.1 Primers design

Homoelogous-specific primer pairs for the isolation of the *TdHsp26* complete gene sequences and promoters were manually designed on the basis of the nucleotide sequence of the retrieved IWGSC contigs and checked for the suitable characteristics (primer length, melting temperature, secondary structure, GC contents) by Primer3 v. 0.4.0 (Untergasser *et al.* 2012). The primers pairs were designed on scaffold-specific polymorphisms to isolate the putative different *TdHsp26* genes and were ordered from Sigma-Aldrich (Gillingham, UK). The full list of the primer pairs sequences used are reported in Table 2.3. The B1-5F primer was designed in a conserved region between all

the retrieved CSS scaffold sequence because the match of the query with the scaffold IWGSC_CSS_4BL_6982788 was not complete. The 12F primer used for the isolation of the promoters was designed on the *TaHsp26* promoter sequence previously isolated by Chauhan *et al.* 2012.

**Table 2.3** Primer pairs used for the *TdHsp26* genes isolation. For each primer the IWGSC scaffold or contig sequence used for its design and the position of the primers in respect to the ATG of the matched CDS are indicated.

| Primer name | Type | Designed on | Primer position (bp to ATG) | Sequence 5'-3' | Amplicon size (bp) | Application |
|---|---|---|---|---|---|---|
| **A1-9F** | Forward | IWGSC_chr4AS_contig_5940868 | -326 | TGTTGGGCCTCCTGATCG | 1,171 | *TdHsp26-A1* isolation and chromosomal localization |
| **A1-4R** | Reverse | IWGSC_chr4AS_contig_5940868 | +845 | AGCCTCAGATGCAGGGTAC | | |
| **A2-24F** | Forward | IWGSC_chr4AS_contig_5995647 | -126 | CCACCAGACAATCACTGCAA | 939 | *TdHsp26-A2* isolation and chromosomal localization |
| **A2-18R** | Reverse | IWGSC_chr4AS_contig_5995647 | +813 | CAGGGTACAGTCTCACACG | | |
| **A3-20F** | Forward | IWGSC_chr4AS_contig_5936815 | -383 | GGCGAAGATCTGCCAAAGTAT | 1,106 | *TdHsp26-A3* isolation and chromosomal localization |
| **A3-29R** | Reverse | IWGSC_chr4AS_contig_5936815 | +723 | AACCAGCACAACCCTCTA | | |
| **B1-5F** | Forward | All the retrieved IWGSC contigs | -63 | GACACTCTCTCGTTTCAATTCTC | 1,182 | *TdHsp26-B1* isolation and chromosomal localization |
| **B1-6R** | Reverse | IWGSC_chr4BL_contig_6982788 | +1119 | GTTATCAGCTTCTTCCGGG | | |
| **12F** | Forward | Chauhan et al 2012 promoter sequence | -1437 | CAATTGGTTCGCACAAACAC | 1,120 | *TdHsp26-A1* promoter isolation |
| **A1-11R** | Reverse | IWGSC_chr4AS_contig_5940868 | -227 | AGTACCTGGTTGGAGAGGCC | | |
| **12F** | Forward | Chauhan et al 2012 promoter sequence | -1437 | CAATTGGTTCGCACAAACAC | 1,268 | *TdHsp26-B1* promoter isolation |
| **B1-PT8R** | Reverse | IWGSC_chr4BL_contig_6982788 | +79 | CCCTCCAGGCACGGATG | | |

### 2.2.3.2 Polymerase Chain Reaction (PCR)

The *TdHsp26-A1, TdHsp26-A2, TdHsp26-A3,* and *TdHsp26-B1* gene sequences were isolated both from Cham1 and Kronos by PCR on Veriti Thermo Cycler (Life Technologies, Paisley, UK) using GoTaq Long PCR 2X Master Mix (Promega, WI, USA) in 30 µl reaction volume with 10-20 ng of Cham1 and Kronos DNA and 0.4 µM of each specific primer. Thermal cycle conditions were: initial denaturation at 94 °C for 2 mins, followed by 40 cycles of 94 °C for 15 s, 60 °C for 30 s and 72°C for 90 s, followed by 72 °C for 10 mins terminal extension.

The same PCR condition, but with an annealing temperature of 64 °C, were used to obtain the DNA sequences located upstream of the gene coding regions *TdHsp26-A1*and *TdHsp26-B1* for the promoters isolation.

Chromosomal localization was determined on nulli tetrasomic and ditelosomic lines by using the same homoelogous-specific primers shown in Table 2.3. The *Triticum aestivum Actin* gene (AB181991) was used as internal control with the primers ACT-Fw 5'-TCCTGTGTTGCTGACTGAGG and ACT-Rev 5'- GGTCCAAACGAAGGATAGCA (Rocchi *et al.* 2012).

## 2.2.4 Nucleic acid extraction and analysis

The genomic DNA of cv. Cham 1 used for the genes isolation was extracted by leaves with the procedure of Tai and Tanksley (1991). The DNA amplicons generated by PCR were separated and visualized by horizontal gel electrophoresis on Tris Acetate EDTA (TAE) Agarose gel. The

amplicons were purified with NZYGelpure kit (NZYtech, Lisbon, Portugal) and sequenced by using the commercial service Eurofins MWG (Eurofins Genomics, Edersberg, Germany).

### 2.2.5 Bioinformatics

BLAST analysis for *TdHsp26* genes identification were performed on the IWGSC database (IWGSC 2014) through the URGI and EnsemblPlants interfaces and on the WGS w7984_Meraculous Scaffolds (Chapman *et al.* 2015) databases through the CerealsDB website. The TREP (the Triticeae Repeat Sequence Database) was also used. Sequence analysis, multiple alignment, Neighbour Joining tree and homology matrix were performed by using DNAMAN software (Lynnon Biosoft, Quebec, Canada). Structure and exons/introns position were determine on the basis of FGENESH and FSPLICE tools of the Softberry suites (Softberry Inc., NY, USA) for gene finding in *Eucaryota*.

The structural protein alignment was carried out with PROMALS3D. The protein localization was predicted by TargetP 1.1 Server and ChloroP 1.1, a neural network based method for identifying chloroplast transit peptides and their cleavage sites (Emanuelsson *et al.* 1999; Emanuelsson *et al.* 2000). The promoter analysis was performed with PlantCARE (Cis-Acting Regulatory Element) database (Lescot *et al.* 2002) and PLACE (Plant Cis-acting Regulatory DNA Element) database (Higo *et al.* 1999).

# 2.3 Results and discussion

## 2.3.1 Identification of wheat *Hsp26* genes

To date the durum wheat genome sequence is not available, even if many effort are done by the scientific community. For this reason to identify *Hsp26* genes in durum wheat, we started the work by searching, in the bread wheat cv. Chinese Spring genome on IWGSC, the only *TdHsp26* cDNA sequence available in NCBI (AJ971373; Rampino *et al.* 2009b). The resulted sequences have been used as reference for generic primers development to retrieve the orthologous sequences in durum wheat by following a PCR-sequence- based approach. The *TdHsp26.5* (AJ971373) was used as query sequence in a BLASTN analysis on EnsemblPlants. Top-scoring matches were assembled on the basis of the chromosome position, scaffold and transcript annotation (Traes) in four group as indicated in Table 2.4 and analysed to reconstruct the putative *TaHsp26* gene sequences.

The assembly of the retrieved CSS scaffolds allowed the identification of four putative *TaHsp26* genes sequences named: *TaHsp26-A1*, *TaHsp26-A2*, *TaHsp26-A3* and *TaHsp26-B1*. According to the retrieved matches the sequence have been putative localized on the short arm of chromosome 4A for *-A1*, *-A2* and *-A3* and on the long arm of chromosome 4B (*-B1*) of the bread wheat cultivar Chinese Spring (Table2.4).

**Table 2.4** *In silico* analysis on EnsemblPlants database and assembly strategy to isolate *Hsp26* genes in durum wheat. The *TdHsp26.5* mRNA sequence (AJ971373) was used as query sequence in the EnsemblPlants BLAST search. Four groups of matches have been assembled to reconstruct four putative *TaHsp26* genes and these have been used as reference for the isolation of four *TdHsp26* genes. The corresponding IWGSC contig retrieved through URGI Seq Repository are also reported.

| IWGSC+popseq(1.0) matches | Transcript on Ensembl Plant and UniProt attribution | Query start | Query end | URGI IWGSC contigs | Putative gene assembly | Correspondent isolated gene | Chromosomal localization |
|---|---|---|---|---|---|---|---|
| 4A:48083230-48083489 | Traes_4AS_8BA1E69CA, Uniprot W5DYH1 | 1 | 260 | IWGSC_chr4AS_ contig_5940868 | *TaHsp26-A1* | *TdHsp26-A1* | 4AS |
| 4A:48082683-48083141 | Traes_4AS_8BA1E69CA, Uniprot W5DYH1 | 259 | 717 | | | | |
| 4A:61082678-61082753 | Traes_4AS_2272D0413, Uniprot W5DWE6 | 29 | 104 | IWGSC_chr4AS_ contig_5995647 | *TaHsp26-A2* | *TdHsp26-A2* | 4AS |
| 4A:61082602-61082677 | Traes_4AS_2272D0413, Uniprot W5DWE6 | 102 | 177 | | | | |
| 4A:61082522-61082586 | Traes_4AS_2272D0413, Uniprot W5DWE6 | 196 | 260 | | | | |
| 4A:61081978-61082436 | Traes_4AS_2272D0413, Uniprot W5DWE6 | 259 | 717 | | | | |
| IWGSC_CSS_4AS_scaff_5936815:1264-1339 | Traes_4AS_049E43B8B, Uniprot W5DVUE | 29 | 104 | IWGSC_chr4AS_ contig_5936815 | *TaHsp26-A3 (putative pseudogene)* | *TdHsp26-A3 (putative pseudogene)* | 4AS |
| IWGSC_CSS_4AS_scaff_5936815:1105-1263 | Traes_4AS_049E43B8B, Uniprot W5DVUE | 102 | 260 | | | | |
| IWGSC_CSS_4AS_scaff_5936815:754-1008 | Traes_4AS_049E43B8B, Uniprot W5DVUE | 259 | 513 | | | | |
| IWGSC_CSS_4BL_scaff_6982788:1-102 | Traes_4BL_3C1C91A9C, Uniprot W5E2Q1 | 159 | 260 | IWGSC_chr4BL_ contig_6982788 | *TaHsp26-B1* | *TdHsp26-B1* | 4BL |
| IWGSC_CSS_4BL_scaff_6982788:199-403 | Traes_4BL_3C1C91A9C, Uniprot W5E2Q1 | 259 | 463 | | | | |
| IWGSC_CSS_4BL_scaff_6982788:475-651 | Traes_4BL_3C1C91A9C, Uniprot W5E2Q1 | 541 | 717 | | | | |

*TaHsp26-A1* and *TaHsp26-A2* are localized on a chromosome pseudomolecules on the short arm of chromosome 4A spanning a region of 13,000,070 bp: *TaHsp26-A1* gene is contained in a region ranging from 48,082,683 bp to 48,083,489 bp, while *TaHsp26-A2* in a region from 61,081,978 bp to 61,082,753 bp on the chromosome. Moreover, the two genes have been retrieved on two WGS Meracolous Scaffolds (4146766 and 2124046, respectively), which map on the same bin at 57.601

cM on Chr 4AS based on the POPSEQ data (Mascher *et al*, 2013). These observations allow us to conclude that *TaHsp26-A1* and *TaHsp26-A2* could be members of a *sHsp26* gene cluster located on the short arm of chromosome 4A (Table 2.4). The presence of a putative *Hsp26* gene cluster suggests that other members of the same family with different sequence features, classes and functions may be present in the same cluster as previously demonstrated in tomato (Goyal *et al.* 2012). Differently, for *TaHsp26-B1* and *TaHsp26-A3,* it was not possible to clearly identify a predicted position on the 4A/4B chromosome due to the absence of the IWGSC scaffolds in the POPSEQ map (Table 2.4). Only very short matches (<35 nt) have been retrieved in EnsemblPlants database using the *TdHsp26*.5 cDNA as query on the other wheat chromosomes; they were not considered due to their low E-value.

## 2.3.2 Isolation and sequence analysis of *TdHsp26* genes

On the basis of the matching sequences retrieved in the databases, different sets of primers were designed to identify the corresponding orthologues in durum wheat (Table 2.3; Table 2.4). Genomic DNA from durum wheat cultivars Cham1 and Kronos was used as template in PCR reaction. From the sequence of the obtained amplicons, four putative orthologoues genes were identified and named *TdHsp26-A1, TdHsp26-A2, TdHsp26-A3*, *TdHsp26-B1* respectively. The genomic alignment of the four genes of Cham1 and Kronos is reported in Figure 2.1; *TdHsp26.5* cDNA sequence (AJ971373) was included in the alignment to better identify the splice junctions. The complete gene sequences of *TdHsp26-A1* (807 bp), *TdHsp26-A2* (8*07 bp*), *TdHsp26-A3* (*623 bp)* and *TdHsp26-B1* (*830 bp)* contained an open reading frame of 717 bp, 720 bp, 525 bp and 732 bp respectively (Figure 2.2A).

The alignments with the *TdHsp26.5* cDNA sequence and the *in silico* prediction of the splicing site with FGENESH and FSPLICE programs (Solovyev *et al.* 2006) revealed that *TdHsp26-A1, -A2, -A3* and *-B1* are characterized by the presence of two exons (Figure 2.1 and Figure 2.2 A,B). The gene structure of *TdHsp26-A1, A2* and *B1* is most conserved with slight differences in exons and intron lengths (Figure 2.2A, B). The *TdHsp26*-A3 sequence shows the presence of a premature termination codon (TAG) within the second exon (623 bp from the start codon) and tcode for a truncated protein of 174 amino acid residues (Figure 2.1 and 2.2A, B) which is missing over half of the α-crystalline domain (ACD). Due to the presence of an in frame stop codon and the interruption of the functionally important ACD domain the *TdHsp26*-A3 sequence was then classified as a putative pseudogene, even if the absence of transcription hasn't been verified yet (Balakirev and Ayala 2003). Association between retrotransposon and HSPs was previously shown in *Arabidopsis*

(Visioli *et al.* 1997). Due to the classification of *TdHsp26-A3* as a pseudogene it was not considered for further analysis.

The homologous genes between the cultivars Cham1 and Kronos shown, as expected, the same gene structures and high sequence conservations (Figure 2.1). The sequence analysis revealed a number of SNPs most likely related to a varietal variation. In particular *TdHsp26-A1* from Cham1 and *TdHsp26-A1* from Kronos share 99.75% identity, with only two SNPs in the first and second exon (C217G and A677G) which leads to a single amino acid substitution (Q73E); the *TdHsp26-A2* gene sequences showed a single SNP in the second exon (G607A) leading to the amino acid substitution E174K and share 99.88% identity at nucleotide level between Cham1 and Kronos; whereas the *TdHsp26-B1* gene shows a perfect identity between the two cultivars. Moreover, as expected most of the variations between the different genes are concentrated on the intronic region.

*TdHsp26-A1* shares 99.72% sequence identity with the *TdHsp26.5* (AJ971373, Rampino *et al.* 2009b) that was used as query sequence in the database search. *TdHsp26-A2* shares 99.72% sequence identity with the Traes 4AS_2272D0413 sequence retrieved in EnsemblPlants and shares with the respective CDS of *T. dicoccoides TdiHSP26.4* (AJ971375) a sequence identity of 99.44%. *TdHsp26-B1* shares 100% identity with the *TaHsp26.6-g* allele (AF097657) previously described by Campbell 1998. *TdHsp26-B1* sequence presents a 9 bp insertion coding for an insertion of 3 Amino Acids (AA) in the ACD domain in respect to the other *TdHsp26* genes (Figure 2.1 and 2.2A), as already noted by Campbell (1998) for the *TaHsp26-g* clone.

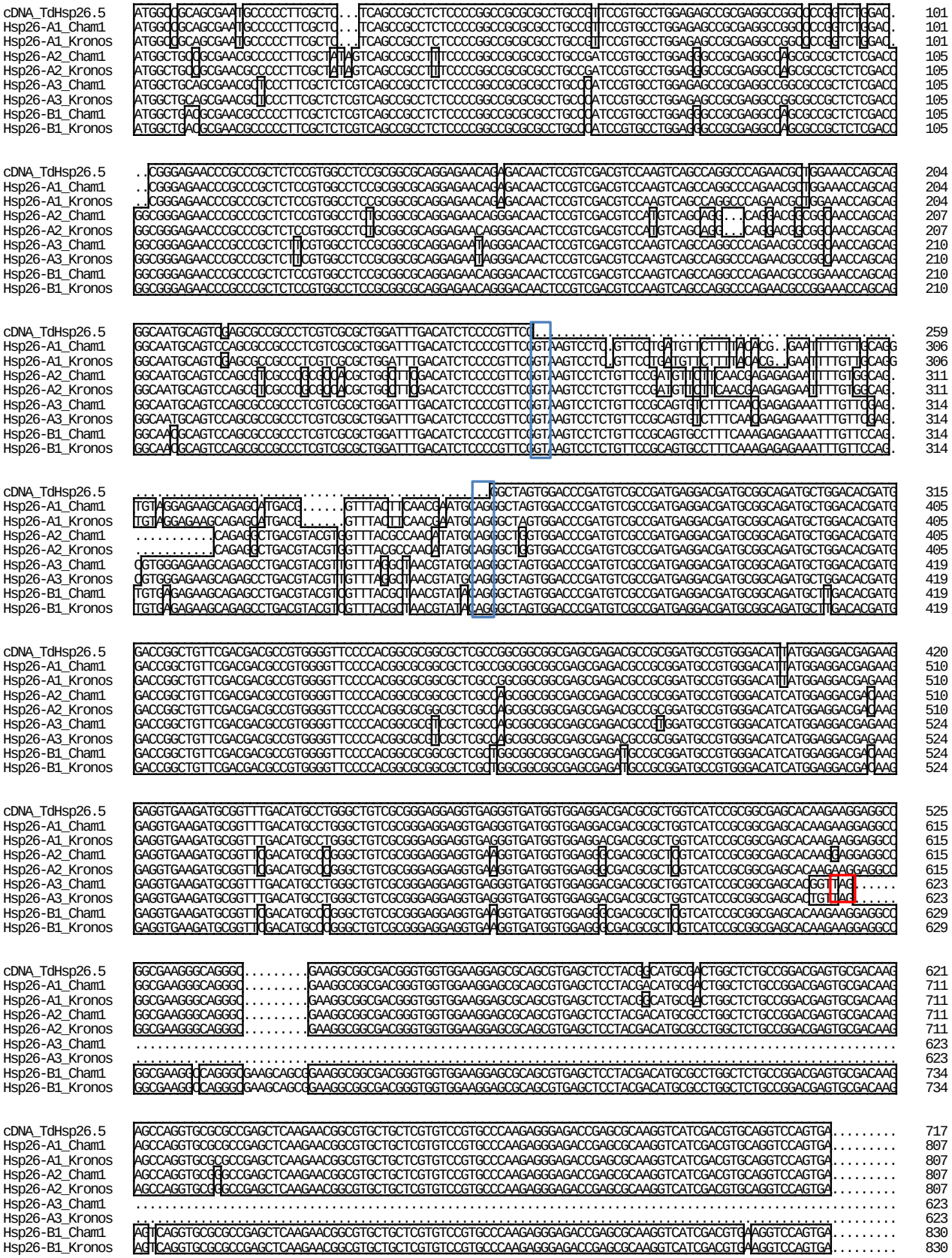



**Figure 2.1** Alignment of the complete genomic sequences of *TdHsp26-A1*, *TdHsp26-A2*, *TdHsp26-A3* and *TdHsp26-B1* with the *TdHsp26.5* cDNA sequence (AJ971373). The conserved blocks on the alignment are boxed in black. The conserved splicing site (GT splice acceptor site and AG splice donor site) are boxed in blue and the premature *TdHsp26*-A3 stop codon TAG is boxed in red.

**A)**

| Sequence | Exon1 | Exon2 | Intron Lenght | Lenght gene sequence | Lenght aminoacidic sequence | Protein weight |
|---|---|---|---|---|---|---|
| *TdHsp26-A1* | 1-259 | 350-807 | 90 bp | 807 bp | 238 AA | 26,528 Da |
| *TdHsp26-A2* | 1-262 | 350-807 | 87 bp | 807 bp | 239 AA | 26,400 Da |
| *TdHsp26-A3* | 1-265 | 364-623 | 98 bp | 623 bp | 174 AA | 19,222 Da |
| *TdHsp26-B1* | 1-265 | 364-830 | 98 pb | 830 bp | 243 AA | 26,826 Da |

**B)**

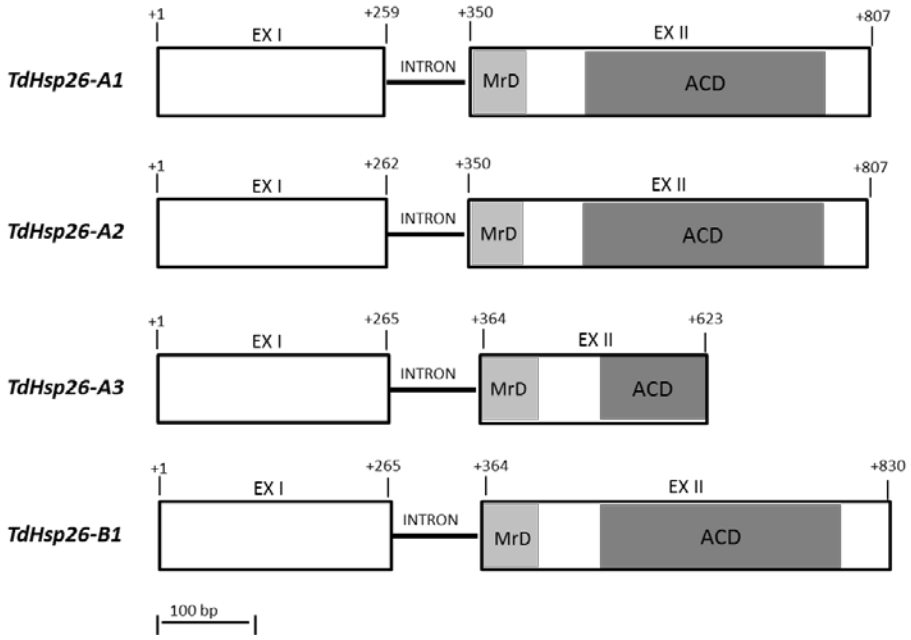



**Figure 2.2** Gene structure of *TdHsp26-A1*, *TdHsp26-A2*, *TdHsp26-A3*, *TdHsp26-B1*. A) Details on the structure composition of *TdHSp26-A1*, *TdHsp26-A2*, *TdHsp26-A3*, *TdHsp26-B1*. The introns and exons position, length of amino acid sequence and predicted molecular weight are reported. B) Gene structure of each member of the durum wheat *Hsp*26 gene family. Start and ending points of exons and introns are indicated. On the basis of literature data (Chauhan *et al.* 2012, Scharf *et al.* 2001, Waters 2013) the conserved HSP26 domains of amphipathic alpha helix rich in Methionine (Met-rich Domain, MrD) and the alpha-crystalline domain (ACD) are indicated (highlighted in bright and dark grey respectively).

## 2.3.3 *TdHsp26* chromosome localization

The chromosomal localization predicted by the anchored scaffolds in the EnsemblPlants database for *TdHsp26-A1* (Chr 4AS), *TdHsp26-A2* (Chr 4AS), *TdHsp26-A3* (Chr 4AS), and *TdHsp26-B1* (Chr 4BL) was confirmed by PCR including as template the Chinese Spring nulli tetrasomic and the ditelosomic lines (Sears and Sears 1978; Sears 1966) for chromosome 4. Nullisomic–tetrasomic (NT) are wheat lines each lacking a different pair of homologous chromosomes, due to the replacement with its homoeologous (i.e. N4AT4D means a line in which the 4A pair is missing and replaced by an additional 4D pair). Ditelosomics (DT) are euploid lines in which one arm of a given chromosome is missing, for example line DT4AL contains only the long arm of chromosome 4A and lacks its short arm. Due to their characteristics, these lines can be used in wheat to map a gene on a specific chromosome arm (Gullì *et al.* 2007). As shown in Figure 2.3, using *TdHsp26-A1*, *-A2* and *-A3* specific primers, no amplification products were obtained for the N4AT4D and DT4AL lines (sample 1 and 4, respectively in Figure 2.3). Using *TdHsp26-B1* specific primers, no amplification products were obtained for the DT4BS line (sample 5 in Figure 2.3).

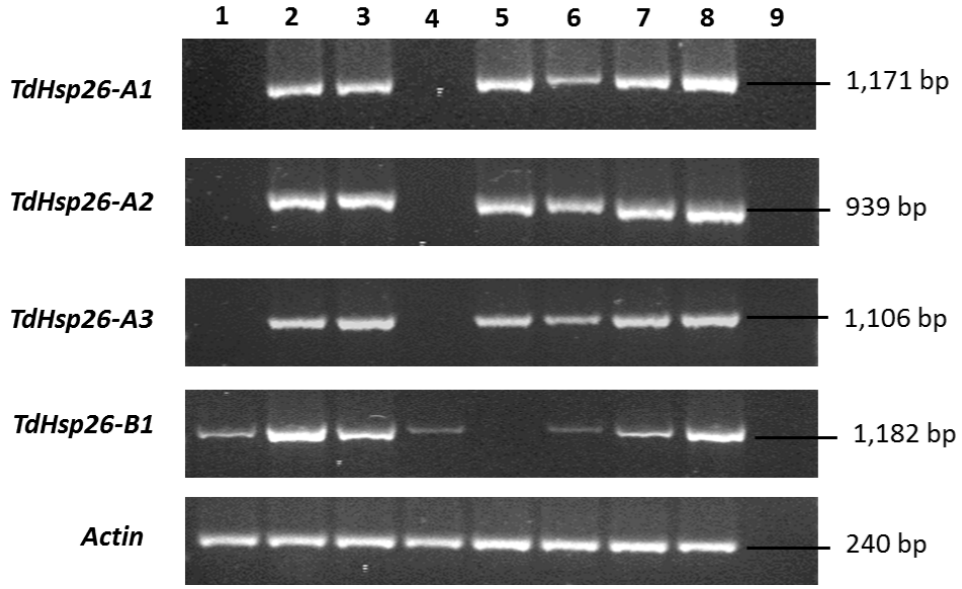



**Figure 2.3** Chromosomal localization of *TdHsp26-A1*, *TdHsp26-A2*, *TdHsp26-A3*, *TdHsp26-B1* and Actin (AB181991). 1) N4AT4D, 2) N4DT4B, 3) DT4AS, 4) DT4AL, 5) DT4BS, 6) DT4DS, 7) DT4DL, 8) Cham1, 9) Negative control. The following primers were used: A1-9F/A1-4R, A2-24F/A2-18R, A3-20F/A3-29R, B1-5F/B1-6R, TaACT-Fw/TaACT-Rev (Table2.3).

The *TdHsp26-B1* gene detected on the 4BL chromosome could be considered as homoeologous to the *TdHsp26-A1* or *-A2* gene sequences localized on the 4AS. This is consistent with Miftahudin *et al.* 2004 reporting that a pericentric inversion within chromosome 4A changed the native

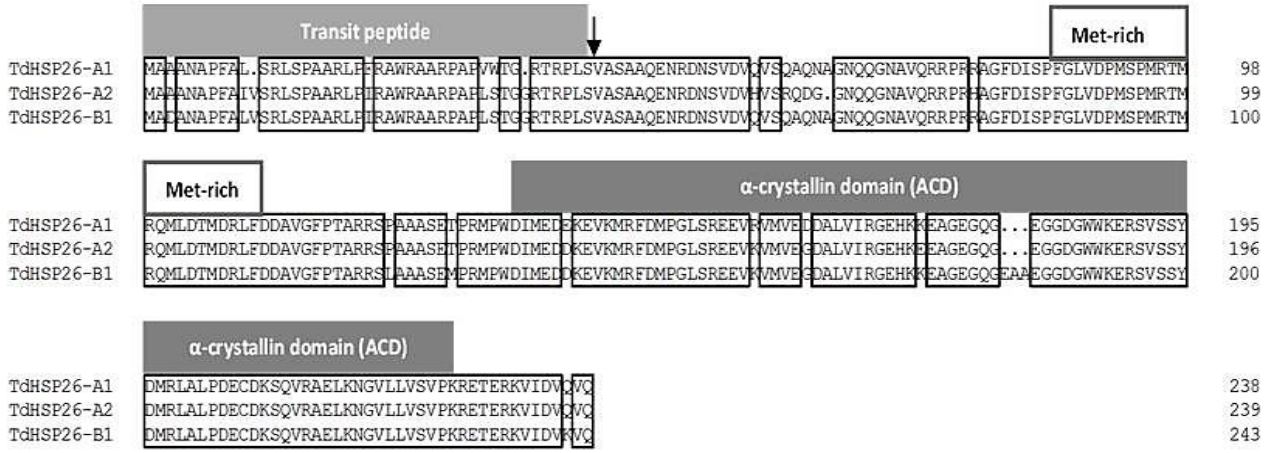
chromosome arm homoeologues of group 4. As a result, the short arm of 4A (4AS) is homoeologous to 4BL and 4DL, while 4AL is homoeologous to 4BS and 4DS (Miftahudin *et al.* 2004; Chen and Gustafson 1997; Devos *et al.* 1995; Mickelson-Young *et al.* 1995; Nelson *et al.* 1995; Liu *et al.* 1992; Naranjo 1990).

## 2.3.4 Analysis of the predicted TdHSP26 proteins

The retrieved *TdHsp*26 gene sequences were analysed for their predicted protein sequences. Since only small varietal differences were observed between Cham1 and Kronos TdHSP26 protein sequences, only the Cham1 sequences were considered for further characterization showed in this section.

For all the TdHSP26 protein sequences the ChloroP prediction software revealed the presence of a chloroplast Transit Peptide (cTP) that should direct the mature protein into the chloroplast. The cTP length was predicted of 42 AA for the TdHSP26-A1 sequence and of 44 AA in the TdHSP26-A2 and B1 sequences, the predicted cleavage site is indicated by the black arrow in Figure 2.4A, in according also to previous literature data (Chauhan *et al.* 2012).

A)




B)

| | |
|---|---|
| TdHSP26-A1 | 100% |
| TdHSP26-A2 | 94.1% 100% |
| TdHSP26-B1 | 95.8% 95.0% 100% |

**Figure 2.4** A) Alignment of the amino acid sequence of identified Cham1 TdHSP26-A1, TdHSP26-A2 and TdHSP26-B1 proteins. cTP (chloroplast transit peptide) and ACD are indicated by grey boxes according to Chauhan *et al.* 2012. The cTP cleavage site predicted by ChloroP 1.1 software is indicated by the black arrow. The highly conserved Met-rich region in the N-terminal domain is indicated with the white box in according to Sharf *et al.* 2001. B) Homology matrix of TdHSP26-A1, TdHSP26-A2 and TdHSP26-B1.

In agreement with the HSP26 chloroplast sub-family structure conservation and previous literature data (Chauhan *et al.* 2012, Scharf *et al.* 2001, Bondino *et al.* 2012, Van Montfort *et al.* 2001, Waters 2013) a typical HSP26 topology characterized by the presence of the N-terminal domain (N-term) containing the conserved Methionine-rich Domain amphipathic α-helix (MrD) and the ACD domain (Figure 2.4A) is detectable in all of the identified proteins.

The three functional members of *TdHsp*26 family encoded proteins of 238 amino acids (TdHSP26-A1), 239 amino acids (TdHSP26-A2) and 243 amino acids (TdHSP26-B1). TdHSP26-A1, TdHSP26-A2 and TdHSP26-B1 share a high amino acid sequence identity that varies from 94.1 to 95.8% (Figure 2.4B) and on the basis of their predicted molecular weight they could be named TdHSP26.5, TdHSP26.4, and TdHSP26.8, respectively (Figure 2.2A).

To assess the sequence variation in the TdHSP26 protein family, an alignment with previous known chloroplast-localized HSP26s belonging to wheat and other plants species was also performed (Figure 2.5). TdHSP26-A1, A2 and B1 (indicated in bold on the alignment) share the conserved MrD domain presents in all the others chloroplast-localized sHSP26, as well as the ACD domain that contains nine highly conserved β-sheets with the Consensus Region I (CRI) and the Consensus Region II (CRII) linked by the β6 domain (Waters 2013, Figure 2.5). The wheat sHSP26 analysed in Figure 2.5 share a high sequence identity and a strong conservation of the key functional domains, especially in the N-terminal domain which usually show the highest rate of divergence between the sHSPs classes but high conservation within the sHSP subfamily (Waters 2013, Waters and Vierling 1999, Bondino *et al.* 2012).

The most relevant difference between TdHSP26-A1 and A2 proteins was observed in the N-term between the residues 60/64 of –A1 and 62/65 of –A2 where TdHSP26-A1 shown the sequence QAQNA (conserved also on the TdHSP26-B1) whereas the TdHSP26-A2 shown the sequence RQDG (Figure 2.4 and Figure 2.5); this difference in the N-term could influence their substrate binding affinity conferring different specificity to the two HSP26 isoforms. Moreover TdHSP26-B1 shown, between the consensus region I and II of the ACD domain, a tripeptide insertion (EAA) described also by Campbell (1998) that is conserved in other HSP26 sequences in Figure 2.5 (as TaHSP26.6g, TasHSP, HvHSP26), which may alter the tertiary structure of the protein conferring more flexibility.

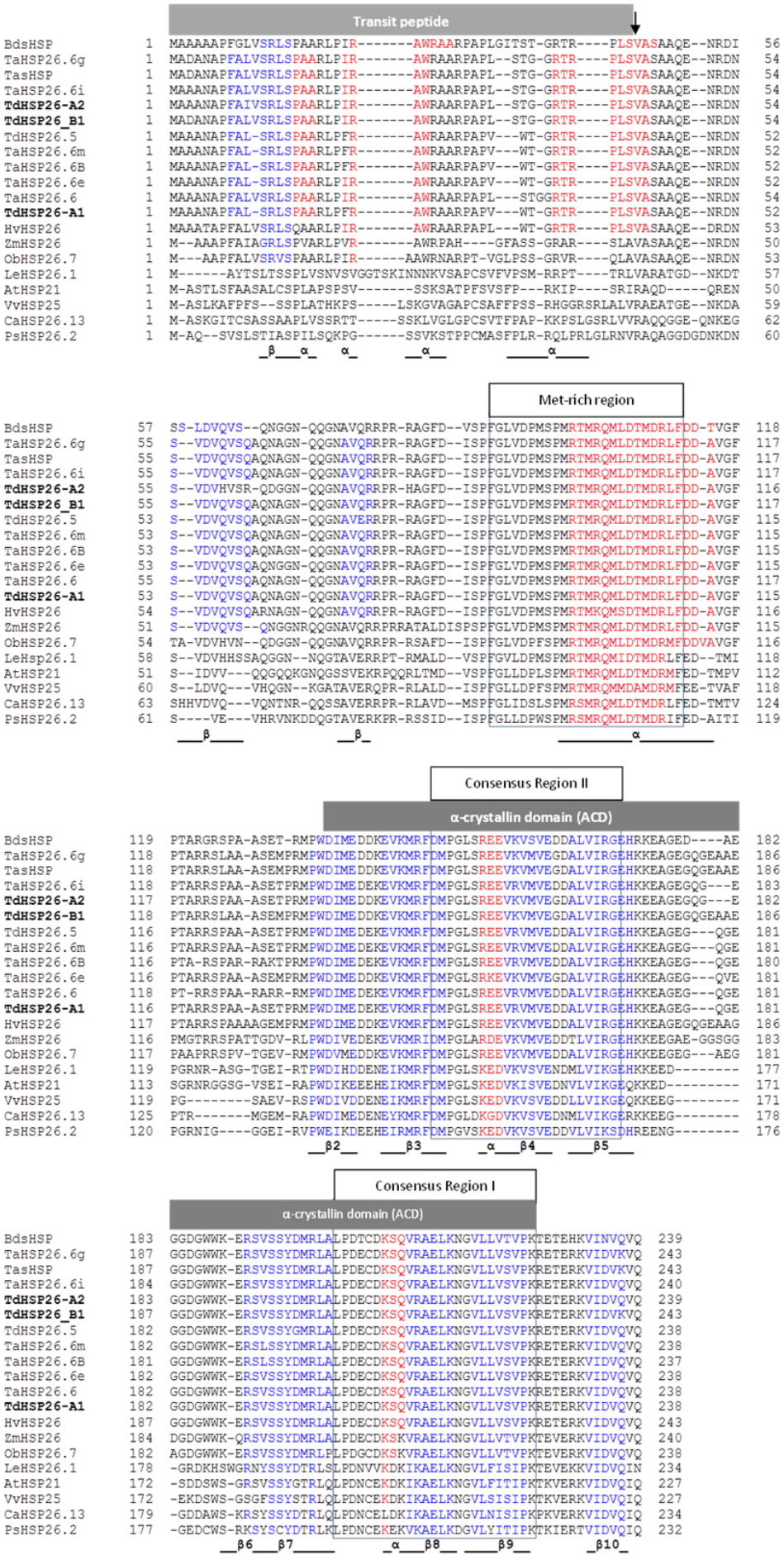

Transit peptide
Met-rich region
Consensus Region II
α-crystallin domain (ACD)
Consensus Region I
α-crystallin domain (ACD)

**Figure 2.5** Alignment of the protein sequence identified in this work with sHSP26 of *T. aestivum*, *T. durum* and of other plant species previously characterized as Chloroplast sHSP sub-family (Sarkar *et al.* 2009; Sharf *et al.* 2001).
BdsHSP (XP_003558381), TaHSP26.6g (AAAC96315), TasHSP (ADN97108), TaHSP26.6i (AAC96316), TdHSP26-A1, TdHSP26-A2, TdHSP26-B1, TdHSP26.5 (CAI96515), TaHSP26.6m (AAC96317), TaHSP26.6B (CAA47745), TaHSP26.6e (AAC96314), TaHSP26.6 (CAA41219), HvHSP26 (AAB28590), ZmHSP26 (AAA33477), ObHSP26.7 (XP_006649731), LeHSP26.1 (AAB49626), AtHSP21 (CAB38279), VvHSP25(CAO48583), CaHSP26.13 (AFZ94855), PsHSP26.2 (CAA30167).
The α-helix and β-sheets predicted with the Promals3D structural alignment program are colored in red and blue respectively and are also indicated th the bottom according to literature data (Waters 2013, Sarkar *et al.* 2009) The transit peptide and the ACD region are indicated with gray boxes according to Chauhan *et al.*, 2012. The cleavage site of the transit peptide is indicated by arrow. The conserved features as the Met-rich region on the N-term, the β-sheets on the ACD and the consensus region I and II are indicated with white boxes according to Scharf *et al.* 2001, Waters 2013.

The alignment of the annotated *Triticum* HSP26 protein sequences with the newly identified TdHSP26 of the cv. Cham1 allow to analyse in details the variations existing between them as reported in the phylogenetic tree (Figure 2.6A) and in the homology matrix (Figure 2.6B) generated from the alignment.

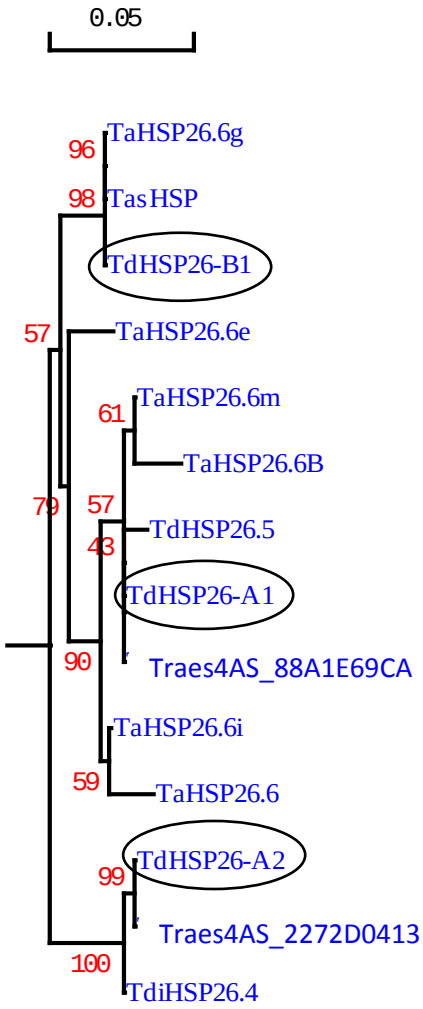
The phylogenetic tree showed that our three sequences (indicated by a circle in Firure 2.6A) grouped, even if with slight differences, in three separate clusters (Figure 2.6A). In particular, TdHSP26-A1 share 99.2% identity (2 AA substitution, one in the N-term domain and one in the ACD) with the sequence TdHSP26.5 isolated by the durum wheat cv. Creso (Rampino *et al.* 2009b), the query sequence used for the allele discovery. This sequence variability observed between TdHSP26-A1 and TdHSP26.5 may be due to varietal variability. Moreover, TdHSP26-A1 shows a 99.6% identity with TaHSP26.6-m sequence (AAC96317) identified from the bread wheat cultivar Mustang (Campbell 1998; Figure 2.6B), but it shows 97.9% sequence identity with TaHSP26.6B sequence (CAA47745) identified in cv. Mustang (Nguyen *et al.* 1993, Figure 2.6B) suggesting that TaHSP26.6B could be coded by a different orthologous gene.

TdHSP26-B1 shows 100% sequence identity with the TasHSP (ADN97108) and TaHSP26.6g (AAC96315) isolated from the bread wheat cv. C306 and Mustang, respectively (Campbell 1998).

Finally, TdHSP26-A2 share 100% identity with the Traes4AS_2272D0413 (abbreviated as T_2272D0413 in Figure 2.6B), a Chinese Spring sequence retrieved in EnsemblPlants database. Moreover, it share 99.6% identity with the *T. dicoccoides* HSP26.4 (CAI96512) that show a 26.4 KDa molecular weight. Since no match was found with previously isolated durum or bread wheat sHSPs the TdHSP26-A2 can be classified as a new durum wheat HSP26.4 (TdHSP26.4).

The TdHSP26-A1, TdHSP26-B1, and TdHSP26-A2 show a slightly different molecular weight of 26.5, 26.8 and 26.4 KDa respectively (as reported in Figure 2.2A), moreover they clusterize in the phylogenetic tree (Figure 2.6A) together with homologous sequence of similar molecular weight (Table 2.1), therefore they could be classified as TdHSP26.5, TdHSP26.4, and TdHSP26.8.

A)




B)

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TdHSP26-A1** | 100% | | | | | | | | | | | |
| **TdHSP26-A2** | 94.1% | 100% | | | | | | | | | | |
| **TdHSP26-B1** | 95.8% | 95.0% | 100% | | | | | | | | | |
| T_88A1E69CA | 100.0% | 94.1% | 95.8% | 100% | | | | | | | | |
| T_2272D0413 | 94.1% | 100.0% | 95.0% | 94.1% | 100% | | | | | | | |
| TdHSP26.5 | 99.2% | 93.2% | 95.0% | 99.2% | 93.2% | 100% | | | | | | |
| TaHSP26.6g | 95.8% | 95.0% | 100.0% | 95.8% | 95.0% | 95.0% | 100% | | | | | |
| TaHSP26.6e | 96.6% | 94.5% | 96.6% | 96.6% | 94.5% | 95.8% | 96.6% | 100% | | | | |
| TaHSP26.6i | 98.7% | 95.4% | 97.1% | 98.7% | 95.4% | 97.9% | 97.1% | 96.2% | 100% | | | |
| TaHSP26.6m | 99.6% | 93.7% | 95.4% | 99.6% | 93.7% | 98.7% | 95.4% | 96.2% | 98.3% | 100% | | |
| TaHSP26.6 | 97.0% | 93.7% | 95.8% | 97.0% | 93.7% | 96.2% | 95.8% | 94.9% | 98.3% | 96.6% | 100% | |
| TdiHSP26.4 | 94.5% | 99.6% | 95.4% | 94.5% | 99.6% | 93.7% | 95.4% | 94.9% | 95.8% | 94.1% | 94.1% | 100% |
| TasHSP | 95.8% | 95.0% | 100.0% | 95.8% | 95.0% | 95.0% | 100.0% | 96.6% | 97.1% | 95.4% | 95.8% | 95.4% | 100% |
| TaHSP26.6B | 97.9% | 91.9% | 93.7% | 97.9% | 91.9% | 97.0% | 93.7% | 94.5% | 96.6% | 98.3% | 97.0% | 92.4% | 93.7% | 100% |

**Figure 2.6** A) Phylogenetic tree of *Triticum* HSP26 proteins. The tree was derived by Neighbour Joining methods with bootstrap analysis (1000 replicates) from the alignment of the entire protein sequence of wheat HSP26 annotated in NCBI database with the newly identified TdHSP-A1, TdHSP26-A2 and TdHSP26-B1 protein sequences. TdHSP26-A1, TdHSP26-A2 and TdHSP26-B1 are signed by circle. B) Homology Matrix of *Triticum* HSP26 proteins. The matrix was derived from the alignment TdHSP26-A1, TdHSP26-A2 and TdHSP26-B1 are in bold. The abbreviations of *Triticum* HSP26 are as follow: TaHSP26.6g (AAAC96315), TasHSP (ADN97108), TaHSP26.6i (AAC96316), TdHSP26-A1, TdHSP26-A2, TdHSP26-B1, TdHSP26.5 (CAI96515), TaHSP26.6m (AAC96317), TaHSP26.6B (CAA47745), TaHSP26.6e (AAC96314), TaHSP26.6 (CAA41219), TdiHSP26.4 (CAI96512), Traes4AS_88A1E69CA (T_88A1E69CA), Traes4AS_2272D0413 (T_2272D0413).

## 2.3.5 Promoter isolation and analysis

For *TdHsp26-A1* and *TdHsp26-B1* two up-stream sequences of 1,379 bp and 1,396 bp respectively, were isolated from genomic DNA of cv. Cham1. These sequences (so far called *Prom_TdHsp26-A1* and *Prom_TdHsp26-B1*, respectively) were *in silico* analysed for the prediction of the regulatory elements. Sequence comparisons between *Prom_TdHsp26-A1*, *Prom_TdHsp26-B1* and the corresponding promoter sequence previously cloned in *T. aestivum* by Chauhan and co-workers (named *TaHSP26-P26*; 1,514 bp) shown that *Prom_TdHsp26-B1* was 98.93% identical to *TaHSP26-P26* whereas 93.39% identity was observed between *Prom_TdHsp26-A1 and TaHSP26-P26.* Between the *TdHsp26-B1* and *TdHsp26-A1* promoter sequences a 95.61% identity was found. As expected *Prom_TdHsp26-B1* shows a higher sequence similarity than *Prom_TdHsp26-A1* with the promoter sequence *TaHSP26-P26* since Chauhan and co-workers (2012) started from the *TaHsp26.6-g* alleles (AF097657) which shows, as previously observed, 100% identity with *TdHsp26-B1* (Figure 2.6 A,B).

All the abiotic stress responsive, developmental and tissue-specific elements predicted by PLACE and PlantCARE on the promoter sequence of *TdHsp26-A1* and *TdHsp26-B1* genes are reported in Table 2.5 and highlighted on the alignment between *Prom_TdHsp26-A1* and *Prom_TdHsp26*-B1 in Figure 2.7. The PLACE analysis reveal the presence in *Prom_TdHsp26-A1* and *Prom_TdHsp26-B1* of all the elements previously identified by Chauhan and co-workers (2012; Table 2.2) (Khurana *et al.* 2013, Chauhan *et al.* 2012). In addiction to these elements, other *cis*-elements were identified as indicated in Table 2.5; in particular:

- four I-box core which were previously detected up-stream a light–regulated genes in *Arabidopsis* (Terzaghi and Cashmore 1995);
- one putative low temperature responsive element (LTRE) found in barley low temperature responsive genes (Nordin *et al.* 1993);
- two PRE (plastid response element) consensus sequence (PRECONSCRHSP70A) found in the promoters of HSP70A in Chlamydomonas and involved in the induction of HSP70A gene by light conditions (Von Gromoff *et al.* 2006);
- five SORLIPs (Sequence Over-Represented in Light-Induced Promoters) element previously reported in *Arabidopsis* SORLIP1AT/ SORLIP2AT (Jiao *et al.* 2005) of which only four where detected on the *Prom_TdHsp26-B1*.
- ten TGAC-containing W-box elements out of which: eight were binding site for rice WRKY71 (WRKY71OS) that is a transcriptional repressor of the gibberellin signaling pathway (Eulgem *et al.* 1999) and two were WBOXHVISO1 conserved in barley iso1 (encoding isoamylase1) promoter (Sun *et al.* 2003). W-box elements are usually found

within Pathogenesis-Related genes promoters suggesting a possible role of *TdHsp26* in the Wheat-Pathogen interaction.

- one ABRE *cis*-acting element involved in the abscissic acid responsiveness in *Arabidopsis* (ABRE) (Narusaka *et al.* 2003);
- two TC-rich repeats elements involved in defense and stress responsiveness in *Nicotiana tabacum* (TC-rich) (Sun *et al.* 2010);
- one GAG-motif, part of a light responsive element in barley (Nalbandi *et al.* 2014);
- a TATA box, that is considered the core of the promoter sequence for the transcription start.

**Table 2.5** *cis*-elements identified by PLACE and PlantCARE on *Prom_TdHsp26-A1* and *Prom_TdHsp26-B1*. The number of motifs on the two homoeologous sequence and their position in respect to the ATG are reported.

| Motif name | Number of motifs found on *Prom_TdHsp26-A1* | Number of motifs found on *Prom_TdHsp26-B1* | Position on *Prom_TdHsp26-A1* (bp) | Position on *Prom_TdHsp26-B1* (bp) | Software used for prediction |
|---|---|---|---|---|---|
| GT1GMSCAM4 | 1 | 1 | -1,344 | -1,361 | PLACE |
| ABRELATERD1 | 2 | 2 | -442/ -1,211 | -458/ -1,227 | PLACE |
| ABRE | 1 | 1 | -878 | -893 | PlantCARE |
| CCAATBOX1 | 3 | 2 | -92/ -704/ -1,191 | -720/ -1,207 | PLACE |
| MYC -rd22 | 5 | 5 | -645/ -777/ -992/ -1,070/ -1,107 | -670/ -793/ -1,000/ -1,086/ -1,123 | PLACE |
| STRE | 1 | 1 | -767 | -783 | PLACE |
| POLLEN1LELAT52 | 4 | 4 | -298/ -390/ -448/-557 | -315 / -407/ -464/-573 | PLACE |
| GCN4 | 1 | 1 | -483 | -499 | PLACE |
| HSE-P | 1 | 1 | -290 | -307 | PLACE |
| HSE-G | 1 | 1 | -206 | -223 | PLACE |
| HSE-S | 1 | 1 | -201 | -218 | PLACE |
| CBFHV | 1 | 1 | -38 | -57 | PLACE |
| I-box core | 4 | 4 | -98/ -282/ -948/ -1,148 | -115/ -299/ -964/ -1,164 | PLACE |
| LTRE | 1 | 0 | -38 | | PLACE |
| PRECONSCRHSP70A | 2 | 2 | -548/ -908 | -564 / -924 | PLACE |
| SORLIP1AT | 2 | 2 | -1,135/ -1,238 | -1,151/ -1,254 | PLACE |
| SORLIP2AT | 3 | 2 | -247/ -322/ -360 | -339/ -377 | PLACE |
| WRKY71OS | 8 | 8 | -314/ -375/ -577/ -645/ -656/ -666/ -935/ -1,072 | -331/ -392/ -593/ -661/ -672/ -682/ -951/ -1,088 | PLACE |
| WBOXHVISO1 | 2 | 2 | -491/ -1,259 | -497/ -1,275 | PLACE |
| MYB | 2 | 2 | -1,361/ -1,379 | -1,211/ -1,227 | PLACE |
| TC-rich | 2 | 2 | -449/ -827 | -465/ -843 | PlantCARE |
| GAG | 1 | 1 | -123 | -140 | PlantCARE |
| TATA box | 1 | 1 | -147 | -164 | PlantCARE |

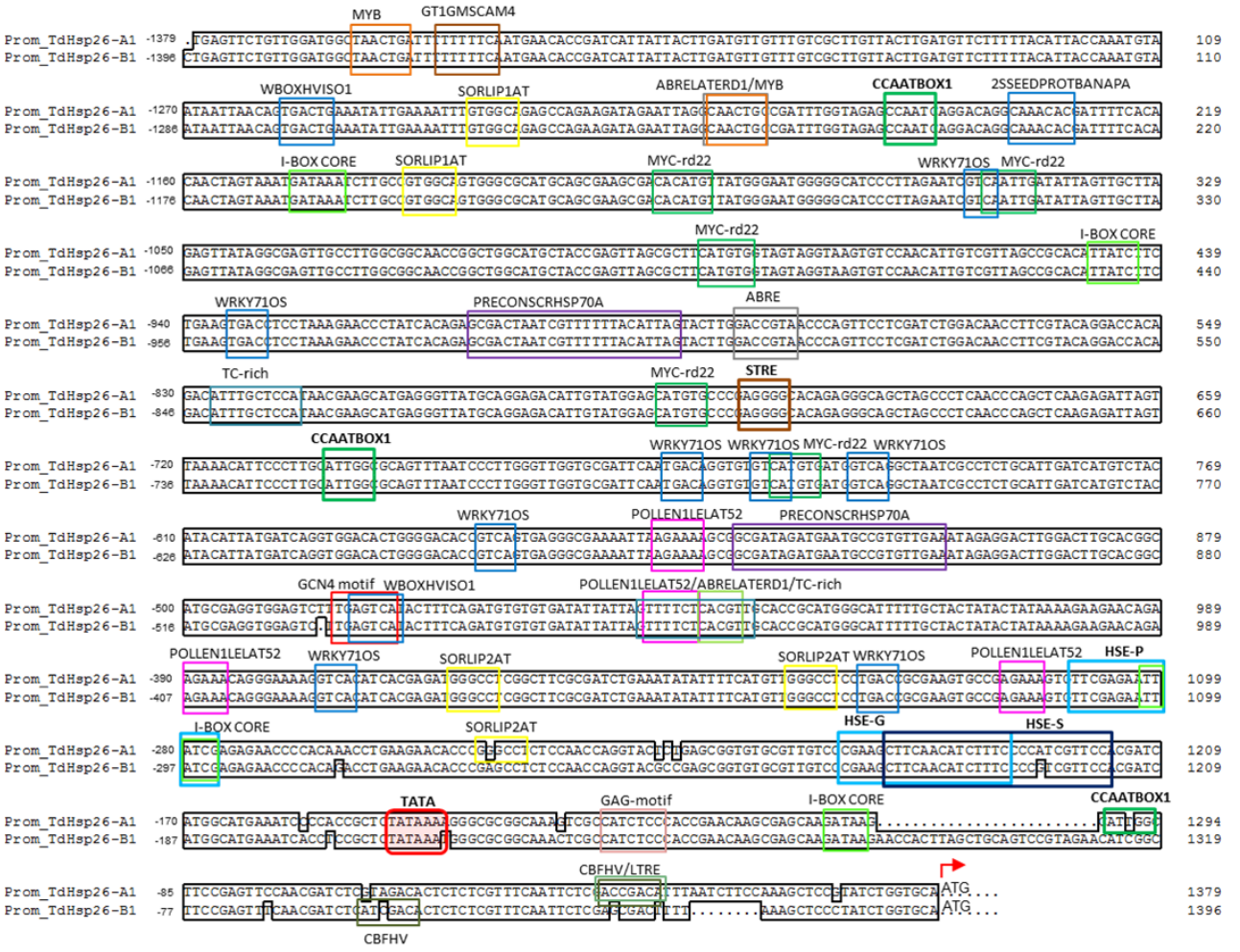



**Figure 2.7** Alignment of the *TdHsp26-A1* and *TdHsp26-B1* promoter sequences of cv. Cham1 with the indication of regulatory elements predicted by PLACE and PlantCARE software. The black box on the alignment highlights the conserved sequence. All the regulatory motifs are highlighted by colored boxes. The TATA motif highlighted with red thick box. The CCAATBOX detected are highlighted with green thick boxes and the additional CCAATBOX detected on *TdHsp26-A1* is highlighted with dark green thick box. The HSEs motifs of perfect-type (HSE-P) and gap-type (HSE-G) are highlighted by light-blue thick boxed whereas the step-type HSE (HSE-S) with blue thick box. The motifs previously discussed by Chauhan *et al.* 2012 (reported in table2.2) are indicated on the alignment as GT1GMSCAM4, ABRELATERD1, CCAATBOX1, 2SSEEDPROTBANAPA, MYC-rd22, STRE, POLLEN1LELAT52, GCN4 motif, CBFHV and boxed with different colours.

Several SNPs and indels were observed in the first 300 bp before the ATG of the two promoter sequences *Prom_TdHsp26-A1* and *Prom_TdHsp26-B1*, where less *cis*- acting elements are localized and conserved between the two promoters. In this region an important motif as the CBFHV is present on both the sequences although in slightly different position due to the presence of SNPs (Figure 2.7). Furthermore, the A/G SNP detected in the step-type HSE (S-HSE) (nTTCn(5bp)nGAAn(5bp)nTTCn) falls within the 5 bp variable position allowing to maintain the a putative functional element. Two of the retrieved elements: the LTRE and one additional SORLIP1AT have been detected only in the *Prom_TdHsp26-A1* sequence (Figure 2.7).

The most interesting difference observed between the two promoters is the generation of an additional CCAAT box in the *Prom_TdHsp26-A1* sequence at position -1,293 bp that was not observed in the *Prom_ TdHsp26-B1*. The CCAAT box elements present in the HSP26 promoter contribute synergistically and in a combinatorial manner with other *cis*-elements to regulate the transcription of heat shock genes (Khurana *et al.* 2013), therefore an addition CCAAT box may positively alter the expression of *TdHsp26-A1* upon heat stress. Moreover, the presence of a different number of CCAAT box elements in the promoters of the *TdHSP26-A1* and *TdHsp26-B1* genes support the idea that *TdHsp26-A1* and *TdHsp26-B1* may play a different roles in the stress defence response.

## 2.4 Conclusion

In the present work, we have identified and characterized *in silico* four new genes encoding for sHSP26 in *T. durum.* In particular three genes *TdHsp26-A1*, *TdHsp26-A2*, *TdHsp26-B1* and one putative pseudogene (*TdHsp26-A3*) have been isolated and characterized. The predicted protein sequence of *TdHsp26-A1, TdHsp26-A2,* and *TdHsp26-B1* showed the typical sHSP topology and were classified on the basis of their molecular weight of putative proteins as TdHSP26.5, TdHSP26.4, and TdHSP26.8, respectively.

Three members of the sHSP26 family in durum wheat are localized on the short arm of chromosome 4A, while a single member of the family *TdHsp26-B1,* that show the highest conservation among *Triticeae,* is located on the long arm of chromosome 4B.

From the bioinformatic research performed on the available databases, no matching sequences were retrieved on the 4BS or 4AL, but is not possible to exclude the presence of other *Hsp26* gene copies. The retrieved protein sequences show a high conservation, in particular in the functional domains, with a peculiar 3 AA insertion in the ACD of the TdHSP26-B1. These finding suggest a redundancy of TdHSP26 proteins in wheat that underline their putative central role in the plant stress resistance. Moreover, a possible different *Hsp26* genes regulation is supported by the finding of different predicted stress responsive motifs and CCAAT boxes in the *TdHsp26-A1* and *TdHsp26-B1* promoter sequence. Many stress responsive elements were retrieved together with pathogen related elements in HSP26 promoters supporting the importance of these genes in the response of durum wheat to abiotic and biotic stresses.

The characterization of the complete genomic sequences of *TdHsp26* genes and the development of genome-specific primers allow to target the different *Hsp26* genes in the following TILLING experiments.

# Chapter 3. Identification of novel *TdHsp26* alleles

## 3.1 Material and methods

### 3.1.1 Vegetal material and growth conditions

A DNA library of M2 cv. Cham1 TILLING population previously obtained (Parry *et al.* 2009) was used as template for the SNPs detection. The wheat online TILLING database (Uauy C. personal communication: www.wheat-tilling.com), based on the exome captured sequencing data of the Kronos TILLING population (Uauy *et al.* 2009), was used to identify the mutations online.

The selected mutant lines (at the M3 generation for Cham1 and at the M4 generation for Kronos) and the corresponding parental wild type were provided by Dr. Cristobal Uauy (JIC, Norwich UK). Sixteen plants for each mutant line and six plants for each parental wild type were sown. Wheat seeds were surface sterilized by immersion in sodium hypochlorite (0.5% v/v) for 30 min, and then rinsed thoroughly in sterile water and sown in Petri dishes in imbibition conditions for 10 days at 5 °C. The seeds, after the germination, were transferred in peat and send mix (85% fine peat, 15% Grit, 2.7 Kg/$m^3$ Osmocote, wetting agent, 4 kg/$m^3$ Maglime, 1 Kg PG mix) in three 96-wells trays. All the plants were grown in glasshouse under environmental temperature and light conditions and daily watered. At Z11-Z12 seedling stage the plants were transplanted in John Innes Cereal Mix (40% medium grade peat, 40% sterilized soil, 20% horticultural grit, 1,3 kg/$m^3$ PG Mix 14-16-18 + Te base fertilizer, 1 kg/$m^3$ Osmocote Mini 16-8-11 2 mg + Te 0.02% B, wetting agent, 3 kg/$m^3$ Maglime, 300 g/$m^3$ Exemptor) and grown till grain harvesting. 2-4 cm of Z11-Z12 seedling leaves were sampled for the DNA extraction.

### 3.1.2 Online TILLING analysis

The wheat TILLING database (Uauy C. personal communication: www.wheat-tilling.com) was used to search allele variations in the *TdHsp26-A1, A2* and *B1* gene copies. This resource allows to identify *in silico* the putative mutations in a target gene sequence present in the database (Uauy *et al.* 2009, Borrill *et al.* 2015) and predict the effects of the mutation retrieved on the protein function based on the protein annotation available in EnsemblPlants.

The Traes and scaffolds retrieved in EnsemblPlants as previously described in Chapter 2 (Table 2.4) (Traes_4AS_8BA1E69CA for the HSP26-A1, Traes_4AS_2272D0413 for the HSP26-A2 and the scaffold information IWGSC_CSS_4BL_scaff_6982788 for the HSP26- B1) were used as query for searching the wheat TILLING database. For each Kronos mutant line identified, the predicted

mutations were manually verified on the sequences and the predicted effects on the nucleotide and protein sequences were evaluated.

## 3.1.3 Conventional screening of TILLING population by High Resolution Melting (HRM)

### 3.1.3.1 Allele-specific primer design

TILLING primers were developed on the basis of the COODLE analysis that predict the most conserved functional domains with the highest probability to cause a loss of function in the target gene. *TdHsp26-B1* was chosen as target. The primers reported in Table 3.1 were manually designed by taking advantage of the SNPs observed in the *TdHsp26* gene sequences and were used for the HRM screening of the Cham1 TILLING library using a nested-PCR strategy.

**Table 3.1** Primer pairs used for the HRM screening of the Cham1 TILLING library. The target sequence used for primer design and the position of the primer in respect to the ATG of *TdHsp26-B1* gene are indicated for each primer.

| Primer name | Type | Designed on | Primer position (bp to ATG) | Sequence 5'-3' | Amplicon size (bp) | Application |
|---|---|---|---|---|---|---|
| **B1-17F** | Forward | Prom_TdHsp26-B1 | -259 | TCTCCAACCAGGTACGCC | 1378 | TdHsp26-B1 1st round HRM screening of Cham1 TILLING library |
| **B1-6R** | Reverse | IWGSC_chr4BL_contig_6982788 | +1119 | GTTATCAGCTTCTTCCGGG | | |
| **B1-PT10F** | Forward | *TdHsp26-B1* | +393 | CGATGCGGCAGATGCTT | 211 | TdHsp26-B1 2nd round HRM screening of Cham1 TILLING library |
| **B1-PT10R** | Reverse | *TdHsp26-B1* | +604 | TGACGAGCGCGTCGC | | |

### 3.1.3.2 HRM screening of the Cham1 TILLING population

The DNA library of Cham1 TILLING population (Parry *et al.* 2009) composed by 960 DNAs was kindly provided by Dr. Cristobal Uauy (JIC, Norwich UK) and used for the conventional TILLING screening by HRM.

High Resolution Melting (HRM) was used to screen the Cham1 library by using the *TdHsp26-B1* gene as target through a nested-PCR strategy with two rounds of PCR as schematized in Figure 3.1. The DNAs of the Cham1 library were randomly pooled 2X. The experimental design used for the HRM screening is reported in Figure 3.1.

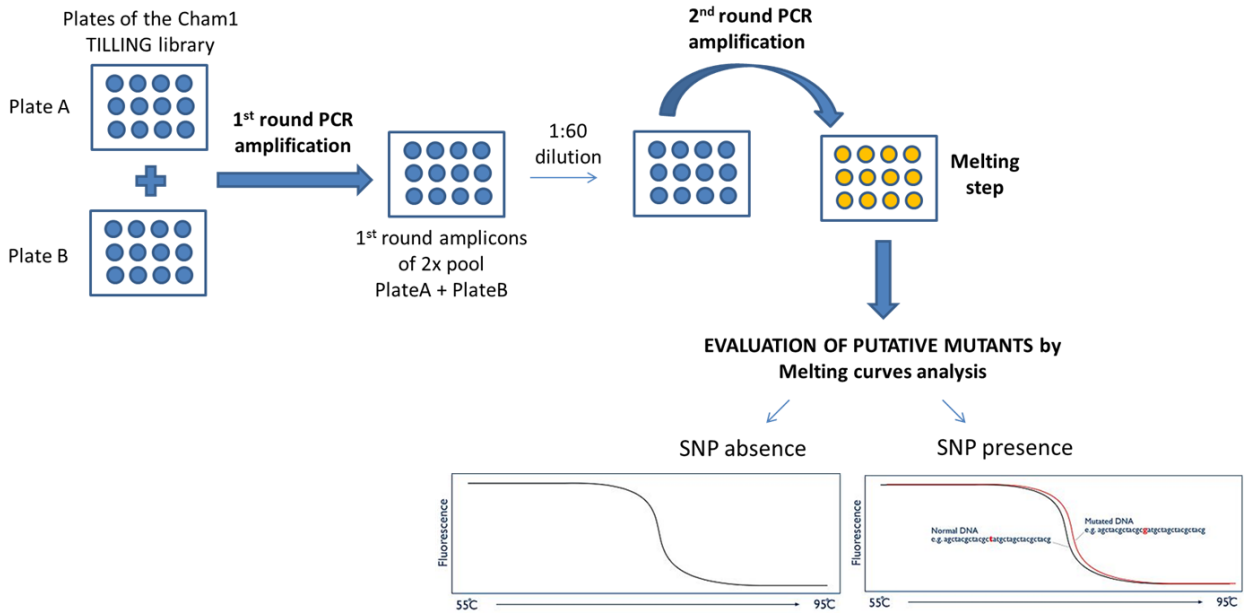



**Figure 3.1** Experimental design used for the HRM screening of the Cham1 TILLING library.

### 1st round PCR

The 2X pooled DNAs were amplified in a 96-wells plates in a Veriti Thermo Cycler (Life Technologies) in a 12.5 µl reaction volume using 10-20 ng of pooled DNA, 2X GoTaq Hot Start Colorless Master Mix (Promega, WI, USA) and 0.25 µM of B1-17F/B1-6R primers (Table 3.1) with the same cycling conditions described for gene isolation (paragraph 2.2.3.2).

### 2nd round PCR

A 1:60 dilution of the 1st round plates were used as template in the 2nd round reactions carried out on RealTime7900HT (Applied Biosystems) in a 10 µl reaction volume using 2X MeltDoctor™ HRM Master Mix (Life Technologies) and 0.3 µM of homoelogous-specific primers B1-PT10F/B1-PT10R (Table 3.1). The 2nd round reactions were set up in 96-wells microplates with optical read covers (Life Technologies), briefly centrifuged and analysed with SDS software v2.4 as recommended by manufacturer. The 2nd round consist of a cycling stage followed by the melt curve/dissociation stage, with the amplification step conditions at 95 °C for 10 mins, 40 cycles of 95 °C for 15 s, 60 °C for 1 min, followed by the melting step cycle 95 °C for 10 s, 60 °C for 1 min, 95 °C for 15 s at 1% ramp rate, 60 °C for 15 mins. During the HRM stage of the melt curve, double-stranded amplicons slowly denature, releasing bound dsDNA-binding dye. The real-time PCR instrument measures this decrease in fluorescence signal, and the HRM software plots fluorescence signal over temperature.

**Melting curves analysis**

The melt curve raw data obtained for each plates with SDS software v2.4 were analysed with the HRM software v2.0.1 by manually setting appropriate active melt region within the pre-melt and post-melt regions, which are the regions used by the software to align the data defining the 100% of fluorescence (where every amplicon is double-stranded) and the 0% fluorescence point (where every amplicon is single-stranded) respectively. The melt curve raw data are generally plotted as fluorescence vs. temperature with an inflection at the Tm of the amplicons and the most common visualization is the negative first derivative plot that shows the Tm as peaks (Figure 3.2A). Once the pre-melt and post-melt regions are set, the software re-analyses the data and defines the Tm of the PCR product as the inflection point of the melt curves within the active melt region. The aligned melt curves plot displays the melt curves as % melt over temperature (Figure 3.2B). The melt curves are aligned to the same fluorescence level using the pre- and post-melt regions enabling easy discrimination of sequence variants that display differences in their melt curve behavior by automatic variant call. Is important to note that Tm alone may not discriminate between some genotypes, as Hom and Het genotype amplicons (green and blue curves in Figure 3.2B) could show virtually identical Tm but can be easily distinguished based on the shape of their melt curves. Small differences between melt curves are best visualized using a difference plot that displays the aligned data as the difference in fluorescence between the melt curves for a reference sample and the other melt curves (Figure 3.2C).

The variant curves were automatically detected in each 2X pool plates by setting the following pre-melt and post-melt regions: 84.39-85.32 °C pre-melt region start-stop and 93.59-94.59 °C post-melt region start-stop. The variant curves were observed in the difference plot setting the most common variant (wild type) as reference. The variant curves with highest confidence were considered as positive pool containing the heteroduplex molecules for the target gene and then sequenced.

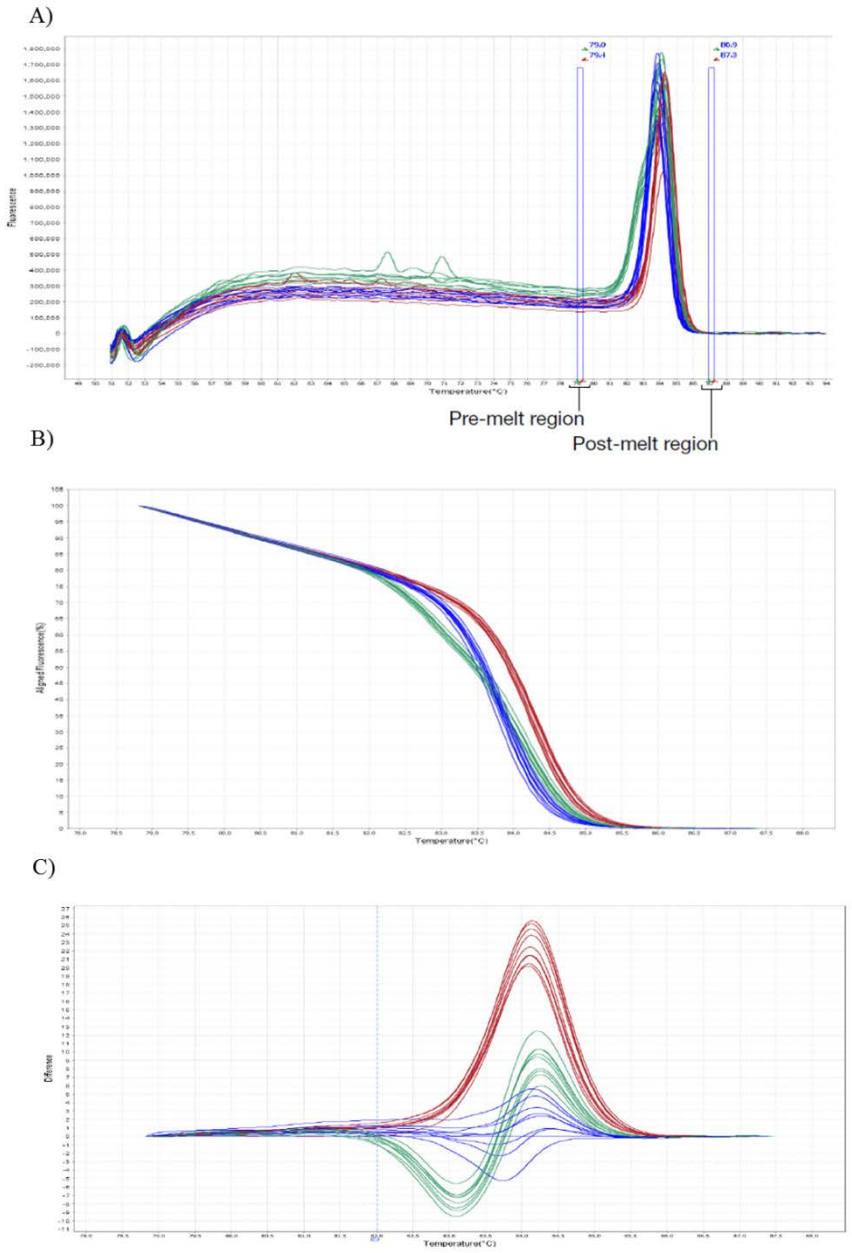



**Figure 3.2** Plots of the melt curves in the HRM analysis. A) Derivative melt curves plot fluorescence vs. temperature with the two sets of vertical bars before and after the data peak that define the pre-melt and post-melt regions. The pre- and post-melt regions circumscribe the active melt region and are used by the software to scale the data in the Aligned Melt Curves and Difference Plot and assign the variant call. B) Example of Aligned Melt curves (aligned fluorescence (%) vs. Temperature) with three variant groups, the curves of the heterozygous genotypes are in blue whereas the curves of the two homozygous genotypes are green and red. C) Example of Difference plot (difference vs. temperature), the heterozygote is selected as the reference (blue curves).

### 3.1.3.3 Mutation identification

The single genotypes of the Cham1 DNA library correspondent to the positive PCR products were separated and visualized by horizontal gel electrophoresis in TAE Agarose gel, purified with NZYGelpure kit (NZYtech, Lisbon, Portugal) and sequenced by using the commercial service Eurofins MWG (Eurofins Genomics, Edersberg, Germany) to identify the mutations (SNPs).

## 3.1.4 Bioinformatics

The putative mutations retrieved by HRM and by the online TILLING database search were analysed with PARSESNP. The software allows the prediction of the putative effect of the mutation on the protein function based on the sequence homology and the physical properties of amino acids through the identification of PSSM and SIFT scores (Kumar *et al.* 2009).

## 3.1.5 SNPs assay by KASP molecular markers

### 3.1.5.1 DNA extraction

Genomic DNA was extracted from 2-4 cm of Z11-Z12 seedlings leaves of the Cham1 and Kronos mutants and also from the corresponding parental wild type. The extractions were conducted in 96-wells 1.5 mL plates by adding 500 µl of extraction buffer composed of 1M Tris-HCl pH 7.5, 0.5 M EDTA, 10% SDS at 65 °C and grounding the samples in the 2010 Geno/Grinder (SPEX Sample Prep, Metuchen, NJ) at 1500 strokes $min^{-1}$ for 90 s. After 1 h incubation at 65 °C followed by 15 mins at 4 °C, 250 µl of 6 M ammonium acetate were added and the plates were leaved 15 mins at 4 °C for protein precipitations. The plates were centrifuged at 4000 rpm for 15 mins and 600 µl of supernatant containing the DNA were recovered in new plates containing 360 µl of cold isopropanol. The plates were than mixed and leaved 5 mins at room temperature for DNA precipitation. After 15 mins centrifugation at 4000 rpm the supernatant was discarded and the DNA pellet was washed in 500 µl of 70% ethanol. The plates were centrifuged again at 4000 rpm for 15 mins, the supernatant was discarded and the DNA pellet was dried and dissolved in 100 µl of sterile water. The DNA was vortex, centrifuged at 4000 rpm for 5 mins and diluted 1:10 before using for the SNP assay.

### 3.1.5.2 KASP markers development

Kompetitive Allele Specific PCR (KASP) (LGC Genomics, Teddington, UK) are codominant PCR based markers that allows to first follow the presence of the mutation detected and secondly to differentiate between homozygote and heterozygotes genotypes. The developed KASP markers were used to validate the SNPs detected both by the online database search and by the HRM screening (Trick *et al.* 2012, King *et al.* 2015).

The primers used in the SNP assay have been designed by using the bioinformatics pipeline PolyMarker (Ramirez-Gonzalez *et al.* 2015) that taking advantage on the information available through IWGSC-CSS automatically retrieve the best primers combination to exploit in the SNP assay. When IWGSC-CSS information where missing, as for the 4B chromosome, the primers were manually designed on the basis of the isolated gene sequence previously described in Chapter 2.

KASP assays are based on a three primers set system with two similar diagnostic primers incorporating the alternative varietal SNP at the 3' end and a third common primer that is preferentially selected to incorporate a genome-specific base at the 3' end assuring both genome and allele-specificity.

Oligonucleotides were ordered from Sigma-Aldrich (Gillingham, UK), with primers carrying standard FAM or HEX compatible tails (FAM tail: 5' GAAGGTGACCAAGTTCATGCT 3'; HEX tail: 5' GAAGGTCGGAGTCAACGGATT 3') with the target SNP at the 3’ base. A full list of primers is provided in Table 3.2.

**Table 3.2** KASP assay primers designed for each mutant line. Each KASP assay is composed of three primers, two allele-specific primers directed to the wild type (WT) or mutated sequence (SNP/Mutation) carrying at the 5' the FAM or VIC/HEX complementary sequence (FAM tail: 5' GAAGGTGACCAAGTTCATGCT 3', HEX tail: 5' GAAGGTCGGAGTCAACGGATT 3'; not reported in the table) and a common primer. The 3' specifically base of each allele-specific primer is in bold. For each primers, information about the design technique, the target mutant line with the target mutation and the target gene are indicated.

| Primer name | Type | Sequence 5'-3' | TAIL | Designed | Target mutant line (mutation) | Target gene |
|---|---|---|---|---|---|---|
| C0844-Fw1 | WT | GGACGACAAGGAGGTGAAGAT**G** | FAM | PolyMarker | Cham1 W4-0844 | *TdHsp26-B1* |
| C0844-Fw2 | SNP/Mutation | GGACGACAAGGAGGTGAAGAT**A** | VIC/HEX | PolyMarker | Cham1 W4-0844 (G536A) | *TdHsp26-B1* |
| C0844_1771-Rev | Common | ACCATCACCTTCACCTCCTC | - | PolyMarker | Cham1 W4-0844/ Cham W4-1771 | *TdHsp26-B1* |
| C1771-Fw1 | WT | AGATGCGGTTCGACATGC**C** | FAM | PolyMarker | Cham1 W4-1771 | *TdHsp26-B1* |
| C1771-Fw2 | SNP/Mutation | AGATGCGGTTCGACATGC**T** | VIC/HEX | PolyMarker | Cham1 W4-1771 (C550T) | *TdHsp26-B1* |
| C0181-Fw | Common | GGACGATGCGGCAGATGCTT | - | Manually | Cham1 W4-0181 | *TdHsp26-B1* |
| C0181-Rev1 | SNP/Mutation | ACGGCATCCGCGGCATCT**T** | FAM | Manually | Cham1 W4-0181 (G483A) | *TdHsp26-B1* |
| C0181-Rev2 | WT | ACGGCATCCGCGGCATCT**C** | VIC/HEX | Manually | Cham1 W4-0181 | *TdHsp26-B1* |
| K866-Fw1 | WT | TGTCGCCGATGAGGACGAT**G** | FAM | PolyMarker | Kronos0866 | *TdHsp26-B1* |
| K866-Fw2 | SNP/Mutation | TGTCGCCGATGAGGACGAT**A** | VIC/HEX | PolyMarker | Kronos0866 (G398A) | *TdHsp26-B1* |
| K866-Rev | Common | CAGCCGGTCCATCGTGTCA | - | PolyMarker | Kronos0866 | *TdHsp26-B1* |
| K1308-Fw1 | WT | CTGTTCGACGACGCCGTG**G** | FAM | Manually | Kronos1308 | *TdHsp26-B1* |
| K1308-Fw2 | SNP/Mutation | CTGTTCGACGACGCCGTG**A** | VIC/HEX | Manual | Kronos1308 (G444A) | *TdHsp26-B1* |
| K1308_265-Rev1 | Common | ATCTCGCTCGCCGCCGCCA | - | Manual | Kronos1308/ Kronos0265 | *TdHsp26-B1* |
| K265-Fw1 | WT | ACCGGCTGTTCGACGACG**C** | FAM | Manually | Kronos0265 | *TdHsp26-B1* |
| K265-Fw2 | SNP/Mutation | ACCGGCTGTTCGACGACG**T** | VIC/HEX | Manually | Kronos0265 (C439T) | *TdHsp26-B1* |
| K2202-Fw | Common | GAGCATGACGGTTTACTTCAAC | - | PolyMarker | Kronos2202 | *TdHsp26-A1* |
| K2202-Rev1 | WT | CCTCATCGGCGACATCG**G** | FAM | PolyMarker | Kronos2202 | *TdHsp26-A1* |
| K2202-Rev2 | SNP/Mutation | CCTCATCGGCGACATCG**A** | VIC/HEX | PolyMarker | Kronos2202 (C361T) | *TdHsp26-A1* |
| K670-Fw1 | WT | CGCTGGATTTGACATCTCC**C** | FAM | PolyMarker | Kronos0670 | *TdHsp26-A1* |
| K670-Fw2 | SNP/Mutation | CGCTGGATTTGACATCTCC**T** | VIC/HEX | PolyMarker | Kronos0670 (C253T) | *TdHsp26-A1* |
| K670-Rev | Common | CAAAATTCCGTGTAAAAGAACATCA | - | PolyMarker | Kronos0670 | *TdHsp26-A1* |
| K2205-Fw | Common | GACGATGCGGCAGATGCTG | - | Manually | Kronos2205 | *TdHsp26-A1* |
| K2205-Rev1 | WT | GGCGTCTCGCTCGCCG**C** | FAM | Manually | Kronos2205 | *TdHsp26-A1* |
| K2205-Rev2 | SNP/Mutation | GGCGTCTCGCTCGCCG**T** | VIC/HEX | Manually | Kronos2205 (G460A) | *TdHsp26-A1* |
| K2206-Fw | Common | AGCTCCTACGGCATGCGA | - | PolyMarker | Kronos2206 | *TdHsp26-A1* |
| K2206-Rev1 | WT | CACGTCGATGACCTTGC**G** | FAM | PolyMarker | Kronos2206 | *TdHsp26-A1* |
| K2206-Rev2 | SNP/Mutation | CACGTCGATGACCTTGC**A** | VIC/HEX | PolyMarker | Kronos2206 (C778T) | *TdHsp26-A1* |
| K2006-Fw1 | WT | GCAAGGTCATCGACGTG**C** | FAM | PolyMarker | Kronos2006 | *TdHsp26-A1* |
| K2006-Fw2 | SNP/Mutation | GCAAGGTCATCGACGTG**T** | VIC/HEX | PolyMarker | Kronos2006 (C796T) | *TdHsp26-A1* |
| K2006-Rev | Common | GGGTACAGAGTCTCGCACAA | - | PolyMarker | Kronos2006 | *TdHsp26-A1* |
| K869-Fw1 | WT | AGCACAAGAAGGAGGCC**G** | FAM | PolyMarker | Kronos0869 | *TdHsp26-A1* |
| K869-Fw2 | SNP/Mutation | AGCACAAGAAGGAGGCC**A** | VIC/HEX | PolyMarker | Kronos0869 (G616A) | *TdHsp26-A1* |
| K869-Rev | Common | CTCTTGTCGCACTCGTCC | - | PolyMarker | Kronos0869 | *TdHsp26-A1* |
| K367-Fw1 | WT | CCGGCTGTTCGACGACG**C** | FAM | Manually | Kronos0367 | *TdHsp26-A1* |
| K367-Fw2 | SNP/Mutation | CCGGCTGTTCGACGACG**T** | VIC/HEX | Manually | Kronos0367 (C425T) | *TdHsp26-A1* |
| K367-Rev | Common | CCTCCTTCTCGTCCTCCATA | - | Manually | Kronos0367 | *TdHsp26-A1* |

### 3.1.5.3 SNP assay

The SNP assay was conducted for each genotype of the putative M4 Kronos and M3 Cham1 mutant lines to confirm the mutation and also the predicted zygosity. The DNA of the mutant and of the Kronos and Cham1 wild type plants were used in the analysis; two technical replicates were done for each genotype.

The KASP assay mix was prepared as recommended by LGC Genomics: 46 µl $dH_2O$, 30 µl of 100 µM common primer (Table 3.2) and 12 µl of each 100 µM allele-specific primers (Table 3.2). Assays were tested in 384-well format and set up with 2 µl mutants or wild type plants DNA [10–20 ng of DNA], 2 µl of 2x KASP Master mix (LGC Genomics, Teddington, UK), and 0.07 µl primer mix). PCR cycling was performed on a Eppendorf Mastercycler pro 384 using the following protocol: hot start at 94 °C for 15 mins, followed by 10 touchdown cycles (94 °C for 20 s; touchdown 65 °C, -1 °C per cycle, 60 s) then followed by 30 cycles of amplification (94 °C 20 s; 57 °C 60 s). Since KASP amplicons are predominantly smaller than 120 bp, an extension step is not needed in the PCR protocol. The plates were read on a Tecan Safire plate reader. Fluorescence was detected at room temperature. Data analysis was performed manually using Klustercaller software (version 2.22.0.5, LGC Genomics). If necessary additional amplification cycles were conducted and the sample where read again.

An example of KASP data plot is shown in Figure 3.3 where the FAM and VIC data are plotted on the *x* and *Y* axes, respectively. The inclusion of a passive reference dye (ROX) for the data normalization leads to tighter clustering and more accurate data calling. The genotypes can be determined according to the FAM and VIC signals distribution.

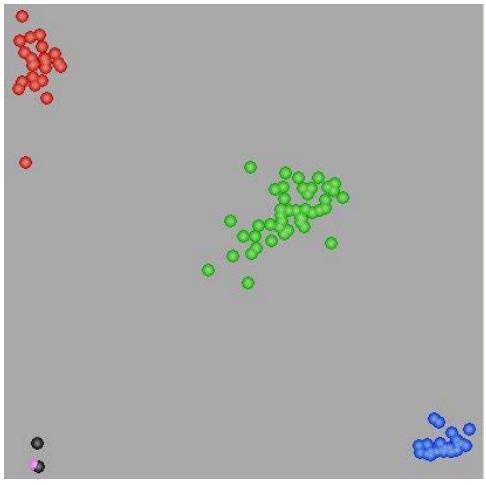


**Figure 3.3** Example of KASP data plotted using KBiosciences Klustercaller software with ROX normalization. In this example the FAM and VIC data are plotted on the *x* and *Y* axes, respectively, and thus the samples marked red are homozygous for the VIC allele, those marked blue are homozygous for the FAM allele, those marked green are heterozygous. This is a conventional colour attribution and is automatically assigned by the software but the colours could be also changed and manually assigned by the operator.

The wild type DNA (Cham1 or Kronos) were used as reference for the Cham1 or Kronos mutants, the genotype and the zygosity of each mutant line was evaluated according to the signals distribution obtained for the single genotypes within the line in respect to the wild type.

## 3.1.6 Phenotypical and physiological evaluation of *TdHsp26s* mutants

### 3.1.6.1 Phenotypical evaluation

The phenotype of the mutants was preliminary observed during the transplant at Z11-Z12 stage observing the number of the germinated seeds per lines or the deleterious phenotype of the seedlings. Then, a general observation of all the plants was done at tillering stage (Z21-Z25).

For a general phenotype evaluation the measurement of plant height (PH), number of spikes per plant (NSPP) and number of spikelets per spike (NSPS) were scored at late-flowering (Z65-Z69) for all the mutant and wild type plants.

### 3.1.6.2 Physiological evaluation: Chlorophyll fluorescence

For all the mutant and wild type plants the Chlorophyll fluorescence was measured at late-flowering (Z65-Z69) stage with Handy PEA (Plant Efficiency Analyzer) fluorimeter (Hansatech Instruments, UK). The measurement of fluorescent emission can be used to evaluate the photosynthetic performance (Krause and Weis 1991). In particular the instrument measure the Kautsky Fluorescence Induction (Kautsky *et al.* 1959), obtained immediately after the suddenly illumination of a dark adapted leaf sample due to the fluorescence quenching phenomenon caused by the closure of the Photosystem II (PSII) reaction center when a leaf is transferred from darkness into a light environment (Maxwell and Johnson 2000).

According to the manual of the instrument, three replicate measures were taken from the same leaf of each mutant and wild type plants after 40 mins of dark adaption. The measurements were taken in the glasshouse always at the same time of the day in the maximum light intensity hours (between 12 A.M. and 2 P.M.). From each measurement the Fv (Variable Fluorescence)/ Fm (Fluorescence Maximum), that is a widely accepted parameter indicating the maximum efficiency of PSII, was analysed.

# 3.2 Results and discussion

## 3.2.1 Online SNPs detection

The recently developed wheat online TILLING database (Uauy C. personal communication: www.wheat-tilling.com) was used to identify mutations in the three *TdHsp26* isolated genes. The EnsemblPlants accession Traes_4AS_8BA1E69CA, Traes_4AS_2272D0413 and IWGSC_CSS_4BL_scaff_6982788 were used as query to gain TILLING data for *TdHsp26-A1*, *TdHsp26-A2* and *TdHsp26-B1*, respectively.

The TILLING data retrieved from the database consist of a list of the Kronos lines carrying the SNPs in the chromosome or scaffold correspondent to the query Traes (and therefore to the target sequence). For each putative mutant line found the database predict the position of the mutation, the zigosity and the nucleotide and amino acid change. The full lists of Kronos mutant lines detected by the wheat TILLING database for *TdHsp26-A1*, *TdHsp26-A2* and *TdHsp26-B1* genes are provided in Table 3.3, Table 3.4 and Table 3.5, respectively.

The SNPs positions predicted in the database were manually verified on the *TdHsp26* genes and protein sequences. The predicted mutations were confirmed by aligning the mutant's sequences with the wild type. The type of mutation was scored and for the missense mutations the PSSM and SIFT values were calculated with PARSESNP software.

A total of 75 mutations were retrieved for all three genes, spanning the entire coding sequence (introns included) with different level of predicted zygosity (heterozygous or homozygous) (Table 3.3, Table 3.4 and Table 3.5).

**Table 3.3** Kronos lines carrying mutation on the *TdHsp26-A1* gene retrieved in the online wheat TILLING database using the Traes_4AS_8BA1E69CA as query.

[1] Chromosome position, confidence, SNP zigosity, consequence, nucleotide change and Amino Acid change have been predicted by the database.

[2] The position on *the TdHsp26-A1* gene, the nucleotide change, the position on gene structure, the Amino Acid Change have been verified manually for all the mutations.

[3] The mutations position in the protein domains were predicted on the basis of the alignment and the available information about HSP26 protein structure.

[4] the PSSM and SIFT values have been calculated on the *TdHsp26-A1* sequence with PARSESNP. The score indicating important effect are colored in red.

[5] The confidence score was automated (coverage of 4 is low, 5 is medium, 6 is high).

| Line | Chromosome position[1] | Position on *TdHsp26-A1* gene[2] | Confidence[1, 5] | Het/Hom[1] | Consequence[1] | Nucleotide change[1, 2] | Position on gene structure[2] | Amino Acid change[1, 2] | Domain[3] | PSSM[4] | SIFT[4] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Kronos1300 | 48083357 | 133 | high | het | missense_variant | Gcg/Acg | Ex I | A45T | N-term | | 0.04 |
| Kronos3076 | 48083354 | 136 | medium | het | missense_variant | Gcg/Acg | Ex I | A46T | N-term | | 1.0 |
| Kronos2193 | 48083349 | 141 | high | het | synonymous_variant | caG/caA | Ex I | Q47= | N-term | | |
| Kronos1179 | 48083340 | 150 | low | het | synonymous_variant | agA/agG | Ex I | R50= | N-term | | |
| Kronos3004 | 48083307 | 183 | high | het | synonymous_variant | gcC/gcT | Ex I | A61= | N-term | | |
| Kronos1179 | 48083298 | 192 | low | het | synonymous_variant | gcT/gcC | Ex I | A64= | N-term | | |
| Kronos2576 | 48083283 | 207 | high | het | synonymous_variant | ggC/ggT | Ex I | G69= | N-term | | |
| Kronos0670 | 48083237 | 253 | high | het | missense_variant | Ccg/Tcg | Ex I | P85S | MrD | 16.4 | 0.09 |
| Kronos2488 | 48083220 | 270 | high | het | intron variant | | INTRON | | | | |
| Kronos2202 | 48083129 | 361 | high | hom | missense_variant | Ccg/Tcg | Ex II | P91S | MrD | 19.7 | 0.01 |
| Kronos2644 | 48083065 | 425 | high | het | missense_variant | gCc/gTc | Ex II | A112V | N-term | | 0.05 |
| Kronos0367 | 48083065 | 425 | high | hom | missense_variant | gCc/gTc | Ex II | A112V | N-term | | 0.05 |
| Kronos0877 | 48083064 | 426 | high | het | synonymous_variant | gcC/gcT | Ex II | A112= | N-term | | |
| Kronos3913 | 48083055 | 435 | high | het | synonymous_variant | ttC/ttT | Ex II | F115= | N-term | | |
| Kronos2205 | 48083030 | 460 | high | hom | missense_variant | Gcg/Acg | Ex II | A124T | N-term | | 0.29 |
| Kronos2711 | 48083021 | 469 | high | het | missense_variant | Gag/Aag | Ex II | E127K | N-term | 0 | 0.92 |
| Kronos3223 | 48082992 | 498 | high | het | missense_variant | atG/atA | Ex II | M136I | ACD | 13.4 | 0.15 |
| Kronos3742 | 48082974 | 516 | high | hom | synonymous_variant | Gtg/gtA | Ex II | V142= | ACD | | |
| Kronos3052 | 48082899 | 591 | high | het | synonymous_variant | atC/atT | Ex II | I167= | ACD | | |
| Kronos0869 | 48082874 | 616 | high | hom | missense_variant | Ggc/Agc | Ex II | G176S | ACD | | 0.53 |
| Kronos3025 | 48082844 | 646 | low | het | missense_variant | Tgg/Agg | Ex II | W186R | ACD | | 0.39 |
| Kronos0243 | 48082830 | 660 | high | het | synonymous_variant | cgC/cgT | Ex II | R190= | ACD | | |
| Kronos0852 | 48082773 | 717 | high | hom | synonymous_variant | caG/caA | Ex II | Q209= | ACD | | |
| Kronos0596 | 48082765 | 725 | high | het | missense_variant | gCc/gTc | Ex II | A212V | ACD | 8.7 | 0.0 |
| Kronos2700 | 48082733 | 757 | high | hom | missense_variant | Gtg/Atg | Ex II | V223M | ACD | 6.3 | 0.01 |
| Kronos0572 | 48082723 | 767 | high | het | missense_variant | aGg/aAg | Ex II | R226K | C-term | 1.4 | 0.78 |
| Kronos2206 | 48082712 | 778 | high | het | missense_variant | Cgc/Tgc | Ex II | R230C | C-term | 12.5 | 0.00 |
| Kronos2006 | 48082694 | 796 | high | het | stop_gained | Cag/Tag | Ex II | Q236* | C-term | | |

**Table 3.4** Kronos lines carrying mutation on the *TdHsp26-A2* gene retrieved in the online wheat TILLING database using the Traes__4AS_2272D0413 as query.

[1] Chromosome position, confidence, SNP zigosity, consequence, nucleotide change and Amino Acid change have been predicted by the database.

[2] The position on *the TdHsp26-A2* gene, the nucleotide change, the position on gene structure, the Amino Acid Change have been verified manually for all the mutations.

[3] The mutations position in the protein domains were predicted on the basis of the alignment and the available information about HSP26 protein structure.

[4] the PSSM and SIFT values have been calculated on the *TdHsp26-A2* sequence with PARSESNP. The score indicating important effect are colored in red.

[5] The confidence score was automated (coverage of 4 is low, 5 is medium, 6 is high).

| Line | Chromosome position[1] | Position on *TdHsp26-A2* gene[2] | Confidence[1, 5] | Het/Hom[1] | Consequence[1] | Nucleotide change[1, 2] | Position on gene structure[2] | Amino Acid change[1, 2] | Domain[3] | PSSM[4] | SIFT[4] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Kronos2221 | 61082722 | 63 | low | het | synonymous_variant | ccG/ccC | EX I | P21= | TRANSIT | | |
| Kronos2517 | 61082719 | 66 | high | het | synonymous_variant | atC/atT | EX I | I22= | TRANSIT | | |
| Kronos3178 | 61082708 | 77 | medium | het | missense_variant | aGg/aAg | EX I | R26K | TRANSIT | -0.8 | 0.7 |
| Kronos2890 | 61082640 | 145 | high | het | stop_gained | Cag/Tag | EX I | Q49* | N-term | | |
| Kronos3846 | 61082632 | 153 | medium | hom | synonymous_variant | aaC/aaT | EX I | N51= | N-term | | |
| Kronos2991 | 61082632 | 153 | high | het | synonymous_variant | aaC/aaT | EX I | N51= | N-term | | |
| Kronos1061 | 61082629 | 156 | low | het | synonymous_variant | agG/agA | EX I | R52= | N-term | | |
| Kronos4250 | 61082629 | 156 | medium | hom | synonymous_variant | agG/agA | EX I | R52= | N-term | | |
| Kronos2453 | 61082626 | 159 | high | hom | synonymous_variant | gaC/gaT | EX I | D53= | N-term | | |
| Kronos0467 | 61082562 | 223 | high | het | missense_variant | Cgt/Tgt | EX I | R75C | N-term | 19.0 | 0.00 |
| Kronos3741 | 61082554 | 231 | high | het | synonymous_variant | ccG/ccA | EX I | P77= | N-term | | - |
| Kronos2901 | 61082547 | 238 | high | het | missense_variant | Gct/Act | EX I | A80T | N-term | 2.7 | 0.65 |
| Kronos0774 | 61082546 | 239 | high | het | missense_variant | gCt/gTt | EX I | A80V | N-term | 5.7 | 0.36 |
| Kronos3884 | 61082533 | 252 | high | het | synonymous_variant | atC/atT | EX I | I84= | N-term | | |
| Kronos3128 | 61082530 | 255 | high | het | synonymous_variant | tcC/tcT | EX I | S85= | N-term | | |
| Kronos2566 | 61082522 | 263 | high | hom | splice_donor_variant | | INTRON | | | | |
| Kronos0401 | 61082439 | 346 | low | het | splice_region_variant, intron_variant | | INTRON | | | | |
| Kronos0111 | 61082434 | 351 | high | het | splice_region_variant, synonymous_variant | ggG/ggA | EX II | G88= | MrD | | |
| Kronos0111 | 61082422 | 363 | high | het | synonymous_variant | ccG/ccA | EX II | P92= | MrD | | |
| Kronos0580 | 61082408 | 377 | low | het | missense_variant | aGg/aAg | EX II | R97K | MrD | 7.5 | 0.12 |
| Kronos0308 | 61082398 | 387 | low | het | synonymous_variant | cgG/cgA | EX II | R100= | MrD | | |
| Kronos2136 | 61082325 | 460 | high | hom | missense_variant | Gcg/Acg | EX II | A125T | N-term | | 0.23 |
| Kronos3001 | 61082316 | 469 | high | het | missense_variant | Gag/Aag | EX II | E128K | N-term | | 1.0 |
| Kronos2286 | 61082212 | 573 | high | het | synonymous_variant | gaG/gaA | EX II | E162= | ACD | | |
| Kronos3174 | 61082197 | 588 | high | het | synonymous_variant | gtC/gtT | EX II | V167= | ACD | | |
| Kronos2898 | 61082176 | 609 | medium | het | synonymous_variant | gaG/gaA | EX II | E174= | ACD | | |
| Kronos2468 | 61082134 | 651 | high | het | stop_gained | tgG/tgA | EX II | W188* | ACD | | |
| Kronos2913 | 61082091 | 694 | high | hom | missense_variant | Ccg/Tcg | EX II | P203S | ACD | 23.6 | 0.00 |
| Kronos2236 | 61082091 | 694 | high | het | missense_variant | Ccg/Tcg | EX II | P203S | ACD | 23.6 | 0.00 |
| Kronos3638 | 61082088 | 697 | high | hom | missense_variant | Gac/Aac | EX II | D204N | ACD | 16.0 | 0.00 |
| Kronos0440 | 61082083 | 702 | high | het | synonymous_variant | gaG/gaA | EX II | E205= | ACD | | |
| Kronos2574 | 61082062 | 723 | high | het | synonymous_variant | cgG/cgA | EX II | R212= | ACD | | |

**Table 3.5** Kronos lines carrying mutation on the *TdHsp26-B1* gene retrieved in the online wheat TILLING database using the Chinese Spring Sequence Survey Scaffold: IWGSC_CSS_4BL_scaff_6982788 as query.
[1] Confidence, SNP zigosity, consequence, nucleotide change and Amino Acid change have been predicted by the database.
[2] The position on *the TdHsp26-B1* gene, the nucleotide change, the position on gene structure, the Amino Acid Change have been verified manually for all the mutations.
[3] The mutations position in the protein domains were predicted on the basis of the alignment and the available information about HSP26 protein structure.
[4] the PSSM and SIFT values have been calculated on the *TdHsp26-B1* sequence with PARSESNP. The score indicating important effect are colored in red.
[5] The confidence score was automated (coverage of 4 is low, 5 is medium, 6 is high)

| Line | Position on *TdHsp26-B1* gene[2] | Confidence[1,5] | Het/Hom[1] | Consequence[1] | Nucleotide change[1, 2] | Position on gene structure[2] | Amino Acid change[1,2] | Domain[3] | PSSM[4] | SIFT[4] |
|---|---|---|---|---|---|---|---|---|---|---|
| Kronos4001 | 212 | low | het | missense_variant | gGc/gAc | Ex I | G71D | N-term | 15.3 | 0.01 |
| Kronos3628 | 259 | high | het | missense_variant | CcG/Tcg | Ex I | P87S | N-term | 15.5 | 0.13 |
| Kronos2574 | 266 | high | hom | intron variant | | INTRON | | | | |
| Kronos3113 | 301 | high | het | intron variant | | INTRON | | | | |
| Kronos0323 | 336 | high | het | intron variant | | INTRON | | | | |
| Kronos0866 | 398 | medium | hom | missense_variant | atG/atA | Ex II | M100I | MrD | 15.5 | 0.03 |
| Kronos0265 | 439 | medium | hom | missense_variant | gCc/gTc | Ex II | A114V | N-term | 9.1 | 0.28 |
| Kronos1308 | 444 | high | hom | missense_variant | Ggg/Agg | Ex II | G116R | N-term | | 0.11 |
| Kronos3487 | 446 | high | het | synonymous_variant | ggG/ggA | Ex II | G116= | N-term | | |
| Kronos3147 | 459 | high | het | missense_variant | Cgg/Tgg | Ex II | R121W | N-term | | 0.02 |
| Kronos3186 | 464 | medium | het | synonymous_variant | cgC/cgT | Ex II | R122= | N-term | | |
| Kronos1146 | 481 | high | het | missense_variant | aGc/aAc | Ex II | S128N | N-term | | 0.45 |
| Kronos2293 | 491 | high | het | synonymous_variant | ccG/ccA | Ex II | P131= | N-term | | |
| Kronos3996 | 660 | high | het | missense_variant | Ggc/Agc | Ex II | G188S | ACD | | 0.24 |
| Kronos1159 | 713 | high | het | synonymous_variant | tgC/tgT | Ex II | C210= | ACD | | |

For *TdHsp26-A1* a total of 28 mutations were detected (Table 3.3) of which only one was in the intron. Eleven were synonymous mutations, whereas 16 were classified as missense mutations. Importantly, one nonsense mutation (Q236*) was detected in the C-terminal domain. The retrieved missense mutations are reported on the TdHSP26-A1 protein scheme in Figure 3.4 where the secondary-structure motives and the functional domains are indicated. A total of 16 missense mutations were identified: two of them (P91S and P85S) were retrieved in the MrD domain which is important for the substrate binding (Waters 2013); five mutations were detected in the ACD: one in correspondence of the β2 sheet (M136I); one before the β6 (W186R); two in the β8 and β9 sheets (A212V and V223M) which are important for the oligomerization and in the hyper-variable region of the ACD domain, respectively. One mutation, A112V, was retrieved in two different mutant lines. Two mutations where detected in the C-terminal region (R230C, R226K) and 6 in the N-terminal region (Figure 3.4).

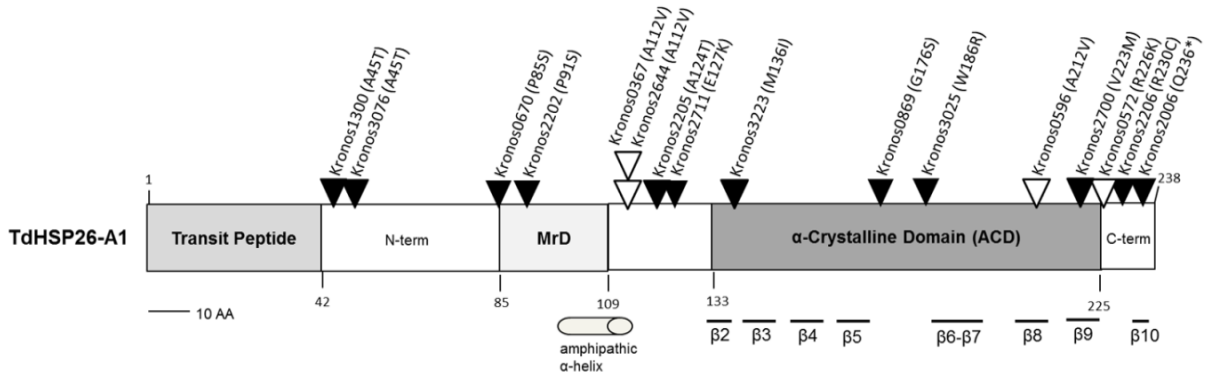



**Figure 3.4** Map of the missense mutations of the Kronos lines, detected by searching the online TILLING database, on the TdHSP26-A1 protein sequence. The biochemical important substitutions are indicated by the black arrows and the similar polarity substitution are indicated by white arrows. The functionally important secondary-structure motives are reported on the basis of the Promals3D structural alignment program. The transit peptide and the ACD domains are indicated in according to Chauhan *et al.* 2012 and the highly conserved Met-rich region (MrD) in the N-terminal domain is indicated in according to Sharf *et al.* 2001.

For *TdHsp26-A2*, a total of 32 mutations were detected (Table 3.4), with two located in the non-coding region; twelve were missense mutations. The retrieved missense mutations are reported on the TdHSP26-A2 protein scheme in Figure 3.5. The more interesting mutations were two nonsense mutations located in the N-terminal region (Q49*) and in correspondence of the ACD right before the β6 sheet (W188*). Moreover, 10 missense mutations were distributed across the N-terminal (5), transit peptide (1), within the MrD domain (R97K) and three (P203S, P203S and D204N) in the ACD both between the β7 and β8 sheets (Table 3.4, Figure 3.5). Again, two mutations were common in two different mutant genotypes.

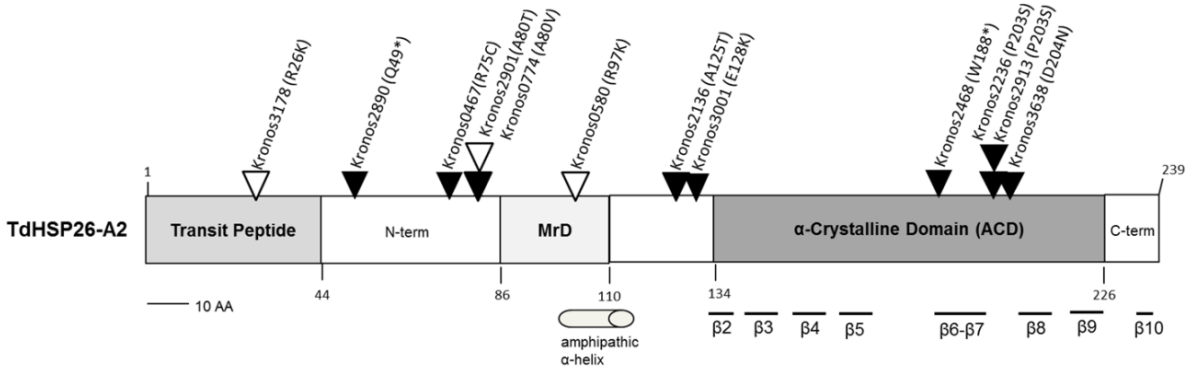



**Figure 3.5** Map of the missense mutations of the Kronos lines, detected by searching the online TILLING database, on the TdHSP26-A2 protein sequence. The biochemical important substitutions are indicated by the black arrows and the similar polarity substitutions are indicated by white arrows. The functionally important secondary-structure motives are reported on the basis of the Promals3D structural alignment program. The transit peptide and the ACD domains are indicated in according to Chauhan *et al.* 2012 and the highly conserved Met-rich region (MrD) in the N-terminal domain is indicated in according to Sharf *et al.* 2001.

For *TdHsp26-B1,* a total of 15 mutations were detected, of which 3 were located on the intron and 8 were missense mutations (Table 3.5). The retrieved mutations are reported on the TdHSP26-B1 protein scheme in Figure 3.6. Of the 12 mutations located on the exons: four were synonymous mutations and 8 were missense mutations located in the N-terminal domain (G71D, P87S, A114V, G116R, R121W, S128N) and within the MrD (M100I) and ACD (G188S) (Table 3.5, Figure 3.6). The lack of mutations in the 5' of the *TdHsp-B1* sequence is due to scaffold 6982788 missing the first 158 bp of the gene.

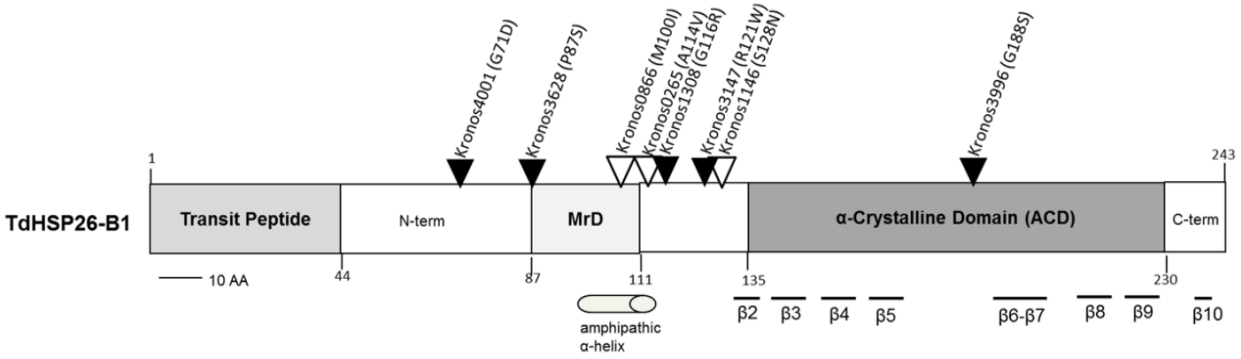



**Figure 3.6** Map of the missense mutations of the Kronos lines, detected by searching the online TILLING database, on the TdHSP26-B1 protein sequence. The biochemical important substitutions are indicated by the black arrows and the similar polarity substitutions are indicated by white arrows. The functionally important secondary-structure motives are reported on the basis of the Promals3D structural alignment program. The transit peptide and the ACD domains are indicated in according to Chauhan *et al.* 2012 and the highly conserved Met-rich region (MrD) in the N-terminal domain is indicated in according to Sharf *et al.* 2001.

For the most interesting mutations showing a SIFT value <0.05 and PSSM value >10, the correlation between the type of SNP, the position in the sequence and the prediction on the protein function (Table 3.3, Table 3.4 and Table 3.5) were also confirmed by analyzing the multiple alignment performed between annotated sHSP26 of *Triticum spp* and other dicots species (previously discussed in Chapter 2, Figure 2.5). In particular for TdHSP26-A1, the proline to serine substitution in position 85 (P85S) and the arginine to cysteine substitution (R230C) in position 230 of the β10 sheet hit highly conserved positions between *Triticum* species and within the dicots. Therefore a mutation in these positions may lead to a severe alteration of the protein function. A similar analysis was performed for TdHSP26-A2 where the proline in position 203 and the glutamine in position 204 of the CRI-ACD region show a high level of conservation between monocots and dicots. Finally, the methionine to isoleucine (M100I) in position 100 of the MrD of TdHSP26-B1 hit a high conserved region between monocots and dicots suggesting that this mutation could have interesting physiological consequence.

In summary, a total of 36 missense mutations in all genes, with a predicted effect on the protein function were identified from the wheat TILLING database: 30% fall in the ACD domain, 65% in the N terminal and of them 18% were localized in the MrD domain; only 9% were located in C-terminal.

## 3.2.2 Mutation detection by High Resolution Melting (HRM)

TILLING database search was associated with HRM analysis to detect mutations in the Cham1 TILLING population DNA library. *TdHsp26-B1*, that code for the most conserved sHSP26 among the *Triticeae*, was chosen as target for the HRM screening focusing on the functional domains as retrieved with COODLE (data not shown) with the aim to detect functional mutations in the target gene.

*TdHsp26-B1* specific primer pairs complementary to the 3' and 5' UTR (B1-17F/B1-6R; Figure 3.7A and Table 3.1) were designed for the first round PCR producing an amplicon of 1,378 bp; more internal specific primers (B1-PT10F/B1-PT10R, Figure 3.7A and Table 3.1) were designed for the second round PCR producing an amplicon of 211 bp suitable for HRM analysis that allowed to obtain a single peak in the derivative melting curves with the expected Tm= 89.9 °C (Table 3.1, Figure 3.7). The target region of the *TdHsp26-B1* in the HRM screening was therefore on the exon II, corresponding to a portion of the N-term, MrD and ACD.

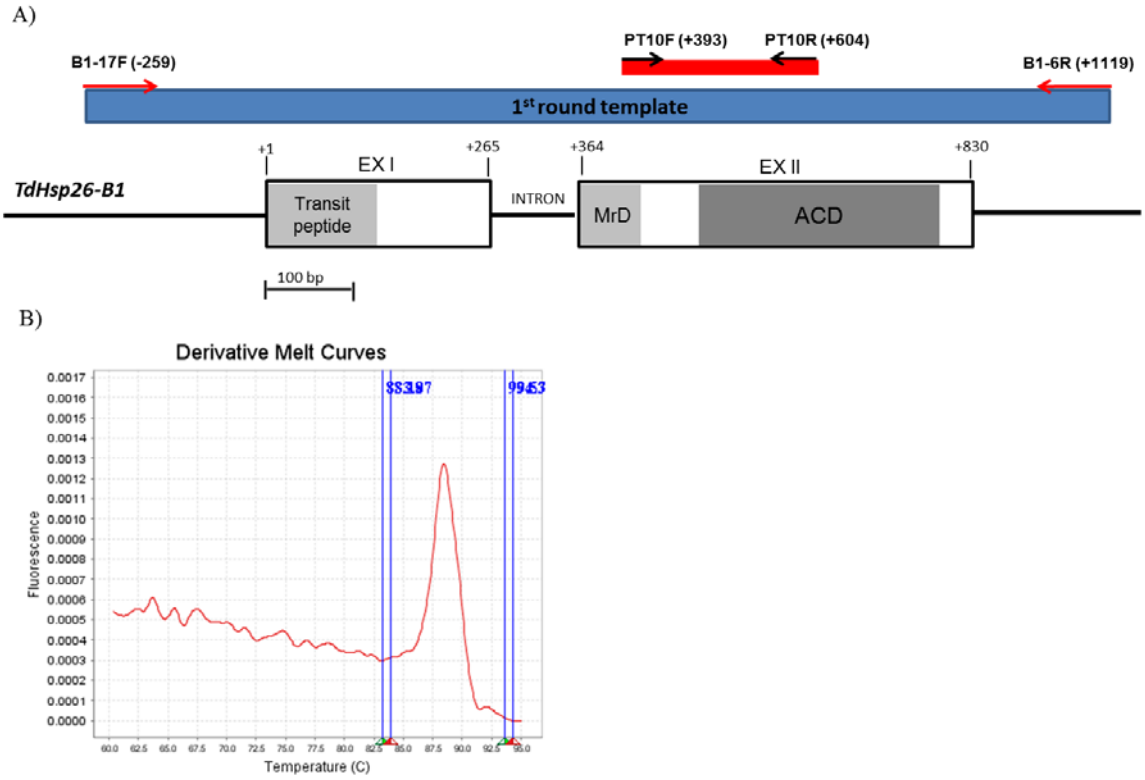



**Figure 3.7** (A) HRM primers position on *TdHSP26-B1* and (B) example of the single peak obtained with the selected primers for the wild type after the melt cycle. The primers pair B1-17F/B1-6R was used for the 1st PCR round while B1-PT10F/B1-PT10R primers pair was used for the 2nd round of the Cham1 TILLING library screening. The primers pair for the second round amplifies specifically a gene region that includes a portion of both the MrD and the ACD domains which are important for the protein function.

The screening of Cham1 TILLING library represented by 960 DNAs was conducted on a 2X pools allowing, in case of SNP presence, to generate eteroduplex in the temperature decrease step and to identify a variant curve after the melt curves analysis. After the 2nd round amplification and the melting step, the HRM software uses the default analysis settings to automatically assign a variant call to each sample. The software determines the variant calls by melt curves characteristics as melt curve shapes and Tm values.

As previously reported in other HRM analysis (Botticella *et al.* 2011) the presence of mutation in heterozygosis state usually cause a different shape of the melt curves, whereas homozygous mutation cause a Tm shift with respect to the wild type curve.

An example of the results obtained for a homozygote and heterozygote mutation detection with HRM analysis is reported in Figure 3.8A-B and C, respectively.

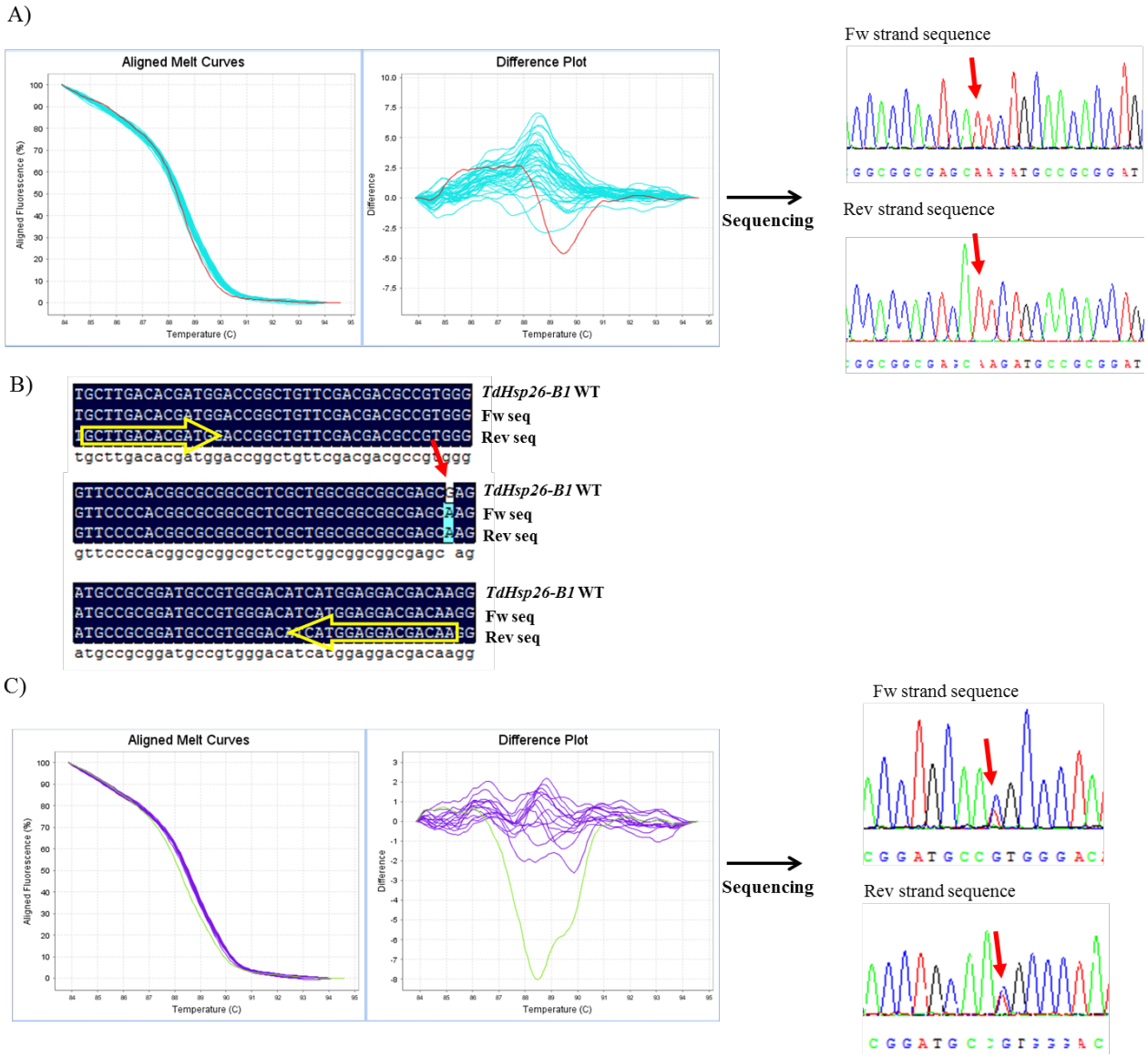



**Figure 3.8** Results of HRM analysis of detected mutations. A) Detection of an homozygote G/A mutation. One genotype within the positive pool of the red variant curves show after sequencing the presence of an homozygote G/A mutation (red arrow). Both 5' and 3' of the amplicon (Fw and Rev strands) were sequenced. B) The homozygote mutant sequence have been aligned with the wild type sequence of *TdHsp26-B1*, the mutation is indicated by the red arrow and the yellow arrows represent the primer used for the sequencing. C) Detection of an heterozygote G/A mutation. One genotype within the positive pool of the yellow variant curves show after sequencing the presence of an heterozygote G/A mutation (red arrows). The heterozygote mutation is observed as a double peak on the electropherogram of both strands.

From the sequence analysis of the positive pools 8 mutations were identified (Table 3.6), five of them were homozygotes and three were heterozygotes, moreover all mutations were transitions (G/C to A/T) as expected by the EMS treatment. Out of 8 mutation identified, 5 were silent and did not lead to amino acid substitutions, no truncations were detected and only three missense were found. These correspond to mutations M146I and P151L in the ACD while E129K was localized in the N-terminal domain, as indicated in the graphical representation (Table 3.6 and Figure 3.9).

PARSESNP analysis suggests that the three non-synonymous mutations are predicted to have severe effects on protein function (PSSM >10 and SIFT score <0.05) (Table 3.6).

**Table 3.6** List of the mutations identified by HRM in the Cham1 TILLING library. In the Amino Acid change columns, the first letter indicates the wild type amino acid, the number its position from the start codon/methionine, and the last letter the mutant amino acid. High PSSM (>10) and low SIFT score (<0.05) predict mutations with severe effects on the protein function**.**

| Mutant line | Position on *TdHsp26-B1* gene sequence | WT_base | Mut_base | Het/Hom | Consequence | Amino Acid change | Domain | PSSM | SIFT |
|---|---|---|---|---|---|---|---|---|---|
| W4-0181 | 483 | G | A | Hom | missense_variant | E129K | N-term | | 0.27 |
| W4-0841 | 425 | G | A | Het | synonymous_variant | R109= | N-term | | |
| W4-0844 | 536 | G | A | Hom | missense_variant | M146I | ACD | 18.4 | 0.02 |
| W4-1152 | 500 | G | A | Het | synonymous_variant | P134= | N-term | | |
| W4-0940 | 554 | G | A | Hom | synonymous_variant | G152= | ACD | | |
| W4-1771 | 550 | C | T | Hom | missense_variant | P151L | ACD | 25.6 | 0.00 |
| W4-1842 | 449 | C | T | Hom | synonymous_variant | F117= | N-term | | |
| W4-2013 | 563 | G | A | Het | Synonimous variant | R155= | ACD | | |

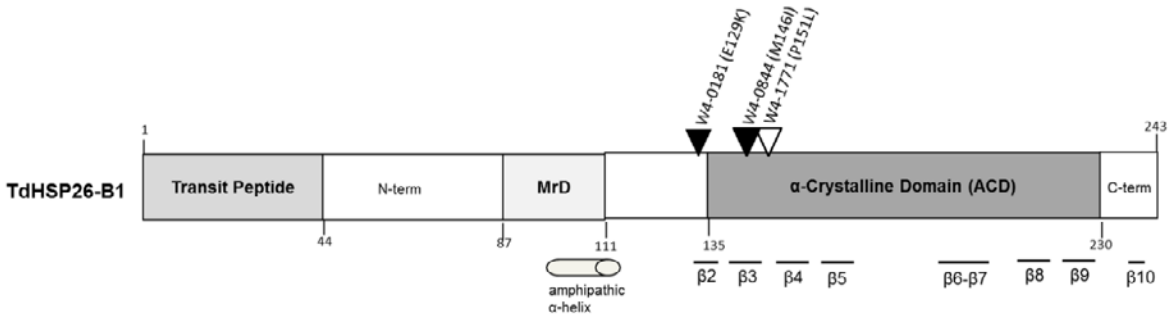



**Figure 3.9** Map of the missense mutations detected by HRM screening of the Cham1 TILLING library on the TdHSP26-B1 protein sequence. The biochemical important substitutions are indicated by the black arrows and the similar polarity substitutions are indicated by white arrow. The functionally important secondary-structure motives are reported on the basis of the Promals3D structural alignment program. The transit peptide and the ACD domains are indicated in according to Chauhan *et al.* 2012 and the highly conserved Met-rich region (MrD) in the N-terminal domain is indicated in according to Sharf *et al.* 2001.

Of particular interest are the M146I and P151L mutations localized respectively in the β3 sheet and in the CRII region of the ACD. Both the methionine and the proline are highly conserved in that positions among monocots and dicots species (Figure 3.9 and Figure 2.5) confirming the relevance of the missense mutations predicted by PARSESNP (Table 3.6).

### 3.2.3 SNPs validation using KASP markers

To further characterize the new identified alleles and to add more information on the gene function, a subset of the retrieved mutations detected by the database search and HRM screening were selected for the *in vivo* validation. The mutant lines were selected on the basis of the SNP position, the SIFT value and the level of conservation of the mutated amino acid. We first selected nonsense mutations independent of their position and secondly we selected missense mutations located in the ACD and MrD domains since these were strictly related with the substrate interaction. We first analysed seven Kronos mutants lines for *TdHsp26-A1* and a total of six lines for *TdHsp26-B1* (three Cham1 detected by HRM and three Kronos mutants lines identified through the database). The mutant lines selected for the *in vivo* validation are reported in Table 3.7 A, B. To confirm *in vivo* the presence of the detected SNPs, 37 KASP homeologous-specific markers were designed (Table 3.2) and tested on the M4 Kronos lines and on the M3 Cham1 lines at seedlings stage; in two cases the same common primers C0844_1771-Rev and K1308_265-Rev1 were used in two different SNP assays (Table 3.2). All the SNPs were confirmed with KASP analysis by evaluating the FAM and HEX fluorescence signal for each of the single genotypes. The distribution of the fluorescence signals within each line was analysed to confirm the predicted level of zygosity (Figure 3.10 and Figure 3.11).

**Table 3.7** Mutant lines selected for *in vivo* validation. A) Selected Kronos mutant lines carrying mutation on the *TdHsp26-A1* gene detected through TILLING database search. B) Selected mutant lines carrying mutation on the *TdHsp26-B1* gene, of these three Kronos mutant lines were detected through TILLING database search and three Cham1 mutant lines were detected through HRM. The SNP variation (mutation), the expected zygosity and the predicted effect on the protein are indicated.

A)

| | **Mutation on *TdHsp26-A1*** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Mutant line** | **Mutation** | **Het/Hom prediction** | **AA change** | **Domain** | **SIFT** | **PSSM** |
| ***In silico* detection** | Kronos0670 | G253A | Het | P85S | MrD | 0.09 | 16.4 |
| | Kronos2202 | G361A | Hom | P91S | MrD | 0.01 | 19.7 |
| | Kronos0367 | G425A | Hom | A112V | MrD | 0.05 | |
| | Kronos2205 | C460T | Hom | A124T | N-term | 0.29 | |
| | Kronos0869 | C616T | Hom | G176S | ACD | 0.53 | |
| | Kronos2206 | G778A | Het | R230C | C-term | 0.00 | 12.5 |
| | Kronos2006 | G796A | Het | Q236* | C-term | | |

B)

| | **Mutation on *TdHsp26-B1*** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Mutant line** | **Mutation** | **Het/Hom prediction** | **AA change** | **Domain** | **SIFT** | **PSSM** |
| ***In silico* detection** | Kronos0866 | G398A | Hom | M100I | MrD | 0.03 | 15.5 |
| | Kronos0265 | C439T | Hom | A114V | N-term | 0.28 | 9.1 |
| | Kronos1308 | G444A | Hom | G116R | N-term | 0.11 | |
| **HRM detection** | Cham1 W4-0181 | G483A | Hom | E129K | N-term | 0.27 | |
| | Cham1 W4-0844 | G563A | Hom | M146I | ACD | 0.02 | 18.4 |
| | Cham1 W4-1771 | C550T | Hom | P151L | ACD | 0.00 | 25.6 |

An example of the results obtained for four Kronos mutant lines is reported in Figure 3.10. Each KASP assay was used to test the correspondent mutants and the parental wild type. The genotypes of the Kronos mutant lines predicted as homozygous by the wheat TILLING database shown as expected an homozygous genotype, in this case the signals of all the genotype within the mutant line clusterized clearly separate from the wild type that are expected homozygous for the wild type nucleotide. An example of this type of results is shown in Figure 3.10 for the Kronos0866 and Kronos0367 lines. For the Kronos mutant lines predicted as heterozygous the expected segregation was observed within the lines with almost 50% heterozygous genotype, 25% homozygous for the wild type nucleotide and 25% homozygous for the SNP mutation. This type of results are shown for the Kronos0670 and Kronos2206 lines in Figure 3.10.

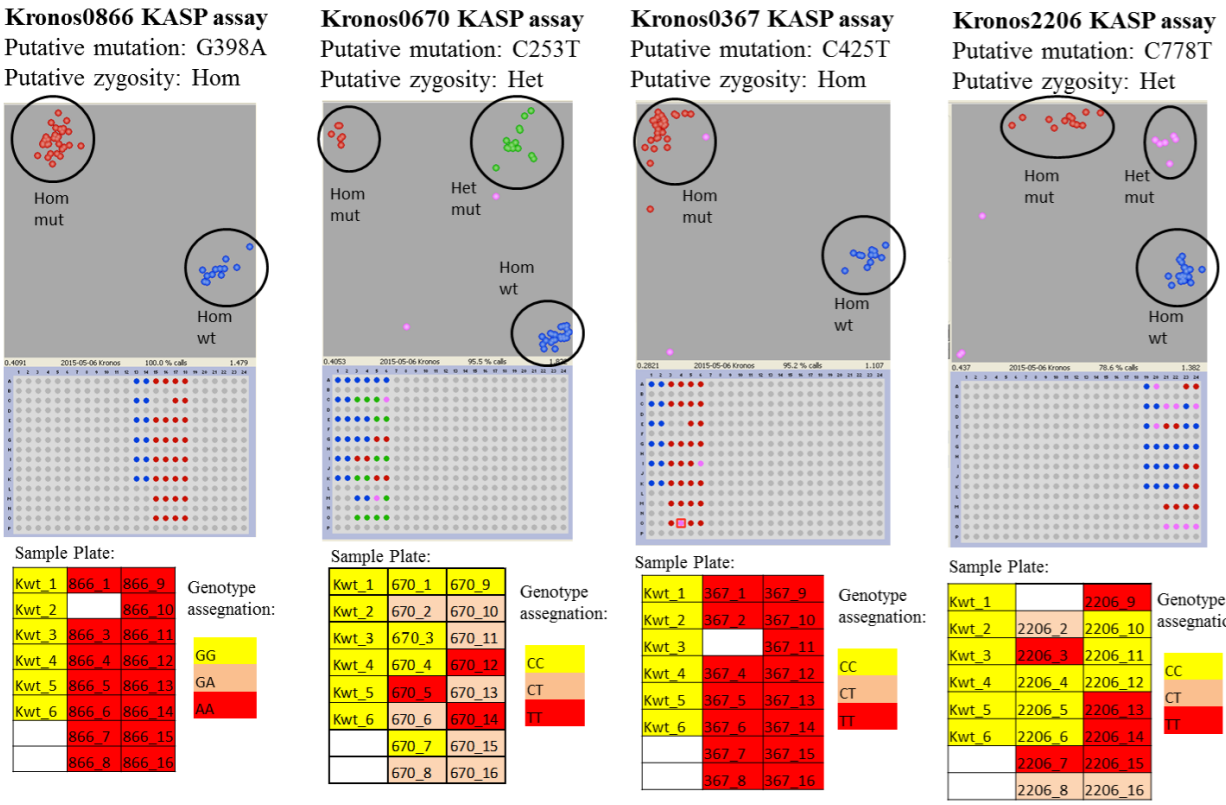



**Figure 3.10** KASP markers performed on the DNA of four Kronos mutant lines detected by searching the TILLING database. The Kronos0866 KASP assay containing K866-Fw1, K866-Fw2 and K866-Rev primers was used to test the Kronos0866 mutants and the Kronos wild type; the Kronos0670 KASP assay containing K670-Fw1, K670-Fw2 and K670-Rev primers was used to test the Kronos0670 mutants and the Kronos wild type. The Kronos0367 KASP assay containing K367-Fw1, K367-Fw2 and K367-Rev was used to test the Kronos0367 mutants and the Kronos wild type. The Kronos2206 KASP assay containing K2206-Fw, K2206-Rev1 and K2206-Rev2 was used to test the Kronos2206 mutants and the Kronos wild type. The putative mutation and zygosity detected *in silico* are reported.
For each result, the plots on the top shows the automatically obtained signal clusterization with the correspondent plate scheme below and the table on the bottom shows the order and the name of the plant tested (each sample in the plate was duplicated as technical replicate). The plants of each mutant line were named with the line number followed by consecutive number. The parental Kronos wild type were named as Kwt followed by consecutive number. The DNA of 6 wild type plants and of the germinated mutant plants were used in the analysis, the white boxes in the sample plate correspond to not germinated plants for which the DNA was missing. The genotypes were assigned on the basis of the signal clusterization (GG, Hom wt in yellow; GA, Het in pink; AA, Hom mut in red).

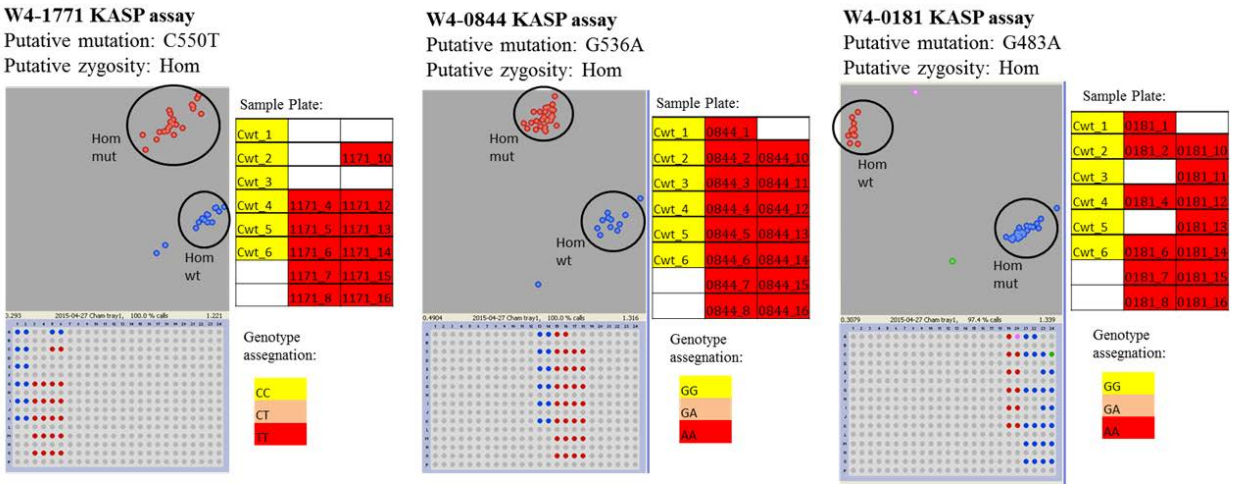



**Figure 3.11** KASP markers performed on the DNA of Cham1 W4-1771, W4-0844 and W4-0181 mutant lines detected by HRM. The W4-1771 KASP assay containing C1771-Fw1, C1771-Fw2 and C0844_1771-Rev primers was used to test the W4-1771 mutants and the Cham1 wild type; the W4-0844 KASP assay containing C0844-Fw1, C0844-Fw2 and C0844_1771-Rev primers was used to test the W4-0844 mutants and the Cham1 wild type. The W4-0181 KASP assay containing C0181-Fw, C0181-Rev1 and C0181-Rev2 was used to test the W4-0181 mutants and the Cham1 wild type. The mutation and zygosity detected by HRM are reported.
For each result, the plots on the left shows the automatically obtained signal clusterization with the correspondent plate scheme below and the table on the right shows the order and the name of the plant tested (each sample in the plate was duplicated as technical replicate). The plants of each mutant line were named with the line number followed by consecutive number. The parental Cham1 wild type were named as Cwt followed by consecutive number. The wells corresponding to the white boxes in the table didn't contain DNA. The genotypes were assigned on the basis of the signal clusterization (Hom wt in yellow; Het in pink; Hom mut in red).

A summary of the results is shown in Table 3.8 and Figure 3.12. For the Cham1 mutants line W4-1771, W4-0181, W4-0844 and Kronos0866, 1308, 0265, 2202, 2205, 0869 and 0367 the wheat TILLING database predicted a homozygose state for the specific mutation and the KASP assay confirms for all the screened plants an homozygous genotype. For the predicted heterozygous lines (Kronos0670, 2206, 2006) the expected segregation, with some slight distortion probably due to the low number of plants tested (Table 3.8, Figure 3.12), was found.

The SNP assay performed by KASP markers allow to confirm for all the selected mutant lines the predicted zygosity obtained from the online database search and to identify homozygous mutated genotype within the selected lines. This is a crucial step because only the homozygous genotype will be considered for the following crossing steps.

**Table 3.8** Results of the SNP assay performed by KASP markers on the mutant plants. A) For each mutant line a number between 11 and 16 plants were analysed, the *in silico* predicted zygosity is indicated and the *in vivo* zygosity of each plant was attributed according to the detected signal.

| Mutants line | Mutation | Predicted zygosity | N screened plants | Confirmed genotypes zygosity | | |
|---|---|---|---|---|---|---|
| | | | | Hom | Het | WT |
| Kronos0670 | C253T | Het | 16 | 3 | 8 | 5 |
| Kronos2202 | C361T | Hom | 16 | 13 | 0 | 0 |
| Kronos0367 | C425T | Hom | 15 | 15 | 0 | 0 |
| Kronos2205 | G460A | Hom | 15 | 15 | 0 | 0 |
| Kronos0869 | G616A | Hom | 16 | 15 | 0 | 0 |
| Kronos2206 | C778T | Het | 15 | 6 | 3 | 6 |
| Kronos2006 | C796T | Het | 11 | 3 | 3 | 5 |
| Kronos0866 | G398A | Hom | 15 | 15 | 0 | 0 |
| Kronos0265 | C439T | Hom | 15 | 15 | 0 | 0 |
| Kronos1308 | G444A | Hom | 13 | 13 | 0 | 0 |
| Cham1 W4-0181 | G483A | Hom | 13 | 13 | 0 | 0 |
| Cham1 W4-0844 | G536A | Hom | 15 | 15 | 0 | 0 |
| Cham1 W4-1771 | C550T | Hom | 11 | 11 | 0 | 0 |

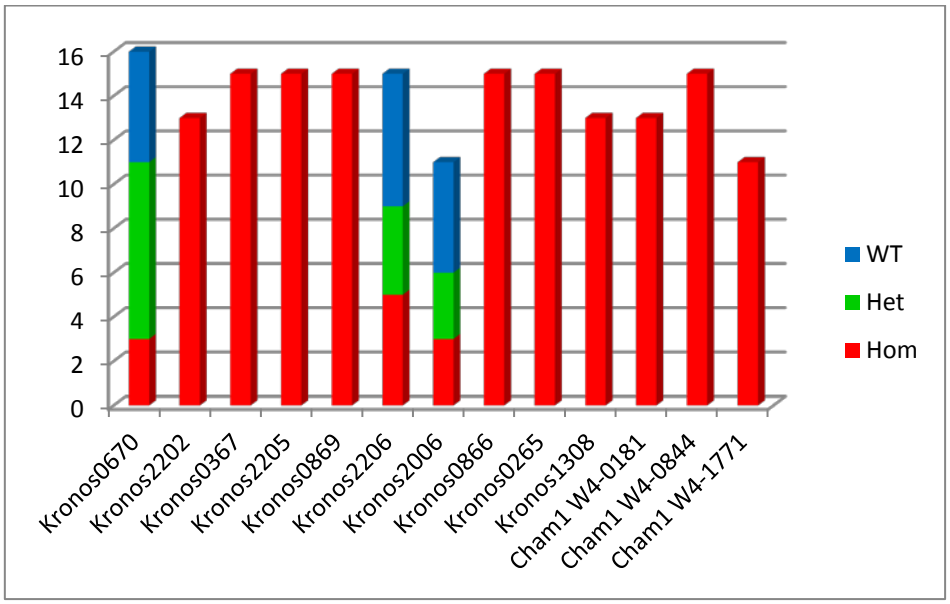



**Figure 3.12** Histogram of the KASP (SNP) assay results showing for each mutant line (x-axis) the genotype attribution for the number of plant tested (y-axis). Different colors indicate the genotype attribution with homozygous for the mutation SNP in red, heterozygous in green and homozygous for the wild type SNP (WT) in blue.

### 3.2.4 Preliminary phenotypical and physiological evaluation of the mutant and wild type lines

Even if the high unintended mutation rate at the M3 and M4 generation of the TILLING population do not allow the direct correlation of the retrieved mutation with the plant phenotype, a preliminary phenotypical evaluation of the validated mutant lines in respect to the wild type was done through the plant observation and the measurement of the plant height (PH), number of spike per plant (NSPP) and number of spikelets per spike (NSPS). The PH, NSPP and NSPS of the mutant and wild type plants grown in the glasshouse are reported in Figure 3.13 as average measure of the genotypes belonging to each mutant line.

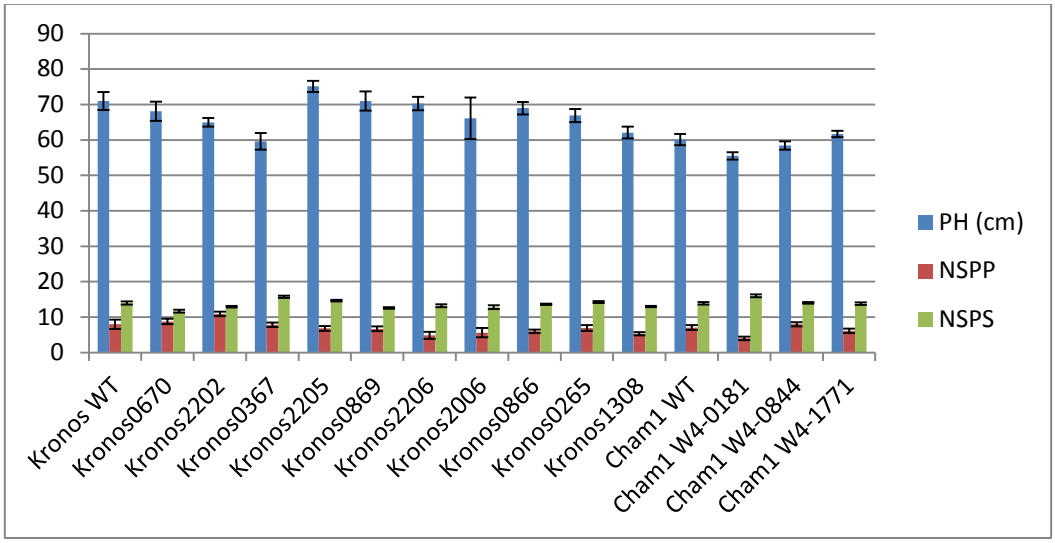



**Figure 3.13** Measure of plant height (PH), number of spike per plant (NSPP) and number of spikelets per spike (NSPS) of the Kronos and Cham1 mutant and wild type plants. The measure are reported for each mutant line and for the wild type as the average of the plants belonging to each line.

The average height in the Cham1 lines (the last four group of bars on the left of the histogram in Figure 3.13) in general is lower than most of the Kronos lines. All the measured traits were highly variable within the different Kronos mutant lines and also between the different Cham1 lines (Figure 3.13). The mutant plants shown variable plant height with respect to the correspondent wild type and also variable fertility, NSPP and NSPS of the mutant lines in comparison to the wild type were observed (Figure 3.13). Within the single plants, belonging to the same mutant line, high variability in the measured traits was observed, especially for the Kronos lines, as indicated by the high standard error deviation reported in the chart of Figure 3.13. As expected, the wide range of random mutations carried by each mutant line account for the huge differences observed in the phenotype of the analysed lines.

A high variability in the number of germinating seeds within each line was scored and suggest the presence of deleterious mutations within the genome (mainly in Kronos2006 and Cham1 W4-1771 for which only 11 plants have been grown and screened; Table 3.8). For the lines that have been confirmed heterozygous for the mutation in a *TdHsp26* gene, that are Kronos0670, Kronos2006 and Kronos2206 was possible to compare the Hom and Het mutant plants with the sibling plants having the same background mutations but different genotype for the mutation in the target gene. None of the observed mutant lines that have been confirmed as homozygous for a mutation in the target genes shown a uniform phenotype that could be related to the presence of the mutation of interest. Contrasting phenotypes were observed within sibling plants (Figure 3.14A) of the same mutant line. Kronos2206_10 confirmed as wild type for the mutation, show an albino phenotype and a slow growth. Moreover Kronos2206_15 and Kronos2206_3 that have been confirmed homozygous for the mutations, show different growth habits, the first monoculm and light green, the latter identical to a wild type Kronos plant with a bright green color and a lot of secondary culms. For some mutants lines no differences were observed if compared with the wild type, as shown for example in Figure 3.14B for Kronos0866_3.

The phenotype data so far collected suggest, as expected, that only preliminary observations can be done due to the high amount of unintended mutations intersperse in the genome. As conventionally assert, at least three cycles of backcross are a prerequisite to clean the background mutations and allow a direct correlation between the mutations and the observed phenotype.

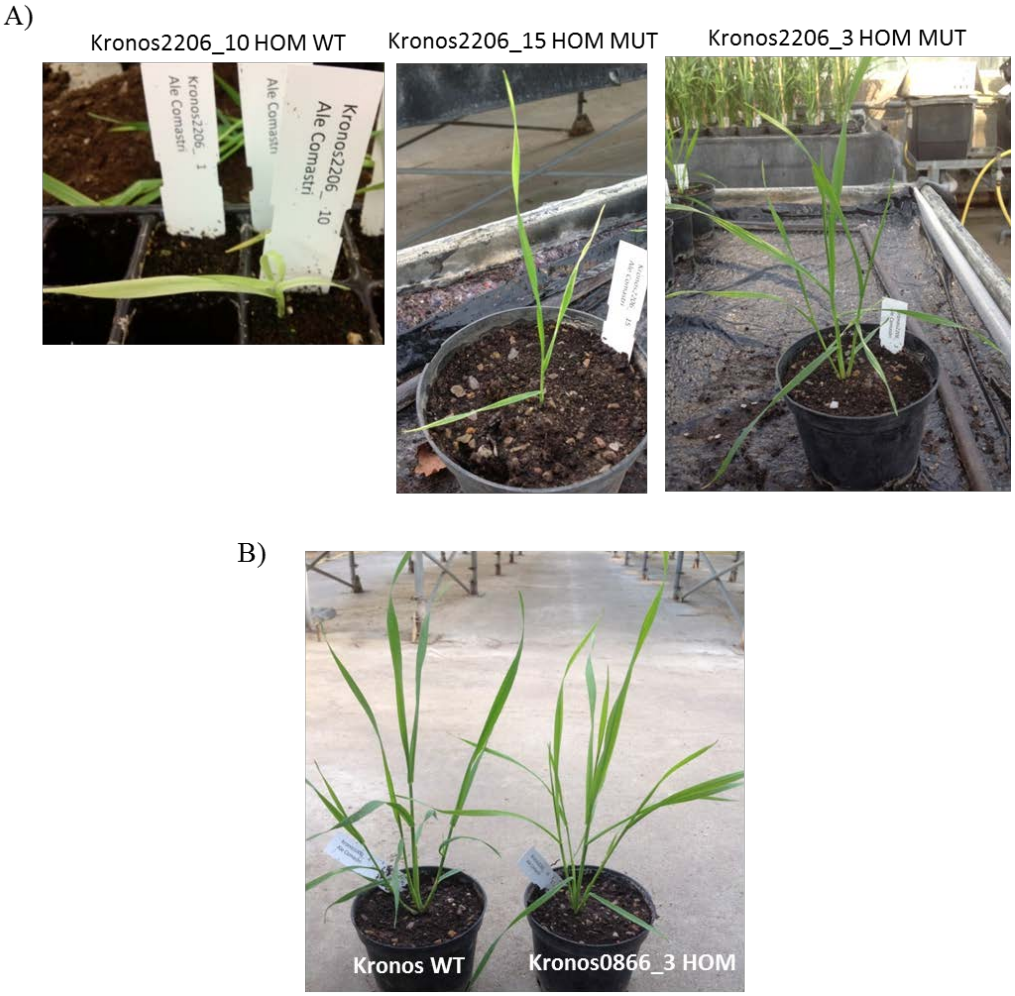



**Figure 3.14** Picture taken during the phenotype observation of the mutant and wild type plants. A) Four plants belonging to the Kronos2206 mutant lines (Kronos2206_10, Kronos2206_15 and Kronos2206_3) are reported. B) Evaluation of the mutant K0866_3 phenotype in respect of the Kronos wild type (WT). For each mutant plant the SNP zygosity confirmed by the SNP assay is indicated.

Nevertheless, to have an idea of the impact of the mutation event on the traits of interest, the mutant lines confirmed to carry the mutation of interest in the *TdHsp26-A1* and *TdHsp26-B1* genes were analysed for the maximum PSII efficiency, indicated by the Fv/Fm ratio, through the chlorophyll fluorescence measurement with the Handy PEA (Photosynthesis Efficiency Analyzer) Instrument (Hansatech Instruments, UK). The maximum photochemical efficiency (Fv/Fm) is considered to be a sensitive indication of the plant photosynthetic performance. As known in literature, healthy wheat sample achieve typically a minimum value of Fv/Fm value around 0.8 and a maximum of 0.85 and lower value were observed in presence of biotic or abiotic stress factor which reduce the capacity for photochemical quenching of energy within PSII (Akram *et al.* 2006, Lu and Zhang 2000). This parameter have been used in the phenotyping of wheat cultivars for heat tolerance and recently have been demonstrated that wheat cultivars selected for high Fv/Fm under heat stress maintain high photosynthesis performance (Sharma *et al.* 2012, Sharma *et al.* 2015). Moreover,

Fv/Fm was used, together with other parameters, to evaluate PSII function in recent studies on the over-expression of TdHSP26 in transgenic *Arabidopsis* plants (Chauhan *et al.* 2012) to evaluate the PSII activity under heat stress. Consistently to these evidence, the Fv/Fm evaluation could be considered an useful tools to evaluate the effect of the mutation in the function of the TdHSP26 and for this reason a preliminary evaluation of the mutant lines was done. The Fv/Fm value of each Kronos and Cham1 mutant lines and of the correspondent parental wild type is reported in Figure 3.15. In general all the Kronos mutant lines shown a lower value of Fv/Fm with respect to the Kronos wild type, but only Kronos2202, Kronos0265 and Kronos1308 lines shown statistically significative decrement in respect to the wild type. The Cham1 mutant lines W4-0181 and W4-1771 shown a decrement of Fv/Fm than the wild type, but only for the W4-0181 it was statistically significant; interestingly the Cham1 W4-0844 line shown a higher Fv/Fm than the wild type. Also for this parameters a high variability between the mutant lines and sometimes (mostly for Kronos0265) a high variability within the different genotype of the same line have been observed. This, as previously observed, could be due to the presence of numerous segregant mutations in the lines and since it was a preliminary test, also to the low number of replicate measurement.

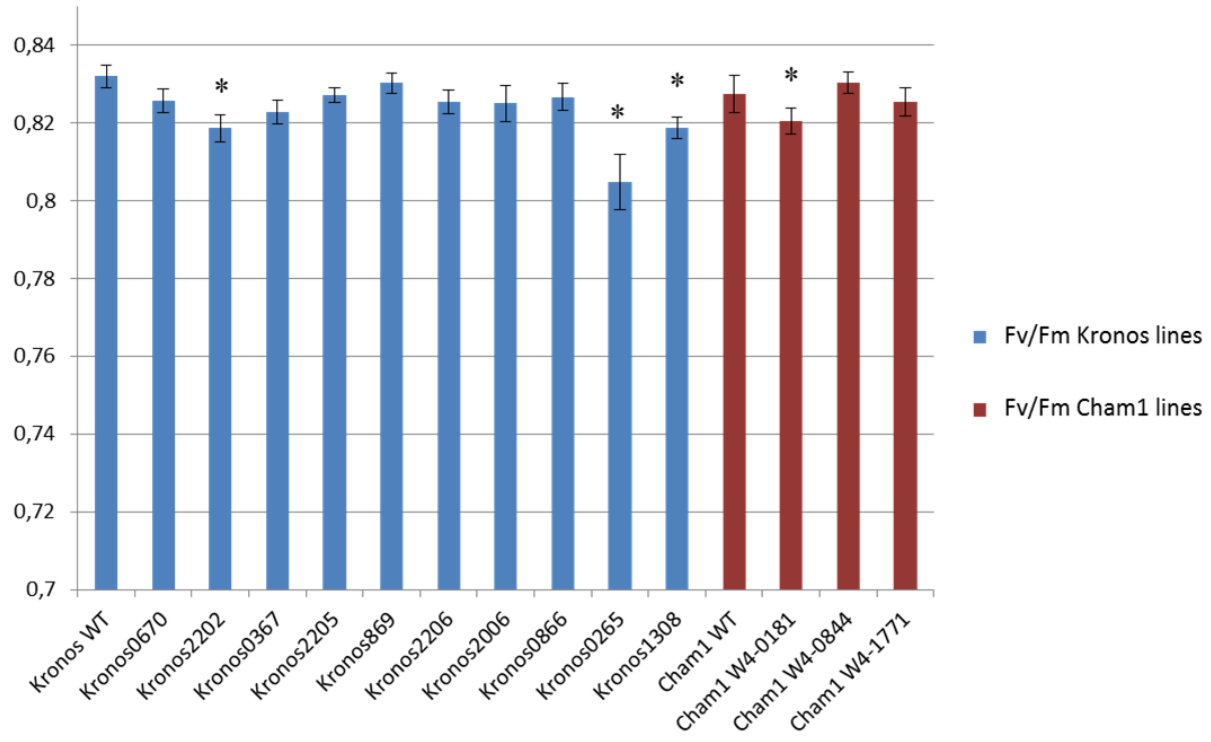



**Figure 3.15** Maximum photochemical efficiency measured as Fv/Fm ratio for the Kronos (blue bars) and Cham1 (red bars) mutant lines and wild type. Three replicates were measured for each plant and the value for each line calculated as the average of the plants of the same mutant line. Standard error bars are indicated. T-student TEST $P<0.05$. Fv= variable fluorescence, Fm= maximum fluorescence.

## 3.3 Conclusion

New allele variants for the wheat genes coding for the chloroplastic TdHSP26-A1, TdHSP26-A2 and TdHSP26-B1 have been successfully identified. In particular, the wheat TILLING database search allowed the identification of 16 mutant lines carrying missense mutations on the *TdHsp26-A1* gene, 12 mutant lines carrying missense mutations on the *TdHsp26-A2* gene and 8 mutant lines carrying missense mutations on the *TdHsp26-B1* gene in the Kronos TILLING population. Three more variants, leading to missense mutations, were identified in the Cham1 TILLING population for the *TdHsp26-B1* gene by conventional TILLING approach through HRM screening; of these mutations 1 falls in the N-term domain, known to be important for the substrate binding and 2 in the functionally important conserved ACD.

The advantage of the online database was a wide discovery of mutations across the entire genomic sequence, whereas the classical HRM screening, due to the difficulties in obtaining homoeologous-specific fragments generating good melt curves, allows to target only a small gene region and further effort are needed to extend the screening to the entire gene sequence. Even if with great time saving, the *in silico* detected allele variants needs an *in vivo* confirmation. The development of specific codominant KASP markers represent the key to fast validate *in vivo* the predicted mutations and their zygosity. Out of all the *in silico* detected variants: 7 lines carrying missense mutations for the *TdHsp26-A1* (of which 3 leading substitutions in the conserved MrD, 1 in the N-term, 1 in the ACD and 2 in the C-term with 1 stop codon) and 3 lines carrying missense mutations for the *TdHsp26-B1* (of which 1 leading substitutions in the conserved MrD and 2 in the N-term) have been confirmed by using the developed KASP markers. This lead to the identification of genotypes carrying interesting SNP variants in homozygose stase. The homozygous mutants identification opened the possibility to undertake backcrossings and also to plan a pyramiding approach, through crossing, for the generation of genotypes carrying multiple homoelogous mutations to assess, contemporarily, the global effect of *TdHsp26* variants on the analysed traits.

# Chapter 4. Work in progress: backcross of the most interesting *TdHsp26* mutant lines, double mutants production and $F_1$ genotyping by KASP markers

## 4.1 Material and methods

### 4.1.1 Crossing

The mutant lines confirmed to carry homozygous mutations in the *TdHsp26-A1* and *TdHsp26-B1* genes were used for backcrosses and to obtain the double mutants. Three lines of the Cham1 M3 TILLING population: W4-0181, W4-0844 and W4-1771; and ten lines of the Kronos M4 TILLING population Kronos 0670, 2202, 0367, 2205, 0869, 2206, 2006, 0866, 0265, 1308 were selected as previously described in Chapter 3 (Table 3.7A,B). From 1 to 6 spikes for each mutant line were backcrossed to generate lines with reduced background mutations; double mutant lines were also generated. The general scheme used for the backcrossing and the crossing between mutant lines is reported in Figure 4.1A and 4.1B respectively.

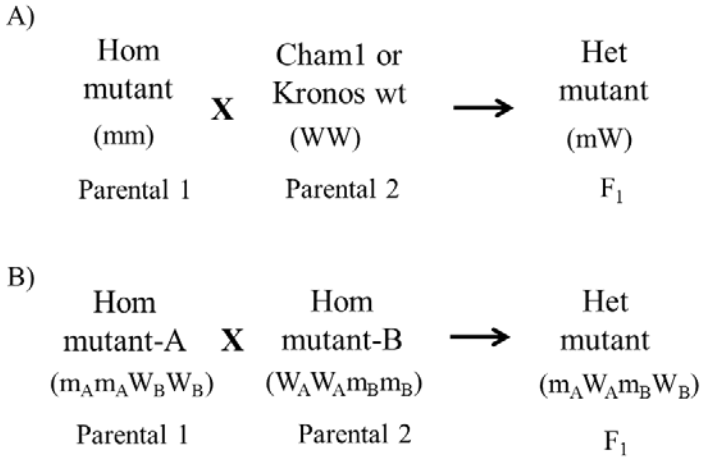



**Figure 4.1** General scheme of the backcrossing (A) and of the crossing between two different mutant lines carrying the mutation on the A or B copy of the *TdHsp26* gene (B). mm= homozygote for the mutant SNP; WW= homozygote for the wild type (wt) SNP; mW= heterozygote. $m_Am_A$ and $m_Bm_B$= homozygote of the mutant SNP on the A or B genome; $m_AW_Am_BW_B$= heterozygote for the mutant SNP on the A and B genome.

The emasculation and the subsequent pollination have been carryed out in the glasshouse and all the crossed spikes have been closed with a cellophane bag to avoid cross-contamination and seeds dispersion. At maturation, the $F_1$ spikes were cutted and dried at 30 °C for 1 month before seed harvesting. The backcrossed and crossed seeds were then sown to confirm the presence of the mutation with KASP markers in the SNP assay.

## 4.1.2 SNP assay by KASP markers

### 4.1.2.1 Vegetal material and DNA extraction

For each $F_1$ plants, 5 to 16 $F_1$ seeds from a single crossed spike were surface sterilized with sodium hypochlorite (0.5% v/v) for 30 mins, rinsed in sterile water and sown in Petri dishes in imbibition conditions at 6-8 °C for 10 days. After the germination seeds were transferred in a mix composed of 25% topsoil, 20% peat, 20% sand with 5% of silicate sand and grown in glasshouse under environmental temperature and light conditions and watered daily. 2-4 cm of Z12 seedling leaves were sampled for the DNA extraction. The same procedure was done for 8 Kronos wild type and 8 Cham1 wild type plants.

The genomic DNA was extracted for the SNP assay as reported in 3.1.5.1 paragraph.

### 4.1.2.2 SNP assay

The SNP assay was performed on the $F_1$ plants to test the presence of the mutation by using the previously generated KASP markers, reported in Chapter 3 (Table 3.2), specific for the target mutation in each mutant line. The KASP endpoint genotyping analysis was conducted on Real Time PCR 7900HT (Applied Biosystems) following the LGC guide "Guide to running KASP genotyping on the ABI7900 instrument". According to the guide a pre-read step and a PCR amplification step followed by the post-read for the data capture were performed. The assays were prepared in 96-wells microplates with optical read covers (Life Technologies) and set up with 5 µl mutants or wild-type DNA [10–20 ng of DNA], 5 µl of 2X KASP Master mix (LGC Genomics, Teddington, UK), and 0.14 µl primer mix. The amplification step was conducted with the following conditions: hot start at 94°C for 15 min, followed by 10 touchdown cycles (94 °C for 20 s; 61 °C for 60 s drop by -0.6 °C per cycle) then followed by 27 cycles of amplification (94 °C 20 s; 55 °C 60 s). Finally the temperature was decreased to 37 °C for 60 s and the post-read was performed. Additional recycling steps of amplification were performed until the obtaining of suitable genotyping clusters. Data analysis was performed using SDS software v2.4, the signal generated by the FAM and VIC primers were plotted on the *x* and *Y* axes to obtain an allelic discrimination plot. When possible the automatic assignation was considered, otherwise the data were manually analysed on the basis of the best clusterization plot. The wild-type DNA (Cham1 or Kronos) and the parental Hom mutant DNA were included in the analysis as reference. The genotype of each mutant was evaluated according to the signal distribution as reported in Chapter 3 (Figure 3.3). The putative double-mutant lines have been tested with two KASP markers, one for each mutation.

# 4.2 Results and discussion

## 4.2.1 Genotyping of the backcrossed and double mutant lines

A total of 21 $F_1$ lines were obtained through crossing, out of which: 13 putative backcrossed lines for *TdHsp26-A1* and *TdHsp26-B1* muant lines, and 8 different putative double mutant lines carrying mutations in the two homoeologous copy of the *TdHsp26* (Table 4.1). A full list of the 21 lines generated and tested by KASP markers are reported in Table 4.1.

To summarize, out of the 13 backcrossed lines obtained: 2 lines carried the mutation in the MrD domain of TdHSP26-A1 and -B1; 6 lines carried the mutation in the N-term domain and among these 3 are localized in TdHSP26-A1 and 3 in TdHSP26-B1; 3 lines carried the mutation in the ACD domain and 2 lines in the C-term domain of TdHSP26-A1. A new set of 8 double mutant lines for *TdHsp26-A1* and *B1* were generated; in particular: 2 lines carried the mutation in the MrD domain, 4 lines carried the mutation in the N-term domain and 2 lines carried the mutation in the ACD domain.

A different number of seeds were obtained for each of the generated lines ($F_1$); between 5 to 16 seeds for each lines were tested for the presence of the mutant SNP by KASP marker to verify the presence of the mutation in heterozygose state, as expected for the $F_1$ (Figure 4.1).

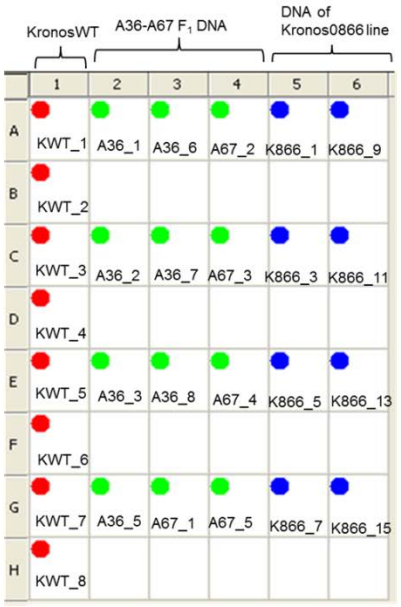
**Table 4.1** Full list of the $F_1$ mutant lines generated through crossing for *TdHsp26-A1* and *TdHsp26-B1* genes and genotyping results. The parental genotype, the mutated gene target and the domain hit by the mutation are indicated for each crossing event. The number of $F_1$ plants for each $F_1$ line that have been tested by KASP markers and the relative genotyping results are reported. For the double mutant lines the $F_1$ genotyping results are relative to the screening of the plants with the KASP markers for both the mutations (the Het attribution indicate the presence of both the mutation in heterozygosis).

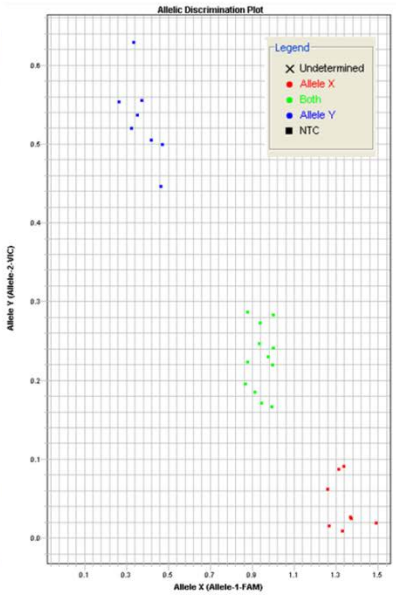
| **$F_1$ mutant lines** | **Crossed lines ♀ X ♂** | **Mutated gene** | **Domain hit by the mutation** | **N seeds obtained for each $F_1$ lines** | **N of $F_1$ plants tested by KASP** | **Genotyping results** | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | **Het mut** | **Hom mut** | **Not assigned** |
| A9 | Kronos0670 X WTKronos | *TdHsp26-A1* | MrD | 15 | 9 | **9** | **0** | **0** |
| A36 | Kronos0866 X WTKronos | *TdHsp26-B1* | MrD | 8 | 7 | **7** | **0** | **0** |
| A6 | Kronos2202 X WTKronos | *TdHsp26-A1* | N-term | 17 | 6 | **6** | **0** | **0** |
| A84 | Kronos0367 X WTKronos | *TdHsp26-A1* | N-term | 13 | 10 | **10** | **0** | **0** |
| A86 | Kronos2205 X WTKronos | *TdHsp26-A1* | N-term | 15 | 10 | 0 | 0 | 10 |
| A35 | Kronos0265 X WTKronos | *TdHsp26-B1* | N-term | 14 | 10 | 0 | 0 | 10 |
| A64 | Kronos1308 X WTKronos | *TdHsp26-B1* | N-term | 10 | 10 | **8** | **1** | **1** |
| A44 | Cham1W4-0181 X WTCham1 | *TdHsp26-B1* | N-term | 17 | 11 | **11** | **0** | **0** |
| A2 | Kronos0869 X WTKronos | *TdHsp26-A1* | ACD | 13 | 9 | **8** | **1** | **0** |
| A73 | Cham1W4-0844 X WTCham1 | *TdHsp26-B1* | ACD | 10 | 6 | **6** | **0** | **0** |
| A22 | Cham1W4-1771 X WTCham1 | *TdHsp26-B1* | ACD | 13 | 10 | **9** | **1** | **0** |
| A15 | Kronos2206 X WTKronos | *TdHsp26-A1* | C-term | 13 | 10 | **7** | **0** | **3** |
| A65 | Kronos2006 X WTKronos | *TdHsp26-A1* | C-term | 13 | 10 | **8** | **0** | **2** |
| A67 | Kronos0670 X Kronos0866 | *TdHsp26-A1 X TdHsp26-B1* | MrD X MrD | 5 | 5 | **5** | **0** | **0** |
| A68 | Kronos0866 X Kronos0367 | *TdHsp26-B1 X TdHsp26-A1* | MrD X N-term | 17 | 13 | **9** | **4** | **0** |
| A85 | Kronos0367 X Kronos0265 | *TdHsp26-A1 X TdHsp26-B1* | N-term X N-term | 7 | 7 | 0 | 0 | 7 |
| A80 | Kronos0367 X Cham1W4-0181 | *TdHsp26-A1 X TdHsp26-B1* | N-term X N-term | 6 | 6 | **6** | **0** | **0** |
| A46 | Kronos0265 X Kronos2202 | *TdHsp26-B1 X TdHsp26-A1* | N-term X N-term | 17 | 16 | 0 | 0 | 17 |
| A62 | Cham1W4-0181 X Kronos2205 | *TdHsp26-B1 X TdHsp26-A1* | N-term X N-term | 13 | 13 | 0 | 0 | 13 |
| A26 | Kronos0869 X Cham1W4-0844 | *TdHsp26-A1 X TdHsp26-B1* | ACD X ACD | 13 | 13 | **13** | **0** | **0** |
| A41 | Cham1W4-1771 X Kronos0869 | *TdHsp26-B1 X TdHsp26-A1* | ACD X ACD | 14 | 14 | **12** | **1** | **1** |

An example of the KASP genotyping for the $F_1$ double mutants line A46 (Kronos0670XKronos0866) and of the $F_1$ backcrossed lines A9 (Kronos0670XKronosWT) and A36 (Kronos0866XKronosWT) is reported in Figure 4.2 A and B.

As reported in Table 4.1, the KASP genotyping on the 13 backcrossed lines and on the 8 double mutant lines confirms the presence of the mutation of interest in heterozygosis for most of the genotype of each lines, with the only exception of the Kronos2205 and Kronos0265 mutant lines. In some cases homozygous mutant have been detected for some of the lines, indicating that the cross pollination between the mutant line and the recurrent parental line did not occur. A not clear clusterization was obtained for Kronos2205 and Kronos0265 preventing any reliable attribution.

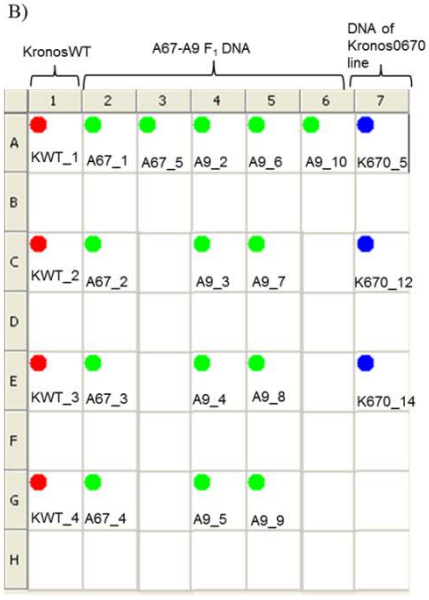
A)

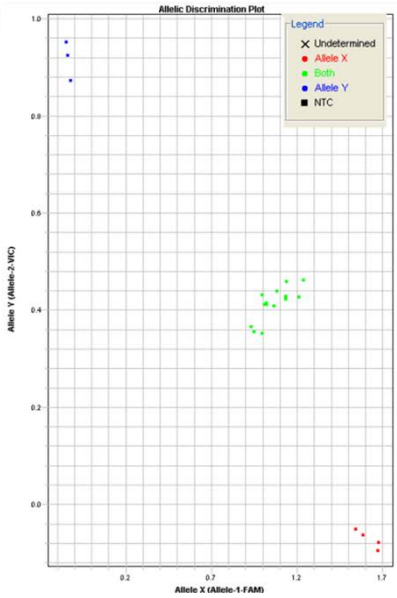



| Plant/DNA | KASP Assay | Genotype |
|---|---|---|
| KWT_1 | K866 Assay mix | HOM WT |
| KWT_2 | K866 Assay mix | HOM WT |
| KWT_3 | K866 Assay mix | HOM WT |
| KWT_4 | K866 Assay mix | HOM WT |
| KWT_5 | K866 Assay mix | HOM WT |
| KWT_6 | K866 Assay mix | HOM WT |
| KWT_7 | K866 Assay mix | HOM WT |
| KWT_8 | K866 Assay mix | HOM WT |
| A36_1 | K866 Assay mix | HET |
| A36_2 | K866 Assay mix | HET |
| A36_3 | K866 Assay mix | HET |
| A36_5 | K866 Assay mix | HET |
| A36_6 | K866 Assay mix | HET |
| A36_7 | K866 Assay mix | HET |
| A36_8 | K866 Assay mix | HET |
| A67_1 | K866 Assay mix | HET |
| A67_2 | K866 Assay mix | HET |
| A67_3 | K866 Assay mix | HET |
| A67_4 | K866 Assay mix | HET |
| A67_5 | K866 Assay mix | HET |
| K866_1 | K866 Assay mix | HOM MUT |
| K866_3 | K866 Assay mix | HOM MUT |
| K866_5 | K866 Assay mix | HOM MUT |
| K866_7 | K866 Assay mix | HOM MUT |
| K866_9 | K866 Assay mix | HOM MUT |
| K866_11 | K866 Assay mix | HOM MUT |
| K866_13 | K866 Assay mix | HOM MUT |
| K866_15 | K866 Assay mix | HOM MUT |

B)




| Plant/DNA | KASP Assay | Genotype |
|---|---|---|
| KWT_1 | K670 Assay mix | HOM WT |
| KWT_2 | K670 Assay mix | HOM WT |
| KWT_3 | K670 Assay mix | HOM WT |
| KWT_4 | K670 Assay mix | HOM WT |
| A67_1 | K670 Assay mix | HET |
| A67_2 | K670 Assay mix | HET |
| A67_3 | K670 Assay mix | HET |
| A67_4 | K670 Assay mix | HET |
| A67_5 | K670 Assay mix | HET |
| A9_2 | K670 Assay mix | HET |
| A9_3 | K670 Assay mix | HET |
| A9_4 | K670 Assay mix | HET |
| A9_5 | K670 Assay mix | HET |
| A9_6 | K670 Assay mix | HET |
| A9_7 | K670 Assay mix | HET |
| A9_8 | K670 Assay mix | HET |
| A9_9 | K670 Assay mix | HET |
| A9_10 | K670 Assay mix | HET |
| K670_5 | K670 Assay mix | HOM MUT |
| K670_12 | K670 Assay mix | HOM MUT |
| K670_14 | K670 Assay mix | HOM MUT |

**Figure 4.2** KASP genotyping for the $F_1$ mutant lines. The scheme of the plates (on the left), the correspondent allelic discrimination plot (on the center) and a list of the results for each plant (on the right) are reported. A) Genotyping of the double mutant line A67 and of the single mutant line A36 by using the Kronos0866 KASP assay containing K866-Fw1, K866-Fw2 and K866-Rev primers, the DNA of Kronos wildtype (Kwt) plants and the DNA of the Hom mutant parental line Kronos0866 (K866) have been included in the analysis. B) Genotyping of the double mutant line A67 and of the single mutant line A9 by using the Kronos0670 KASP assay containing K670-Fw1, K670-Fw2 and K670-Rev primers, the Kwt DNA and the DNA of the Hom mutant plants of the parental line Kronos0670 (K670) have been included in the analysis.

The genotyping results indicate that for most of the lines a successful crossing between mutant and parental lines occur and generate 11 backcrossed lines for the Kronos0670, 2202, 0367, 0869, 2206, 2006, 0866, 1308 lines and for the Cham1W4-0181, W4-0844 and W4-1771 lines. These lines represent the first step of the background cleaning of the detected mutant lines. Moreover the double mutant lines for *TdHsp26-A1* and *TdHsp26-B1* shown as expected the presence of both the mutation in heterozygosis. Of particular interest could be considered the A26 line (Kronos0869XCham1W4-0844) carrying the mutation in the ACD domains of the HSP26 involved in the oligomerization; and also the double mutant lines A80 (Kronos0367XCham1W4-0181) and A67 (Kronos0670XKronos0866) carrying the mutation in the N-terminal and in the conserved MrD respectively, that are important for the substrate binding.

## 4.3 Conclusion

The work presented in this last section of the thesis is still ongoing and concern the production of near isogenic lines (NILs) for *TdHsp26-A1* and *TdHsp26-B1* mutant lines carrying the mutations of interest in homozygose state in a cleaned background. The production of double mutants lines is fundamental to demonstrate the role played by *Hsp26* in durum wheat during heat stress.

The first step to achieve this goal is currently in progress. Several mutant lines for *TdHsp26-A1* and *TdHsp26-B1* have been backcrossed and double mutant lines have been generated.

Despite the necessity of several consecutive backcrosses to clean the background of the mutant lines prior to obtain functional knowledge, this resources represent the first step required to obtain NILs to understand the molecular mechanisms of the chloroplast small HSPs regarding their role in the thermotolerance acquisition.

A first ready to use product generated and tested in this work is represented by the KASP markers developed to follow the target mutations in the *TdHsp26* genes through generations in marker assisted selection (MAS).

# Chapter 5. General discussion and conclusions

An allele mining approach was used to retrieve new genetic variability in durum wheat. The overall objective of the work was to mine new alleles of the chloroplast *sHsp26* coding genes to investigate the role of HSP26 during wheat adaptation to heat stress and also to identify new resources for wheat breeding.

Three main results have been reached with this thesis work:

- Four *TdHsp26* genes have been identified and the complete genomic sequence have been characterized; their predicted protein sequences were confirmed as chloroplast localized sHSP26 on the basis of the conserved domains retrieved in the sequence.
- Three of the isolated genes were used as target in TILLING analysis to mine new *Hsp26* alleles in durum wheat; this was done by coupling a classical detection methods with an online database search to find the mutations. A wide number of SNPs variation that should differently affect the protein function have been identified for all genes, spanning the entire gene sequences.
- 13 mutants lines undergo the backcross process to obtain near isogenic lines (NILs) and 8 double mutant lines were generated to test for heat-related traits and eventually to be rapidly introgressed in breeding programs.

Taking all together these results represent a big step forward for the comprehension of the role played by *Hsp26* in the durum wheat adaptation to heat stress. The wide number of mutants identified open new perspectives for functional genomics studies on heat stress and represent the first step for the development of “non GM” NILs to be introduced in wheat breeding programs.

The production of new wheat varieties with increased thermotolerance and higher yield in heat stress conditions will help to address the rapid increases in global demand for agricultural commodities for food and animal feed.

# Web Sites

Cereals DataBase website: http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php

ChloroP 1.1: http://www.cbs.dtu.dk/services/ChloroP/

COODLE: http://blocks.fhcrc.org/proweb/input/

EnsemblPlants IWGSC 1.0+popseq: http://plants.ensembl.org/Triticum_aestivum/Info/Index

EnsemblPlants website: http://plants.ensembl.org/index.html

FAOSTAT http://faostat3.fao.org/home/E

Mutant Variety Database (MVD): http://mvd.iaea.org/

NCBI database BLAST: http://blast.ncbi.nlm.nih.gov

PolyMarker: http://polymarker.tgac.ac.uk/

PARSESNP: http://blocks.fhcrc.org/proweb/parsesnp/

PLACE (Plant Cis-acting Regulatory DNA Element) database: https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?lang=en&pj=640&action=page&page=newplace.

PlantCARE (Cis-Acting Regulatory Element)

database: http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

Primer3 v. 0.4.0: http://bioinfo.ut.ee/primer3-0.4.0/

PROMALS3D software online: http://prodata.swmed.edu/promals3d/promals3d.php

Softberry suites for gene finding in Eucaryota:

http://www.softberry.com/berry.phtml?topic=index&group=programs&subgroup=gfind

TargetP 1.1 Server: http://www.cbs.dtu.dk/services/TargetP-1.1/

TREP (the Triticeae Repeat Sequence Database): http://wheat.pw.usda.gov/ITMI/Repeats/

URGI (Unité de Recherche Génomique)

Blast: https://urgi.versailles.inra.fr/blast/?dbgroup=wheat_survey&program=blastn

WGS w7984_Meraculous Scaffolds BLAST

page: http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/blast_WGS.php

WheatBP: http://bio-gromit.bio.bris.ac.uk/cerealgenomics/WheatBP/Documents/DOC_WheatBP.php

Wheat TILLING database: http://www.wheat-tilling.com/

# References


Abe H., Urao T., Ito T., Seki M., Shinozaki K., Yamaguchi-Shinozaki K. (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. The Plant Cell, 15:63–78.

Abe H., Yamaguchi-Shinozaki K., Urao T., Iwasaki T., Hosokawa D., Shinozaki K. (1997) Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. The Plant Cell, 9: 1859-1868.

Agarwal M., Katiyar-Agarwal S., Sahi C., Gallie DR., Grover A. (2001) *Arabidopsis thaliana* Hsp100 proteins: kith and kin. Cell stress & chaperones, 6(3): 219-224.

Agarwal PK. and Jha B. (2010) Transcription factors in plants and ABA dependent and independent abiotic stress signaling. Biologia Plantarum 54: 201-212.

Akram M., Farooq S., Afzaal, M., Naz F., Arshad R. (2006) Chlorophyll fluorescence in different wheat genotypes grown under salt stress. Pakistan Journal of Botany, 38(5): 1739-1743.

Allakhverdiev I.S., Kreslavski D.V., Klimov V.V., Los A.D., Carpentier R., Mohanty P. (2008) Heat stress: an overview of molecular responses in photosynthesis. Photosynthesis research, 541-550 98:541–550.

Altenbach SB. (2012) New insights into the effects of high temperature, drought and post-anthesis fertilizer on wheat grain development. Journal of cereal science, 56:39-50.

Al-Whaibi MH. (2011) Plant heat shock proteins: a minireviews. Journal of King Saud University-Science, 23:139-150.

Andersen JR. and Lübberstedt T. (2003). Functional markers in plants. Trends in plant science, 8: 554–60.

Aoki K., Kragler F., Xoconostle-Cázares B., Lucas WJ. (2002) A subclass of plant heat shock cognate 70 chaperones carries a motif that facilitates trafficking through plasmodesmata. Proceedings of the National Academy of Sciences, 99(25): 16342-16347.

Bagge M., and Lübberstedt T. (2008) Functional markers in wheat: Technical and economic aspects. Molecular Breeding, 22(3): 319-328.

Balakirev ES. and Ayala FJ. (2003) Pseudogenes: are they "junk" or functional DNA? Annual review of genetics, 37:123-51.

Basha E., Friedrich KL., Vierling E. (2006) The N-terminal arm of small heat shock proteins is important for both chaperone activity and substrate specificity. Journal of biological chemistry, 281:39943–39952.

Basha EM., Waters ER., Vierling E. (1999) Triticum aestivum cDNAs homologous to nuclear-encoded mitochondrion- localized small heat shock proteins Plant Science, 141(1): 93-103.

Bate N. and Twell D. (1998) Functional architecture of a late pollen promoter : pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. Plant Molecular Biology, 37: 859–869.

Bolon YT., Haun WJ., Xu WW., Grant D., Stacey MG., Nelson RT., Orf, J. H. (2011). Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant physiology, 156(1): 240-253.

Bondino HG., Valle EM., Ten Have AT. (2012) Evolution and functional diversification of the small heat shock protein/$\alpha$-crystallin family in higher plants. Planta, 235:1299–1313.

Boston RS., Viitanen PV., Vierling E. (1996) Molecular chaperones and protein folding in plants. Plant Molecular Biology, 32: 191–222.

Botticella E., Sestili F., Hernandez-Lopez A., Phillips A., Lafiandra D. (2011) High resolution melting analysis for the detection of EMS induced mutations in wheat SbeIIa genes. BMC Plant Biology, 11(1):156.

Borrill P., Adamski N., Uauy, C. (2015) Genomics as the key to unlocking the polyploid potential of wheat. New Phytologist, 208(4): 1008-1022.

Bovina R., Brunazzi A., Gasparini G., Sestili F., Palombieri S., Botticella E., Lafiandra D., Mantovani P., Massi A. (2014) Development of a TILLING resource in durum wheat for reverse- and forward-genetic analyses. Crop and Pasture Science, 65 (1): 112-124.

Brown AH. (2010) Variation under domestication in plants: 1859 and today. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1552): 2523-2530.

Bucher P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. Journal of molecular biology, 212(4): 563-578.

Campbell JL. (1998) Cloning of novel low and high molecular weight heat shock genes in wheat. Thesis (M.S.)- Texas Tech University. Available at: https://ttu-ir.tdl.org/ttu-ir/handle/2346/19824

Carver BF. (2009) Wheat science and trade. Wiley, Danvers, p 569.

Chapman JA., Mascher M., Buluç A., Barry K., Georganas E., Session A., Strnadova V., Jenkins J., Sehgal S., Oliker L., Schmutz J., Yelick K. A., Scholz U., Waugh R., Poland JA., Muehlbaue GJ., Stein N., Rokhsar DS. (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biology, 16:26.

Chauhan H., Khurana N., Tyagi AK., Khurana JP., Khurana P. (2011) Identification and characterization of high temperature stress responsive genes in bread wheat (*Triticum aestivum L.*) and their regulation at various stages of development. Plant molecular biology, 75: 35-51.

Chauhan H., Khurana N., Nijhavan A., Khurana J.P., Khurana P. (2012) The wheat chloroplastic small heat shock protein (sHSP26) is involved in seed maturation and germination and imparts tolerance to heat stress. Plant, Cell and Environment, 35: 1912–1931.

Chen J. and Gustafson JP. (1997) Chromosomal rearrangement of wheat (*T. aestivum*) chromosome 4A by *in situ* hybridization. Chinese journal of genetics, 24: 39–47.

Chen L., Hao L., Parry MA., Phillips AL., Hu YG. (2014). Progress in TILLING as a tool for functional genomics and improvement of crops. Journal of integrative plant biology, 56(5): 425-443.

Choulet F., Alberti A., Theil S., Glover N., Barbe V., Daron J., Pingault L., Sourdille P., Couloux A., Paux E. (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science 345: 1249721.

Colbert T,. Till BJ., Tompa R., Reynolds S., Steine MN., Yeung AT., McCallum CM., Comai L., Henikoff S. (2001) High-throughput screening for induced point mutations. Plant Physiology, 126: 480–484.

Comai L. and Henikoff S. (2006) TILLING: practical single-nucleotide mutation discovery. The Plant Journal, 45(4): 684-694.

Comai L., Young K., Till BJ., Reynolds SH., Greene EA., Codomo CA., Enns LC., Johnson JE., Burtner C., Odden AR., Henikoff S. (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. The Plant Journal, 37(5): 778–786.

Curtis BC., Rajaram S., Macpherson HG. (2002) Bread Wheat: Improvement and Production (FAO Plant Production and Protection Series No. 30). Rome, FAO, pp. 1–19.

Devos KM., Dubcovsky J., Dvorák J., Chinoy CN., Gale MD. (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theoretical and Applied Genetics 91: 282–288.

Ding XZ., Tsokos GC., Kiang JG. (1998) Overexpression of HSP-70 inhibits the phosphorylation of HSF1 by activating protein phosphatase and inhibiting protein kinase C activity. FASEB J. 12(6): 451–459.

Dong C., Vincent K., Sharp P. (2009) Simultaneous mutation detection of three homoeologous genes in wheat by high resolution melting analysis and mutation surveyor ® . BMC Plant Biology 9(1): 143.

Dubcovsky J., Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science, 316: 1862–1866.

Dubcovsky J., Krasileva K., Vasquez-Gross H., Paraiso F., Howell T., Wang X,. Bailey PC., Ayling S., Uauy C. (2015) GENOMICS PLATFORMS FOR DURUM WHEAT. Proceedings of International Conference FROM SEED TO PASTA & BEYOND Bologna, Italy 31 May - 2 June 2015.

Dupont FM., Hurkman WJ., Vensel WH., Tanaka CK., Kothari KM., Chung OK., Altenbach SB. (2006) Protein accumulation and composition in wheat grains: effects of mineral nutrients and high temperature. European Journal of Agronomy 25:96–107.

Elshire RJ., Glaubitz JC., Sun Q., Poland JA., Kawamoto K., Buckler ES., Mitchell SE. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, 6(5), e19379.

Emanuelsson O., Nielsen H., Von Heijne G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science, 8(05), 978-984.

Emanuelsson O., Nielsen H., Brunak S., von Heijne G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology, 300(4), 1005-1016.

Eulgem T., Rushton PJ., Schmelzer E., Hahlbrock K., Somssich IE. (1999) Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors. The EMBO Journal, 18(17): 4689-4699.

Foley JA., Ramankutty N., Brauman K A., Cassidy ES., Gerber JS., Johnston M., Mueller DN., O'Connell C., Ray KD., West CP., Balzer C., Bennett ME., Carpenter RS., Hill J., Monfreda C., Polasky S., Rockström J., Sheehan J., Siebert S., Tilman D., Zaks PMD. (2011). Solutions for a cultivated planet. Nature, 478(7369), 337-342.

Frydman J. (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. Annual review of biochemistry 70, 603–647.

Gilmour SJ., Sebolt AM., Salazar MP., Everard JD., Thomashow MF. (2000) Overexpression of the *Arabidopsis* CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant Physiology 124:1854–1865.

Goyal RK., Kumar V., Shukla V., Mattoo R., Liu Y., Chung SH., Giovannoni JJ., Mattoo AK. (2012) Features of a unique intronless cluster of class I small heat shock protein genes in tandem with box C/D snoRNA genes on chromosome 6 in tomato (*Solanum lycopersicum*) Planta, 235 (3): 453-471.

Goymer P. (2007). Synonymous mutations break their silence. Nature Reviews Genetics, 8(2): 92-92.

Garcia M. and Mather ED. (2014) From Genes to Markers: Exploiting Gene Sequence Information to Develop Tools for Plant Breeding. Chapter2. In: Crop Breeding: Methods and Protocols, Methods in Molecular Biology, 1145: 21-36. Delphine Fleury and Ryan Whitford (eds). Springer New York.

Grigorova B., Vaseva II., Demirevska K., Feller U. (2011) Expression of selected heat shock proteins after individually applied and combined drought and heat stress. Acta Physiologiae Plantarum, 33(5): 2041-2049

Gupta PK., Mir RR., Mohan A., Kumar J. (2008) Wheat genomics: present status and future prospects. International Journal of Plant Genomics, ID896451:896451_1-896451_36.

Gullì M., Corradi M., Rampino P., Marmiroli N., Perrotta C. (2007) Four members of the HSP101 gene family are differently regulated in *Triticum durum* Desf. FEBS Letters, 581(25): 4841–4849.

Gururani, M.A., Mohanta, T.K., Bae, H. (2015) Current understanding of the interplay between phytohormones and photosynthesis under environmental stress International Journal of Molecular Sciences, 16(8): 19055-19085.

Gustavsson N., Kokke BPA., Anzeilius B., Boelens WC., Sundby C. (2001) Substitution of conserved methionines by leucines in chloroplast small heat shock protein results in loss of redox-response but retained chaperone-like activity. Protein Science, 10: 1785–1793.

Haake V., Cook D., Riechmann JL., Pineda O., Thomashow MF., Zhang JZ. (2002) Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiology, 130: 639–648.

Hamilton EW. and Heckathorn SA. (2001) Mitochondrial adaptations to NaCl. Complex I is protected by anti-oxidants and small heat shock proteins, whereas complex II is protected by proline and betaine. Plant Physiology, 126(3): 1266-1274.

Haralampidis K., Milioni D., Rigas S. & Hatzopoulos P. (2002) Combinatorial interaction of *cis* elements specifies the expression of the *Arabidopsis AtHsp90-1* gene. Plant Physiology, 129: 1138–1149.

Härndahl U., Hall RB., Osteryoung KW., Vierling E., Bornman JF., Sundby C. (1999) The chloroplast small heat shock protein undergoes oxidation-dependent conformational changes and may protect plants from oxidative stress. Cell Stress Chaperones, 4:129–138.

Hartl FU. (1996) Molecular chaperones in cellular protein folding. Nature 381: 571–580.

Hasanuzzaman M.., Nahar K., Alam MM., Roychowdhury R., Fujita M. (2013) Physiological, Biochemical, and Molecular Mechanisms of Heat Stress Tolerance in Plants. International Journal of Molecular Sciences, 14: 9643-9684.

Haslbeck M., Franzmann T., Weinfurtner D., Buchner, J. (2005) Some like it hot: the structure and function of small heat-shock proteins. Nature structural & molecular biology, 12(10): 842-846.

Hazard B., Zhang X., Colasuonno P., Uauy C., Beckles DM., Dubcovsky J. (2012) Induced mutations in the Starch Branching Enzyme II (SBEII) genes increase amylose and resistant starch content in durum wheat. Crop Science, 52 (4): 1754-1766.

Heckathorn SA., Downs CA., Sharkey TD., Coleman JS. (1998) The small, methionine-rich chloroplast heat-shock protein protects photosystem II electron transport during heat stress. Plant Physiology, 116(1): 439-444.

Henikoff S., Till BJ., Comai L. (2004) TILLING. Traditional mutagenesis meets functional genomics Plant Physiology, 135 (2): 630-636.

Henry IM., Nagalakshmi U., Lieberman MC., Ngo KJ., Krasileva KV., Vasquez-Gross H., Akhunova A., Akhunov E., Dubcovsky J., Tai TH., Comai L. (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using Exome capture and next-generation sequencing. The Plant Cell, 26 (4): 1382-1397.

Higo K., Ugawa Y., Iwamoto M., Korenaga T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic acids research, 27(1): 297-300.

Hu X., Yang Y., Gong F., Zhang D., Zhang L., Wu L., Li C., Wang W. (2015) Protein sHSP26 improves chloroplast performance under heat stress by interacting with specific chloroplast proteins in maize (*Zea mays*). Journal of Proteomics, 115: 81-92.

ICG, Grain market report (2014) http://www.thebioenergysite.com/reports/?category=33

IPCC (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

IWGSC (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science, 345(6194). doi: 10.1126/science.1251788.

Jantasuriyarat C., Vales MI., Watson CJW., Riera-Lizarazu O. (2004) Identification and mapping of genetic loci affecting the freethreshing habit and spike compactness in wheat (*Triticum aestivum L.*). Theoretical and Applied Genetics, 108: 261–273.

Jaglo KR., Kleff S., Amundsen KL., Zhang X., Haake V., Zhang JZ., Deits T., Thomashow MF. (2001) Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in Brassica napus and other plant species. Plant Physiology, 127: 910–917.

Jaya N., Garcia V., Vierling E. (2009) Substrate binding site flexibility of the small heat shock protein molecular chaperones. Proceedings of the National Academy of Sciences, 106(37): 15604-15609.

Jiao Y., Ma L., Strickland E., Deng XW. (2005) Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and *Arabidopsis*. The Plant Cell, 17(12): 3239-3256.

Jones E., Chu W-C., Ayele M., *et al.*. (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (Zea mays L.) germplasm. Molecular Breeding, 24: 165–76.

Jung SH. and Chory J. (2010) Signaling between chloroplasts and the nucleus: can a systems biology approach bring clarity to a complex and highly regulated pathway? Plant Physiology 152: 453–459.

Kadirvel P., Senthilvel S., Geethanjali S., Sujatha M., Varaprasad KS. (2015) Genetic Markers, Trait Mapping and Marker-Assisted Selection in Plant Breeding. In Plant Biology and Biotechnology, pp. 65-88. Springer India.

Kage U., Kumar A, Dhokane D, Karre S, Kushalappa AC. (2015). Functional molecular markers for crop improvement. Critical reviews in biotechnology, 1-14.

Kang JY., Choi HI., Im MY., Kim SY. (2002) *Arabidopsis* basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. The Plant Cell, 14(2): 343-357.

Kasuga M., Liu Q., Miura S., Yamaguchi-Shinozaki K., Shinozaki K. (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nature Biotechnology, 17: 287–291.

Kautsky H., Appel W., Amann H. (1959) Chlorophyll fluorescence and carbon assimilation. Part XIII. The fluorescence and the photochemistry of plants. Biochemische Zeitschrift, 332: 277-292.

Kersey JP., Allen EJ., Armean I., Boddu S., Bolt JB., Carvalho-Silva D., Christensen M., Davis P., Falin FL., Grabmueller C., Humphrey J., Kerhornou A., Khobova J., Aranganathan KN., Langridge N., Lowy E., McDowall DM., Maheswari U., Nuhn M., Ong KC., Overduin B., Paulini M., Pedro H., Perry E., Spudich G., Tapanari E., Walts B., Williams G., Tello–Ruiz M., Stein J., Wei S., Ware D., Bolser MD., Howe LK., Kulesha E., Lawson D., Maslen G., Staines MD. (2015) Ensembl Genomes 2016: more genomes, more complexity**.** Nucleic Acids Research, gkv1209.

Khurana N., Chauhan H., Khurana P. (2013) Wheat Chloroplast Targeted sHSP26 Promoter Confers Heat and Abiotic Stress Inducible Expression in Transgenic *Arabidopsis* Plants. PLoS ONE, 8(1), art. no. e54418.

King R., Bird N., Ramirez-Gonzalez R., Coghill JA., Patil A., Hassani-Pak K., Uauy C., Phillips AL. (2015) Mutation Scanning in Wheat by Exon Capture and Next-Generation Sequencing. PLoS ONE, 10(9).

Kirschner M., Winkelhaus S., Thierfelder JM., Nover L. (2000) Transient expression and heat-stress-induced co-aggregation of endogenous and heterologous small heat-stress proteins in tobacco protoplasts. The Plant Journal, 24(3): 397-412.

Kotak S., Larkindale J., Lee U., von Koskull-Döring P., Vierling E., Scharf KD. (2007) Complexity of the heat stress response in plants.Current opinion in plant biology, 10(3): 310-316.

Krause GH. and Weis E. (1991) Chlorophyll fluorescence and photosynthesis: the basics. Annual review of Plant Physiology and Plant Molecular Biology, 42: 313-249.

Krishna P. and Gloor G. (2001) The Hsp90 family of proteins in *Arabidopsis thaliana*. Cell Stress Chaperones, 6: 238–246.

Kumar P., Henikoff S., Ng PC. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocol, 4(7): 1073-81.

Kumar RR., and Rai RD. (2014) Can Wheat Beat the Heat: Understanding the Mechanism of Thermotolerance in Wheat (*Triticum aestivum L.*). Cereal Research Communications, 42(1): 1–18.

Kurowska M., Daszkowska-Golec A., Gruszka D., Marzec M., Szurman M., Szarejko I., Maluszynski M. (2011) TILLING - a shortcut in functional genomics. Journal of applied genetics, 52(4): 371-390.

Landsberger N. and Wolffe AP. (1995) Role of chromatin and *Xenopus laevis* heat shock transcription factor in regulation of transcription from the *X. laevis hsp70* promoter in vivo. Molecular and cellular biology, 15(11): 6013-6024.

Lata C., Yadav A., Prasad M. (2011) Chapter10. Role of Plant Transcription Factors in Abiotic Stress Tolerance. In: Abiotic Stress Response in Plants – Physiological, Biochemical and Genetic Perspectives, Edited by Arun Kumar Shanker and B. Venkateswarlu. InTech: 269-296.

Larkindale J., Hall JD., Knight MR, Vierling E. (2005) Heat stress phenotypes of *Arabidopsis* mutants implicate multiple signaling pathways in the acquisition of thermotolerance. Plant Physiology, 138(2): 882-897.

Lee JH., Hübel A., Schöffl F. (1995) Derepression of the activity of genetically engineered heat shock factor causes constitutive synthesis of heat shock proteins and increased thermotolerance in transgenic *Arabidopsis*. The Plant Journal, 8(4): 603-612.

Lee LS., Till B., Hill H., Huynh O, Jankowicz-Cieslak J (2014) Mutation and Mutation Screening. In: Henry RJ, Furtado A (eds) Cereal Genomics, vol 1099. Methods in Molecular Biology. Humana Press, pp 77-95.

Leone C., Perrotta C., Maresca B. (2003) Plant tolerance to heat stress: current strategies and new emergent insights. Chapter1. L.Sanità di Toppi and B.Pawlik-Skowrońska (eds.). Abiotic Stresses in Plants, pp 1-22.

Lescot M., Déhais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y., De Moor B., Rouzé P., Rombauts S. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic acids research, 30(1): 325-327.

Liu CJ., Devos KM., Chinoy CN., Atkinson MD., Gale MD. (1992) Homoeologous translocations between group 4, 5 and 7 chromosomes in wheat and rye. Theoretical and Applied Genetics, 83(3): 305–312.

Liu Y., He Z., Appels R., Xia X. (2012). Functional markers in wheat: current status and future prospects. Theoretical and Applied Genetics, 125(1): 1-10.

Lochlainn SO., Amoah S., Graham NS. Alamer K., Rios JJ., Kurup S., White PJ. (2011) High Resolution Melt (HRM) analysisis an efficient tool to genotype EMS mutants in complex crop genomes. Plant Methods, *7*(1): 1-9.

Lu CM. and Zhang JH. (2000) Heat-induced Multiple Effects on PSII in Wheat Plants. Journal of Plant Physiology, 156(2): 259-265.

Maestri E., Klueva N., Perrotta C., Gullì M., Nguyen HT., Marmiroli N. (2002) Molecular genetics of heat tolerance and heat shock proteins in cereals. Plant Molecular Biology, 48: 667-681.

Marmiroli N., Di Cola G., Komjanc M., Terzi V., Stanca AM., Martiniello P. Lorenzoni C. (1989a) Molecular and physiological parameters as aids in selection for temperature tolerance. In: R. Cavalloro and V. Delucchi (Eds.) Parasistis 88, Boletin de Sanidad Vegetal 17, pp. 49–55.

Marmiroli N., Lorenzoni C., Cattivelli L., Stanca AM., Terzi V. (1989b) Induction of heat shock proteins and acquisition of thermotolerance in barley (*Hordeum vulgare* L.). Variations associated with growth habit and plant development. Journal of plant physiology, 135(3): 267-273.

Marmiroli N., Maestri E., Terzi V., Gullì M., Pavesi A., Raho G., Lupotto E., Di Cola G., Sinibaldi R., Perrotta, C. (1994) Genetic and molecular evidences of the regulation of gene expression during heat shock in plants. In: J.H. Cherry (Ed.) Biochemical and Cellular Mechanisms of Stress Tolerance in Plants, NATO ASI Series H: Cell Biology 86, Springer-Verlag.

Martinez-Pastor MT., Marchler G., Schuller C., Marchler-Bauer A., Ruis H., Estruch F. (1996) The *Sacccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). The EMBO Journal, 15: 2227–2235.

Mascher M., Muehlbaue GJ., Rokhsar DS., Chapman JA., Schmutz J., Barry K., Munoz-Amatriain M., Close T., Wise R., Schulman A.H., Himmelbach A., Mayer KFX., Scholz U., Poland JA., Stein N., Waugh R. (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). The Plant Journal. 76: 718–727.

Maxwell K. and Johnson NG. (2000) Chlorophyll fluorescence-a practical guide. Journal of experimental botany, 51(345): 659-668.

McCallum CM., Comai L., Greene EA., Henikoff S. (2000) Choosing optimal regions for TILLING. Plant physiology, 123(2): 439-442.

McElwain EF. and Spiker S. (1989) A wheat cDNA clone which is homologous to the 17kd heat-shock protein gene family of soybean. Nucleic Acids Research, 17: 1764.

Mejlhede N., Kyjovska Z., Backes G., Burhenne K., Rasmussen SK., Jahoor A. (2006) Ecotilling for the identification of allelic variation in the powdery mildew resistance genes mlo and Mla of barley. Plant Breeding, 125: 461–467.

Metzker ML. (2010) Sequencing technologies -the next generation. Nature Reviews Genetics, 11: 31-46.

Meryem E., Mustapha L., Fatiha B., Mohamed A., Mouna T., Issa K., Perry M., Fatima G., Nessrelah N., Miloud N. (2014). Characterization of new allele variation for glutenin in EMS-mutant durum wheat population (*Triticum turgidum* L. subsp. *durum* (Desf.)). Journal of Life Sciences, 8: 880-888.

Mickelson-Young L., Endo TR. Gill BS. (1995) A cytogenetic ladder-map of the wheat homoeologous group *4* chromosomes. Theoretical and Applied Genetics, 90: 1007–1011.

Miftahudin, Ross K., Ma XF., Mahmoud AA., Layton J., Rodriguez Milla MA., Chikmawati T., Ramalingam J., Feril O., Pathan MS., Surlan Momirovic GS., Kim S., Chema K., Fang P., Haule L., Struxness H., Birkes J., Yaghoubian C., Skinner R., McAllister J., Nguyen V., Qi LL., Echalier B., Gill BS., Linkiewicz AM., Dubcovsky J., Akhunov ED., Dvořák J., Dilbirligi M., Gill KS., Peng JH., Lapitan NLV., Bermudez-Kandianis CE., Sorrells ME., Hossain KG., Kalavacharla V., Kianian SF., Lazo GR., Chao S., Anderson OD., Gonzalez-Hernandez J., Conley EJ., Anderson JA., Choi DW., Fenton RD., Close TJ., McGuire PE., Qualset CO., Nguyen HT., Gustafson JP. (2004) Analysis of expressed sequence tag loci on wheat chromosome group 4. Genetics, 168(2):651-663.

Mishra SK., Tripp J., Winkelhaus S., Tschiersch B., Theres K., Nover L., Scharf KD. (2002) In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. Genes & Development, 16(12): 1555-1567.

Mittler R., Finka A., Goloubinoff P. (2012) "How do plants feel the heat?." Trends in biochemical sciences, 37(3): 118-125.

Mogk A., Schlieker C., Friedrich KL., Schönfeld HJ., Vierling E., Bukau B. (2003) Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. Journal of Biological Chemistry, 278(33): 31033-31042.

Morimoto RI. (1998) Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes & Development, 12: 3788–3796.

Murata N., Takahashi S., Nishiyama Y., Allakhverdiev IS. (2007) Photoinhibition of photosystem II under environmental stress. Biochimica et Biophysica Acta 1767: 414–421.

Nakashima K., Fujita Y., Katsura K., Maruyama K., Narusaka Y., Seki M., Shinozaki K., Yamaguchi-Shinozaki K. (2006) Transcriptional regulation of ABI3- and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of *Arabidopsis*. Plant Molecular Biology, 60: 51–68.

Nalam VJ., Vales MI., Watson CJW., Kianian SF., Riera-Lizarazu O. (2006) Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum L.*). Theoretical and Applied Genetics, 112: 373–381.

Nalbandi K., Kohnehrouz BB., Saeed KA., Gholizadeh, A. (2014). Isolating Barley *(Hordeum vulgare L.*) B1 Hordein Gene Promoter and Using Sequencing Analaysis For The Identification of Conserved Regulatory Elements By Bioinformatic Tools. African Journal of Biotechnology, 11(29).

Naranjo T. (1990) Chromosome structure of durum wheat. Theoretical and Applied Genetics, 79: 397-400.

Narusaka Y., Nakashima K., Shinwari ZK., Sakuma Y., Furihata T., Abe H., Yamaguchi-Shinozaki K. (2003) Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis* rd29A gene in response to dehydration and high-salinity stresses. The Plant Journal, 34(2): 137-148.

Neelam K., Brown-Guedira G., Huang L. (2013) Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus Lr21. Molecular Breeding, 31: 233–237.

Nelson JC., Sorrells ME., Van Deynze AE., Lu YH., Atkinson M. (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics 141: 721–731.

Ng SB., Buckingham KJ., Lee C., Bigham AW., Tabor HK., Dent KM., Huff CD., Shannon PT., Jabs EW., Nickerson DA., Shendure J., Bamshad MJ. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nature genetics, 42(1): 30-35.

Ng SB., Turner EH., Robertson PD., Flygare SD., Bigham AW., Lee C., Bamshad M. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature, 461(7261): 272-276.

Nguyen HT., Weng J., Joshi CP. (1993) A wheat (*Triticum aestivum*) cDNA clone encoding a plastid-localized heat-shock protein. Plant Physiology 103: 675–676.

Nordin K., Vahala T., Palva, ET. (1993). Differential expression of two related, low-temperature-induced genes in *Arabidopsis thaliana (L.) Heynh*. Plant molecular biology, 21(4): 641-653.

Nover L., Bharti K., Doring P., Mishra S.K., Ganguli A., Scharf K.D. (2001) *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress & Chaperones 6, 177–189.

Oladosu Y., Rafii MY., Abdullah N., Hussin G., Ramli A., Harun A., Rahim A., Miah G., Usman M. (2015) Principle and application of plant mutagenesis in crop improvement: a review, Biotechnology & Biotechnological Equipment, 1-16.

Onodera Y., Suzuki A.,Wu C.Y.,Washida H. & Takaiwa F. (2001) A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. Journal of Biological Chemistry, 27: 14139-14152.

Osteryoung KW. and Vierling E. (1994) Dynamics of small heat shock protein distribution within the chloroplasts of higher plants. Journal of Biological Chemistry 269: 28676–28682.

Pandey B., Kaur A., Gupta OP., Sharma I., Sharma P. (2015) Identification of *HSP20* gene family in wheat and barley and their differential expression profiling under heat stress. Applied biochemistry and biotechnology, 175(5): 2427-2446.

Park HC., Kim ML., Kang YH., Jeon JM., Yoo JH., Kim MC., Yoon HW. (2004) Pathogen-and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. Plant Physiology, 135(4): 2150-2161.

Parry MAJ., Madgwick PJ., Bayon C., Tearall K., Hernandez-Lopez A., Baudo M., Rakszegi M., Hamada W., Al-Yassin A., Ouabbou H., Labhilili M., Phillips AL. (2009) Mutation discovery for crop improvement. Journal of Experimental Botany, 60(10): 2817-2825.

Peng JH., Sun D., Nevo E. (2011) Domestication evolution, genetics and genomics in wheat. Molecular Breeding, 28: 281–301.

Poland JA. and Rife TW. (2012) Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome, 5(3): 92-102.

Poulain P., Gelly JC., Flatters D. (2010) Detection and architecture of small heat shock protein Monomers. PLoS ONE, 5(4).

Prändl R., Hinderhofer K, Eggers-Schumacher G., Schöffl F. (1998) HSF3, a new heat shock factor from *Arabidopsis thaliana*, derepresses the heat shock response and confers thermotolerance when overexpressed in transgenic plants. Molecular and General Genetics MGG, 258(3): 269-278.

Qu A., Ding Y., Jiang Q., Zhu C. (2013) Molecular mechanisms of the plant heat stress response. Biochemical and Biophysical Research Communications, 432: 203–207.

Queitsch C., Hong SW., Vierling E., Lindquist S. (2000) Heat shock protein 101 plays crucial role in thermotolerance in *Arabidopsis*. The Plant Cell 12: 479–492.

Queitsch C., Sangster TA., Lindquist S. (2002) Hsp90 as a capacitor of phenotypic variation. Nature, 417: 618-624.

Rafalski A. (2002) Applications of single nucleotide polymorphisms in crop genetics. Current opinion in plant biology, 5(2): 94-100.

Ramirez-Gonzalez RH., Uauy C., Caccamo M. (2015) PolyMarker: A fast polyploid primer design Pipeline. Bioinformatics 1–2.

Ramirez-Gonzalez RH., Segovia V., Bird N., Fenwick P., Holdgate S., Berry S., Jack P., Caccamo M., Uauy C. (2014) RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant biotechnology journal, 13(5): 613-624.

Rampino P., Mita G., Assab E., De Pascali M., Giangrande E., Treglia AS., Perrotta C. (2009a) Two sunflower 17.6HSP genes, arranged in tandem and highly homologous, are induced differently by various elicitors. Plant Biology, 12:13–22

Rampino P., Mita G., Pataleo S., De Pascali M., Di Fonzo N., Perrotta C. (2009b) Acquisition of thermotolerance and HSP gene expression in durum wheat (*Triticum durum Desf.)* cultivars. Environmental and experimental Botany, 66(2): 257-264.

Rampino P., Mita G., Fasano P., Borrelli MG., Aprile A., Dalessandro G., De Bellis R., Perrotta C. (2012) Novel durum wheat genes up-regulated in response to a combination of heat and drought stress. Plant physiology and biochemistry, 56: 72-78.

Rawland JG., Pang X., Suzuki I., Murata N., Simon WJ., Slabas AR. (2010) Identification of components associated with thermal acclimation of photosystem II in Synechocystis sp. PCC6803. PLoS ONE 5(5), e10511.

Rezaei EE., Webber H., Gaiser T., Naab J., Ewert F. (2015) Heat stress in cereals: Mechanisms and modelling. European Journal of Agronomy, 64: 98-113.

Richter K., Haslbeck M., Buchner J. (2010). The heat shock response: life on the verge of death. Molecular cell, 40(2): 253-266.

Rieping M. and Schoffl F. (1992) Synergistic effect of upstream sequences, CCAAT box elements, and HSE sequences for enhanced expression of chimaeric heat shock genes in transgenic tobacco. Molecular & General Genetics, 231: 226–232.

Rigola D., van Oeveren J., Janssen A., Bonne A., Schneiders H., van der Poel HJ., van Orsouw NJ., Hogers RC., de Both MT., van Eijk MJ. (2009) High-throughput detection of induced mutations and natural variation using KeyPoint technology. PLoS ONE 4:e4761.

Ririe KM., Rasmussen RP., Wittwer CT. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Analytical Biochemistry 245: 154–160.

Risch NJ. (2000). Searching for genetic determinants in the new millennium. Nature, 405: 847–56.

Rocchi V., Janni M., Bellincampi D., Giardina T., D'Ovidio R. (2012) Intron retention regulates the expression of pectin methylesterase inhibitor (Pmei) genes during wheat growth and development. Plant Biology, 14: 365–373.

Rossel JB., Wilson IW., Pogson BJ. (2002) Global changes in gene expression in response to high light in *Arabidopsis*. Plant Physiology, 130: 1109–1120.

Roxas VP., Smith RK., Allen ER., Allen RD. (1997) Overexpression of glutathione S-transferase/glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress. Nature biotechnology, 15(10): 988-991.

Rushton PJ., Somssich IE., Ringler P., Shen QJ. (2010) WRKY transcription factors. Trends in plant science, 15(5): 247-258.

Saibil HR. and Ranson NA. (2002) The chaperonin folding machine. Trends in biochemical sciences, 27(12): 627-632.

Saidi Y., Finka A., Goloubinoff P. (2011) Heat perception and signalling in plants: a tortuous path to thermotolerance. New Phytologist 190: 556–565.

Sarkar NK., Kim YK., Grover A. (2009) Rice sHsp genes: genomic organization and expression profiling under stress and development. BMC Genomics 10(1): 393.

Sauna ZE. and Kimchi-Sarfaty C. (2011) Understanding the contribution of synonymous mutations to human disease. Nature Reviews Genetics, 12(10): 683-691.

Scharf KD., Siddique M., Vierling E. (2001) The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing alpha-crystallin domains Acd proteins. Cell Stress and Chaperones 6: 225–237.

Schirmer EC., Glover JR., Singer MA., Lindquist S. (1996) HSP100/Clp proteins: a common mechanism explains diverse functions. Trends in biochemical sciences, 21(8): 289-296.

Schlotterer C. (2004) The evolution of molecular markers—just a matter of fashion? Nature Reviews Genetics, 5(1): 63–69.

Schöffl F., Prändl R., Reindl A. (1998) Regulation of the heat-shock response. Plant Physiology, 117(4): 1135-1141.

Sears ER. (1966) Nullisomic–tetrasomic combinations in hexaploid wheat in: (Riley, R. and Lewis, K.R., Eds.), Chromosome Manipulations and Plant Genetics: The Contributions to a Symposium held during the Tenth International Botanical Congress, Edinburgh, 1964, p. vii, Plenum Press, New York.

Sears ER. and Sears MS. (1978) The telocentric chromosomes of common wheat. In: Proceedings of the 5th International Wheat Genetics Symposium of the Indian Society of Genetics and Plant Breeding (Ramanujam, S., Ed.), pp. 389–407, New Delhi, India.

Semagn K., Babu R., Hearne S., Olsen M. (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Molecular Breeding 33(1): 1–14.

Sestili F., Botticella E., Bedo Z., Phillips A., Lafiandra D. (2010) Production of novel allelic variation for genes involved in starch biosynthesis through mutagenesis. Molecular Breeding, 25(1): 145-154.

Sestili F., Palombieri S., Botticella E., Mantovani P., Bovina R., Lafiandra D. (2015) TILLING mutants of durum wheat result in a high amylose phenotype and provide information on alternative splicing mechanisms. Plant Science, 233: 127-133.

Svensson J.T., Crosatti C., Campoli C., Bassi R., Stanca A.M., Close T.J. & Cattivelli L. (2006) Transcriptome analysis of cold acclimation in barley albina and xantha mutants. Plant Physiology, 141: 257–270.

Sharma DK., Andersen SB., Ottosen CO., Rosenqvist E. (2012) Phenotyping of wheat cultivars for heat tolerance using chlorophylla fluorescence. Functional Plant Biology, 39(11): 936-947.

Sharma DK., Andersen SB., Ottosen CO., Rosenqvist E. (2015) Wheat cultivars selected for high Fv/Fm under heat stress maintain high photosynthesis, total chlorophyll, stomatal conductance, transpiration and dry matter. Physiologia plantarum, 153(2): 284-298.

Shakeel SN., Ul Haq N., Heckathorn S., Luthe DS. (2011) Analysis of gene sequences indicates that quantity not quality of chloroplast small HSPs improves thermotolerance in C4 and CAM plants. Plant Cell Reports, 31(10): 1943-1957.

Sharp P. and Dong C. (2014) TILLING for Plant Breeding. Chapter13. In: Crop Breeding: Methods and Protocols. Methods in Molecular Biology, 1145:155-165. Delphine Fleury and Ryan Whitford (eds). Springer New York.

Shewry PR. (2009) Wheat. Journal of Experimental Botany, 60(6): 1537–1553.

Simons KJ., Fellers JP., Trick HN., Zhang Z., Tai Y-S., Gill BS., Faris JD. (2006) Molecular characterization of the major wheat domestication gene Q. Genetics, 172(1): 547–555.

Slade AJ., Fuerstenberg SI., Loeffler D., Steine MN., Facciotti D. (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology, 23(1): 75–81.

Slade AJ., McGuire C., Loeffler D., Mullenberg J., Skinner W., Fazio G., Holm A., Brandt MK., Steine MN., Goodstal JF., Knauf VC. (2012) Development of high amylose wheat through TILLING.BMC plant biology, 12(1): 69.

Solovyev V., Kosarev P., Seledsov I., Vorobyev D. (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biology 7(Suppl 1): S10.

Soreng RJ., Peterson PM., Romaschenko K., Davidse G., Zuloaga FO., Judziewicz EJ., Morrone O. (2015) A worldwide phylogenetic classification of the Poaceae (Gramineae). Journal of Systematics and Evolution, 53(2): 117-137.

Sorrells ME., Gustafson JP., Somers D., Chao S., Benscher D., Guedira-Brown G., Huttner E., Kilian A., McGuire PE., Ross K., Tanaka J,. Wenzl P,. Williams K,. Qualset OC. (2011) Reconstruction of the Synthetic W79849 Opata M85 wheat reference population. Genome, 54: 875–882.

Stalberg K., Ellerstom M., Ezcurra I., Ablov S., Rask L. (1996) Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. Planta, 199: 515–519.

Stapper M. (2007) Crop monitoring and Zadoks growth stages for wheat. CSIRO Plant Industry. Canberra, ACT. Available at:
http://www.biologicagfood.com.au/wp-content/uploads/STAPPER-Crop-Monitoring-v21.pdf

Storozhenko S., Pauw PD., Montagu MV., Inze D., Kushnir S. (1998) The Heat-Shock element is a functional component of the *Arabidopsis* APX1 gene promoter. Plant Physiology 118: 1005–1014.

Sun W., Montagu MV., Verbruggen N. (2002) Small heat shock proteins and stress tolerance in plants. Biochimica et Biophysica Acta 1577(1): 1–9.

Sun C., Palmqvist S., Olsson H., Borén M., Ahlandsberg S., Jansson, C. (2003). A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. The Plant Cell, 15(9): 2076-2092.

Sun Q., Gao F., Zhao L., Li K., Zhang J. (2010). Identification of a new 130 bp cis-acting element in the TsVP1 promoter involved in the salt stress response from Thellungiella halophila. BMC plant biology, 10(1): 90.

Sundby C., Harndahl U., Gustavsson N., Ahrman E., Murphy DJ. (2005) Conserved methionines in chloroplasts. Biochimica et Biophysica Acta, 1703: 191–202.

Tai TH. and Tanksley SD. (1991) A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. Plant Molecular Biology Reporter, 8: 297-303.

Taylor NE. and Greene EA. (2003) PARSESNP: a tool for the analysis of nucleotide polymorphisms. Nucleic acids research, 31(13): 3808–11.

Terzaghi WB. and Cashmore AR. (1995) Light-regulated transcription. Annual review of plant biology, 46(1): 445-474.

Till BJ., Zerr T., Comai L., Henikoff S. (2006) A protocol for TILLING and Ecotilling in plants and animals. Nature Protocols, 1(5): 2465-2477.

Trick M., Adamski NM., Mugford SG., Jiang CC., Febrer M., Uauy C. (2012) Combining SNP discovery from next generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC plant biology, 12(1): 14.

Trnka M., Rotter RP., Ruiz-Ramos M., Kersebaum CK., Olesen EJ., Zalud Z., Semenov AM. (2014) Adverse weather conditions for European wheat production will become more frequent with climate change. Nature Climate Change, 4:637-643.

Tsai H., Howell T., Nitcher R., Missirian V., Watson B., Ngo KJ., Lieberman M., Fass, J., Uauy C., Tran RK., Khan, AA., Filkov V., Tai TH., Dubcovsky J., Comai L. (2011) Discovery of rare mutations in populations: Tilling by sequencing. Plant Physiology, 156(3): 1257-1268.

Tucker JE. and Lam Huynh B. (2014) Genotyping by High-Resolution Melting Analysis. Chapter5. In: Crop Breeding: Methods and Protocols. Methods in Molecular Biology, 1145: 59-66. Delphine Fleury and Ryan Whitford (eds). Springer New York.

Tyedmers J., Mogk A., Bukau B. (2010) Cellular strategies for controlling protein aggregation. Nature Reviews Molecular Cell Biology 11: 777-788.

Uauy C., Paraiso F., Colasuonno P., Tran R.K., Tsai H., Berardi S., Comai L., Dubcovsky J. (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biology, 9(1): 115.

Udvardi MK., Kakar K., Wandrey M., Montanri O, Murray J., Andraiankaja A., Zhang JY., Benedito V., Hofer JMI., Cheng F., Town CD. (2007) Legume transcription factors: global regulators of plant development and response to the environment. Plant Physiology 144: 538-549.

Ul Haq N., Ammar M., Bano A., Luthe DS., Heckathorn SA., Shakeel SN. (2013) Molecular Characterization of *Chenopodium album* Chloroplast Small Heat Shock Protein and Its Expression in Response to Different Abiotic Stresses. Plant molecular biology reporter, 31(6): 1230-1241.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth BC., Remm M., Rozen SG. (2012) Primer3-new capabilities and interfaces. Nucleic Acids Research, 40(15): e115.

Urao T., Yamaguchi-Shinozaki K., Urao S., Shinozaki K. (1993) An *Arabidopsis* myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. The Plant Cell, 5(11): 1529-1539.

Van Montfort RL., Basha E., Friedrich KL., Slingsby C., Vierling E. (2001) Crystal structure and assembly of a eukaryotic small heat shock protein. Nature Structural Biology 8: 1025–1030.

Van Montfort RL., Slingsby C., Vierling E. (2002) Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. Advances in Protein Chemistry 59: 105-156.

Varshney RK., Graner A., Sorrells ME. (2005) Genomics-assisted breeding for crop improvement. Trends in plant science, 10(12): 621-630.

Vierling E. and Kimpel JA (1992) Plant responses to environmental stress. Current Opinion in Biotechnology, 3(2): 164-170.

Visioli G., Maestri E., Marmiroli N. (1997) Differential display-mediated isolation of a genomic sequence for a putative mitochondrial LMW HSP specifically expressed in condition of induced thermotolerance in *Arabidopsis thaliana(L.) Heynh*. Plant Molecular Biology, 34(3): 517–527.

Volkov RA, Panchuk II, Mullineaux PM, Schöffl F. (2006) Heat stress-induced $H_2O_2$ is required for effective expression of heat shock genes in *Arabidopsis*. Plant molecular biology, 61(4-5): 733-746.

Von Gromoff ED., Schroda M., Oster U., Beck CF. (2006). Identification of a plastid response element that acts as an enhancer within the Chlamydomonas HSP70A promoter. Nucleic acids research, 34(17): 4767-4779.

Vrána J., Šimková H., Kubaláková M., Číhalíková J., Doležel J. (2012). Flow cytometric chromosome sorting in plants: The next generation. Methods, 57(3): 331-337.

Wahid A., Gelani S., Ashraf M., Foolad MR (2007) Heat tolerance in plants: An overview. Environmental and Experimental Botany, 61(3): 199-223.

Wang WX., Pelah D., Alergand T., Shoseyov O., Altman A. (2002) Characterization of SP1, a stress-responsive, boiling-soluble, homo-oligomeric protein from aspen. Plant physiology 130(2): 865-875.

Wang D. and Luthe DS. (2003) Heat sensitivity in a bentgrass variant. Failure to accumulate a chloroplast heat shock protein isoform implicated in heat tolerance. Plant physiology, 133(1): 319-327.

Wang W., Vinocur B., Altman A. (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta, 218: 1–14.

Wang W., Vinocur B., Shoseyov O., Altman A. (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in plant science 9(5): 245-252.

Wang J., Sun J., Liu D., Yang W., Wang D., Tong Y., Zhang A. (2008) Analysis of *Pina* and *Pinb* alleles in the micro-core collections of Chinese wheat germplasm by Ecotilling and identification of a novel *Pinb* allele. Journal of Cereal Science, 48(3): 836–842.

Wang T., Uauy C., Till B., Liu CM. (2010) TILLING and associated technologies. Journal of integrative plant biology, 52(11): 1027-1030.

Wang TL., Uauy C., Robson F., Till B. (2012) TILLING in extremis. Plant Biotechnology Journal, 10(7): 761-772.

Waters ER., Lee GJ., Vierling E. (1996) Evolution, structure and function of the small heat shock proteins in plants. Journal of Experimental Botany, 47(3): 325-338.

Waters ER, Aevermann BD, Sanders-Reed Z. (2008) Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. Cell Stress & Chaperones, 13: 127–142.

Waters ER. (2013) The evolution, function, structure, and expression of the plant sHSPs. Journal of Experimental Botany, 64(2): 391–403.

Waters ER. and Vierling E. (1999) The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. Molecular Biology and Evolution 16: 127–139.

Weng J., Wang ZF., Nguyen HT. (1991) A *Triticum aestivum* cDNA clone encoding a low-molecular weight heat shock protein. Plant Molecular Biology, 17: 273–275.

Wicker T., Matthews DE., Keller B. (2002) TREP: A database for Triticeae repetitive elements. Trends in Plant Science, 7(12): 561-562.

Wise RR., Olson AJ., Schrader SM., Sharkey TD. (2004) Electron transport is the functional limitation of photosynthesis in field-grown Pima cotton plants at high temperature. Plant, Cell & Environment, 27(6): 717-724.

Xia Y., Li R., Ning Z., Bai G., Siddique HMK., Yan G., Baum M., Varshney KR., Guo P. (2013) Single Nucleotide Polymorphisms in HSP17.8 and Their Association with Agronomic Traits in Barley. PLoS ONE 8(2): e56816.

Yamamoto A., Mizukami Y., Sakurai H. (2005) Identification of a novel class of target genes and a novel type of binding sequence of heat shock transcription factor in *Saccharomyces cerevisiae*. Journal of Biological Chemistry, 280: 11911–11919.

Zadoks JC., Chang TT., Konzak CF. (1974) A decimal code for the growth stages of cereals. Weed Research, 14: 415-421.

Zhang L., Zhang Q., Gao Y., Pan H., Shi S., Wang Y. (2014) Overexpression of heat shock protein gene PfHSP21.4 in *Arabidopsis thaliana* enhances heat tolerance, 36(6): 1555-1564.

# Acknowledgments

The author would to thank the Life Sciences Department of the University of Parma for providing financial support and all the equipment and facilities used in this project. Moreover I gratefully acknowledge Prof. Nelson Marmiroli as my supervisor for giving me the opportunity to conduct my PhD on this interesting project and for supporting me with advices and useful suggestions. I would like to thank my fantastic tutor Dr. Michela Janni of IBBR CNR of Bari (IT) for her assistance and guidance, for giving me a great example of research leadership imparting at every moment tenaciousness and determination. I would to thank Dr. Domenico Pignone of IBBR CNR of Bari (IT) and Dr. Giovanni Giuseppe Vendramin director of IBBR. I'd like to express my gratitude to Dr. Cristobal Uauy of the Crop Genetics Department of John Innes Centre (JIC), Norwich (UK) for his hospitality at JIC, for his supervision and for providing the mutant lines. Thank to Dr. James Simmonds for the precious supervision and support during my working period at JIC, for teaching me the wheat crossing and the glasshouse work.

I would to thank CIB (Consorzio Interuniversitario Biotecnologie) and Erasmus+ CON.C.E.R.T.O (CONsortium for Certified Emilia Romagna Traineeship Opportunities) for the mobility grant and contribution that sustained the stage abroad at JIC (Norwich) and the congress participation.

I would to thank Prof. Mario Pagnotta and Prof. Renato D'Ovidio of the University of Tuscia, Viterbo (IT) and Prof. Antonio Blanco of the University of Bari, Bari, (IT) for providing seeds and nulli tetrasomic and ditelosomic lines.

To all the Professor, Researcher, Post-doc, technician and PhD Student of my group at the Life Science Department of the University of Parma Dr. Elena Maestri, Dr. Giovanna Visioli, Dr. Mariolina Gullì, Dr. Marina Caldara, Dr. Roberta Ruotolo, Dr. Caterina Agrimonti, Urbana Bonas, Dr. Alessio Malcevschi, Dr. Annamaria Sanangelantoni, Dr. Marta Marmiroli, Dr. Davide Imperiale, Dr. Sara Graziano, Dr. Luca Pagano, Dr. Francesca Mussi, Dr. Francesco Pasquali, Dr. Graziella Pira, Dr. Stefania Cadonici, Dr. Valentina Gallo, Dr. Giacomo Lencioni and Dr. Nicola Cavirani: thanks for the critical and scientific support exhibited in every occasions of meeting. I would to special thank my colleagues and friends Dr. Laura Paesano, Dr. Valentina Buffagni and Dr. Veronica Pigoni for their help and support and for making unique and simply likable our everyday work during this last three year.

Thanks also to all the brilliant people of the Uauy team at JIC, in particular: Dr. Peter Scott, Dr. Nikolay Adamski, Dr. Philippa Borrill, Dr. Oluwaseyi Shorinola, Dr. Amelia Frizell-Armitage, Dr. Clare Lewis, Dr. Jemima Brinton for the help and support during my lab work at JIC. Thanks to Dr. Marianna Pasquariello and Dr. Laurie David for their help and friendship during my period at JIC.

In the end, I would to thank Simone for supporting and motivating me throughout all my studies and expecially in the last three years. Thanks also to my parents because they have been at every moment an important support.

# List of publications

The scientific production carried out during my PhD include: three national conference proceedings related to the doctorate research project presented in this PhD thesis; one peer-reviewed article and two national conference proceeding related to the activity conducted for my Master Degree internship (not presented in this thesis).

The publications are listed and attached following:

- COMASTRI A., JANNI M., SIMMONDS J., UAUY C., PIGNONE D., MARMIROLI N. Allele Mining for small HSP genes in durum wheat by applying different TILLING strategies to increase wheat adaptation to heat stress. Poster Communication Abstract – 2.41 Proceedings of the Joint Congress SIBV-SIGA. *Milano, Italy – 8/11 September, 2015.* ISBN 978-88-904570-5-0.

- COMASTRI A., JANNI M., PIGNONE D., MARMIROLI N. Characterization of small heat shock protein genes (sHSP) in durum wheat. Poster Communication Abstract 3.18, Proceedings of the 58$^{th}$ Italian Society of Agricultural Genetics Annual Congress. *Alghero, Italy – 15/18 September, 2014.* ISBN 978-88-904570-4-3.

- JANNI M., COMASTRI A., PIGNONE D., MARMIROLI N. Structural characterization of HSP101 genes and of their promoter in durum wheat. Poster Communication Abstract 3.19, Proceedings of the 58$^{th}$ Italian Society of Agricultural Genetics Annual Congress. *Alghero, Italy – 15/18 September, 2014.* ISBN 978-88-904570-4-3.

- VISIOLI G., COMASTRI A., IMPERIALE D., PAREDI G., FACCINI A., MARMIROLI N. (2016) Gel-based and gel-free analytical methods for the detection of HMW-GS and LMW-GS in wheat flour. *Food Analytical Methods*, 9:469-476.

- VISIOLI G., GULLÌ M., COMASTRI A., IMPERIALE D., CAMPIONI D., PAREDI G., MARMIROLI N. Proteomics of reserve protein in cereals. Italian Proteomics Association (ItPA) 8$^{th}$ National Congress, Abstract Volume pp .112. *Padova, Italy – 18/21 September, 2013.*

- COMASTRI A., GULLÌ M., VISIOLI G., CAMPIOLI D., IMPERIALE D., MARMIROLI N. Analisi proteomiche nei semi di cereali a paglia. Libro degli Atti del IX Convegno 9$^{th}$ AISTEC (Associazione Italiana di Scienza e tecnologia dei Cereali) Congress, Book of abstract pp.113-117. *Bergamo, Italy – 12/14 June, 2013.* ISBN:978-88-906680-1-2.

Proceedings of the Joint Congress SIBV-SIGA
*Milano, Italy – 8/11 September, 2015*
ISBN **978-88-904570-5-0**

**Poster Communication Abstract – 2.41**

# ALLELE MINING FOR SMALL *HSP* GENES IN DURUM WHEAT BY APPLYING DIFFERENT TILLING STRATEGIES TO INCREASE WHEAT ADAPTATION TO HEAT STRESS

COMASTRI A.*, JANNI M.**, SIMMONDS J.***, UAUY C.***, PIGNONE D.**, MARMIROLI N.*,

*) Department of Life Sciences, University of Parma, Parco Area Delle Scienze 11/A, 43124 Parma (Italy)
**) Institute of Biosciences and Bioresources – CNR, Via G. Amendola 165/A, 70126 Bari (Italy)
***) John Innes Center, Norwich Research Park, Norwich NR4 7UH (UK)

*Heat stress, thermotolerance, durum wheat, small Heat Shock Protein, TILLING*

The oncoming climate changes are projected to have negative impact on wheat and other crops production (IPCC report 2014), moreover the projection of an increase of 60% of the total food world consumption by 2050 (FAO, 2012) claim for the need of finding new adaptation traits to both abiotic and biotic stresses to be introduced into modern élite varieties.

Heat stress is one of the main abiotic stress affecting crop yield and many research efforts are now focused on breeding strategies to improve heat tolerance. Understanding the mechanisms and the genes involved in the thermo tolerance acquisition in wheat is therefore a crucial issue.

The HSPs (Heat Shock Proteins) are central effectors and regulators of the plants stress response and within them the low molecular weight HSP (smallHSPs) played an important role allowing the organisms to cope with stress. The photosynthesis is one of the process most affected by heat stress in plants. The chloroplast localized smallHSPs, named *TdHsp26*, are massively expressed at high temperature and function in the protection of PSII during heat stress.

In this work the complete gene sequence of the two homeologous genes *TdHsp26-A1* and *TdHsp26-B1* have been isolated in durum wheat and a conventional reverse genetic approach of TILLING has been apply to identify mutation in this genes by using the HRM as detection techniques. Moreover taking advantage of the new developed technique of exome capture, a new set of mutants were identify *in silico*.

The development of new genetic KASPar markers, the crossing pipeline and preliminary physiological tests on the mutants will be presented.

Proceedings of the 58th Italian Society of Agricultural Genetics Annual Congress
*Alghero, Italy – 15/18 September, 2014*
ISBN **978-88-904570-4-3**

**Poster Communication Abstract – 3.18**

# CHARACTERIZATION OF SMALL HEAT SHOCK PROTEIN GENES (sHSP) IN DURUM WHEAT

COMASTRI A.*, JANNI M.**, PIGNONE D.**, MARMIROLI N.*

*) Department of Life Sciences, University of Parma, Parco Area Delle Scienze 11/A, 43124 Parma (Italy)
**) Institute of Biosciences and Bioresources – CNR, Via G. Amendola 165/A 70126 Bari (Italy)

*Small Heat Shock Protein, Triticum durum, stress response, thermotolerance*

The Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) states that by 2100 in the Mediterranean area we will assist to an increase of the average temperatures of about 2 to 5°C, and to a decrease of the annual precipitation (about 4-27%). Therefore, the current challenge of basic and applied plant research is to shed light on the molecular mechanisms guiding thermotolerance and heat stress response.

Small Heat Shock Proteins (sHSP) belongs to a multigenic family exceptionally widespread and variable in plants. They have a crucial role in thermotolerance therefore the characterization of the corresponding genes in *Triticum durum* is a premise for understanding the specific effects of individual sHSP in stress response in durum wheat.

In order to investigate the implication of sHSPs in thermotolerance we focused our interest in two nuclear-encoded organellar classes of sHSP, in particular those localized in the chloroplast and in the mitochondria since they are known to interact with the Photosystem II upon heat stress and to increase their transcript level in thermotolerant genotype respectively.

In the present work the isolation and the genomic characterization of two plastidic and one mitochondrial sHSP, will be discuss.

Proceedings of the 58th Italian Society of Agricultural Genetics Annual Congress
*Alghero, Italy – 15/18 September, 2014*
ISBN **978-88-904570-4-3**

**Poster Communication Abstract – 3.19**

# STRUCTURAL CHARACTERIZATION OF HSP101 GENES AND OF THEIR PROMOTER REGION IN DURUM WHEAT

JANNI M.**, COMASTRI A.*, PIGNONE D.**, MARMIROLI N.*

*) Department of Life Sciences, University of Parma, Parco Area Delle Scienze 11/A, 43124 Parma (Italy)
**) Institute of Biosciences and Bioresources – CNR, Via G. Amendola 165/A, 70126 Bari (Italy)

*TdHSP101, Durum wheat, thermotolerance, HSP101 promoter*

In agriculture, heat stress is one of the most important constraints on crop yield.

Many efforts have been made to understand the molecular and physiological mechanisms involved in the thermotolerance process in order to find new traits to use in breeding programs.

Heat shock proteins, mainly chaperones or proteases, play the essential role of preventing or minimizing the deleterious effects of heat at the cellular and molecular levels.

Clear evidence of the protective role of HSP101 and its involvement in thermotolerance was previously demonstrated in yeast and Arabidopsis.

In many plant species many cDNA sequences have been isolated suggesting that HSP101 is a member of a small gene family strongly induced by heat.

So far in durum wheat two isoforms of HSP101(*TdHSP101-B* and *C*) with two forms for each genome, have been isolated and characterized at cDNA level (Gulli et al 2007).

The oncoming rising of the temperature worldwide claimed for new traits to introduce in breeding programs therefore the complete coding sequence of candidate genes for heat resilience is necessary for functional genetics studies.

In the present work the isolation and the structural characterization of the complete coding sequence and of the promoter region of *TdHSP101-B* and *C* isoforms will be presented.

Food Anal. Methods (2016) 9:469–476
DOI 10.1007/s12161-015-0218-3

# Gel-Based and Gel-Free Analytical Methods for the Detection of HMW-GS and LMW-GS in Wheat Flour

**Giovanna Visioli[1] · Alessia Comastri[1] · Davide Imperiale[1] · Gianluca Paredi[2] · Andrea Faccini[3] · Nelson Marmiroli[1]**

Received: 17 April 2015 / Accepted: 1 June 2015 / Published online: 9 June 2015


**Abstract** Durum wheat (*Triticum turgidum* L.) flour is instrumental for the production of pasta worldwide. The quality of this food rests on flour processing and on its protein content and composition. Gluten proteins as high and low-molecular weight glutenins (GS) are important to predict the flour technological property in pasta making. Different methods were compared to separate, identify and quantify GS in flours from two wheat cultivars. Sodium dodecyl sulphate-polyacrilamide gel electrophoresis (SDS-PAGE) gave in a fast way information about the GS assets. Two-dimensional gel electrophoresis (2D-GE) allowed for the highest resolution in detecting and quantifying single GS, subsequently identified by liquid chromatography-electrospray ionization mass spectrometry (LC-ESI-MS/MS). Reversed-phase high-performance liquid chromatography (RP-HPLC) is a non-gel alternative system for separation and quantification of single GS that when combined with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF/MS) gave information about their exact masses. This method gives also quantitative indications of each individual GS. Different GS patterns and contents were detected in the flour of the two cultivars, underlining the importance of these analytical methods before determining the best flour processing procedure in pasta making. The different methods were evaluated with a modular approach consisting of a grid of different parameters and a non-linear score within each module.

**Keywords** Durum wheat flour · High and low-molecular weight glutenins · Quality · Proteomics methods

**Abbreviations**

| | |
|---|---|
| GS | Glutenins |
| SDS-PAGE | Sodium dodecyl sulphate-polyacrilamide gel electrophoresis |
| 2D-GE | Two-dimensional gel electrophoresis |
| LC-ESI-MS/MS | Liquid chromatography-electrospray ionization mass spectrometry |
| RP-HPLC | Reversed-phase high-performance liquid chromatography |
| MALDI-TOF/MS | Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry |

## Introduction

Durum wheat (*Triticum turgidum* L., var. *durum*) is an important staple food mainly used for pasta and bread making in Europe and North America. The economical and technological value of its flour rests on the combined characteristics of seed proteins and starch (Sissons 2008). Gluten proteins consist of monomeric gliadins soluble in acqueous alcohols and of polymeric glutenins insoluble in alcohol: solubility being

**Electronic supplementary material** The online version of this article (doi:10.1007/s12161-015-0218-3) contains supplementary material, which is available to authorized users.

✉ Giovanna Visioli
giovanna.visioli@unipr.it

[1] Department of Life Sciences, University of Parma, Parco Area delle Scienze 11/a, 43124 Parma, Italy

[2] Department of Pharmacy, Center Siteia.Parma, University of Parma, Parco Area delle Scienze 23/a, 43124 Parma, Italy

[3] Interdepartmental Measure Centre "Giuseppe Casnati", University of Parma, Parco Area delle Scienze 24/a, 43124 Parma, Italy

## P-III-12

# Proteomics of reserve proteins in cereals

**Giovanna Visioli[a], Mariolina Gullì [a], Alessia Comastri [a], Davide Imperiale [a], Daniel Campioli [a], Gianluca Paredi [b], Nelson Marmiroli [a]**

a) Department of Life Sciences, University of Parma, Parco Area delle Scienze 11/A, 43124 Parma, Italy
b) Department of Pharmacy, Interdepartmental Center Siteia, University of Parma, Parco Area delle Scienze 24/a, 43124 Parma, Italy

Durum wheat (*Triticum durum* Desf.) and rice are important cereal crops in Italy, with different cultivars showing significant behaviours in adaptation to environmental and agronomical conditions. We have developed methods for characterising the main seed storage proteins, in order to identify diagnostic markers for cultivar characterisation.
Proteomic approaches were applied to the analysis of seed protein content and diversity in four different varieties of Italian durum wheat (Biensur, Aureo, Ariosto, Liberdur) and in four different varieties of Italian rice (Carnaroli, Karnak, Volano, Arborio) using 1D and 2D-gel electrophoresis (GE), RP-HPLC, 2D liquid chromatography (2D-LC), and mass spectrometry (MALDI-TOF/MS) analyses.
The results obtained in durum wheat by 2D-GE and RP-HPLC followed by MS analyses showed that the composition of seed storage proteins can be correlated with the mineral nutrition status of the plants, in particular concerning N fertilisation. In addition protein polymorphisms have also been evidenced among cultivars, useful for the development of diagnostic markers.
In rice the proteomic analyses focused on the seed storage proteins which were analysed by 2D-GE. The major spots were excised and treated with trypsin for peptide mass fingerprinting analysis with a MALDI-TOF/MS. The comparison of the different cultivars allowed for the identification of quality markers. Proteomic analyses were also performed, on the same samples, with 2D-LC for the separation of water-soluble proteins, with the advantage of being a gel-less system. MS has allowed the identification of several proteins which were differentially abundant in the four cultivars (Arborio, Carnaroli, Karnak and Volano).
The methods developed will be further exploited for the characterisation of cultivar diversity and adaptation to environmental conditions.

*Corresponding author*: Giovanna Visioli, Dr.
University of Parma, Dept. of Life Sciences, Parco Area delle Scienze 33/A 43124 Parma, Italy;
Tel.: +39 0521 905692
Fax: +39 0521 906222
e-mail: giovanna.visioli@unipr.it

CRA
CONSIGLIO PER LA RICERCA
E LA SPERIMENTAZIONE
IN AGRICOLTURA

AISTEC

# Atti del 9° Convegno AISTEC

## Un mondo di cereali

## Potenzialità e sfide

Bergamo 12 - 14 giugno 2013

a cura di

**R. Acquistucci, M. G. D' Egidio, G. Panfili, R. Redaelli**

**Analisi proteomiche in semi di cereali a paglia**

*A. Comastri*, M. Gullì, G. Visioli, D. Campioli, D. Imperiale, N. Marmiroli*

Dipartimento di Bioscienze, Università degli Studi di Parma, Parco Area delle Scienze 11/A - 43124 Parma.
*E-mail: alessia.comastri86@gmail.com

**Abstract**
Durum wheat (*Triticum durum* Desf.) and rice (*Oryza sativa)* are important cereal crops in Italy. Cultivars have been selected and characterised for the technological properties of grain and flour, and for their adaptation capacity to environmental and agronomical conditions. The molecular characterization of seed storage proteins can help to identify diagnostic biomarkers for these characters. A proteomic approach was used to analyse protein content and diversity in four Italian durum wheat varieties and in four Italian rice varieties which are considered to be agronomically interesting within the AGER projects ("Environmental and economic sustainability for yield and quality production of durum wheat supply chain" grant n° 2010-0278 and "Crop Improvement for sustainable rice production in Italy- "RISINNOVA" grant n° 2010-2369). After implementation of the protein extraction methodologies, rice and wheat storage proteins were analyzed by gel electrophoresis and liquid chromatography, taking advantages of proper software to obtain qualitative and quantitative data. The protein components were also characterized and identified by mass spectrometry analyses.

**Riassunto**
Il frumento duro (*Triticum durum* Desf.) e il riso (*Oryza sativa* L.) sono importanti colture cerealicole; numerose varietà italiane sono state selezionate e studiate per le caratteristiche tecnologiche della granella, delle farine derivate e per l'adattamento della pianta a diverse condizioni ambientali. La caratterizzazione delle varietà dal punto di vista molecolare mediante analisi proteomiche può essere utile a determinare biomarcatori di qualità tecnologica per individuare caratteristiche interessanti di risposta a trattamenti agronomici. Nell'ambito dei due progetti di ricerca finanziati da AGER ("Sostenibilità produttivo-ambientale, qualitativa ed economica della filiera frumento duro" grant n° 2010-0278 e "Sistemi integrati genetici e genomici mirati al rinnovo varietale nella filiera risicola Italiana-"RISINNOVA" grant n° 2010-2369) diverse varietà italiane di frumento duro e di riso sono state analizzate con metodologie di indagine proteomica. In seguito all'implementazione di adeguate tecniche di estrazione, le proteine di riserva di frumento e di riso sono state separate e confrontate dal punto di vista qualitativo e quantitativo mediante elettroforesi, cromatografia liquida e appositi software di analisi. Le componenti proteiche analizzate sono state inoltre caratterizzate e identificate mediante spettrometria di massa.

**Introduzione**
L'importanza delle colture cerealicole di frumento e riso risiede principalmente nelle caratteristiche tecnologiche e nutrizionali della granella. Per il frumento è di particolare interesse la composizione della frazione di proteine di riserva del seme, infatti sia la presenza di determinate subunità proteiche che l'abbondanza relativa delle diverse classi possono condizionare le caratteristiche degli impasti e la loro destinazione d'uso (Flagella, 2006). La frazione proteica più interessante è data dalle proteine del glutine appartenenti alla classe delle prolammine alcol-solubili che si dividono in gliadine e glutenine, queste ultime si

suddividono a loro volta in componenti ad alto e a basso peso molecolare. Per quanto riguarda il riso invece la frazione proteica più abbondante nel seme è costituita dalle gluteline, solubili in soluzioni alcaline. Queste sono le componenti più interessanti per la caratterizzazione qualitativa di varietà di riso, infatti le proteine Waxy e la subunità A delle gluteline possono essere utili marcatori non solo per distinguere tra varietà di japonica e indica, ma anche per una discriminazione intraspecifica. Inoltre le gluteline di tipo B potrebbero essere coinvolte nelle risposte di protezione in condizioni di carenza idrica (Farinha *et al.*, 2011).
Il contenuto proteico della granella matura dipende dalla varietà considerata e generalmente è inversamente proporzionale alla resa, pertanto la caratterizzazione varietale per la frazione di proteine di riserva è importante ai fini di individuare e selezionare genotipi produttivi e di buona qualità tecnologica (Flagella, 2006). In questo lavoro sono state utilizzate metodologie analitiche innovative per la caratterizzazione del proteoma del seme in varietà italiane di interesse agronomico, non ancora caratterizzate dal punto di vista molecolare, allo scopo di identificare marcatori di qualità. In particolare sono state utilizzate tecniche di analisi qualitative, come separazione in elettroforesi denaturante (SDS-PAGE) e spettrometria di massa (MS), e tecniche semi-quantitative come elettroforesi bidimensionale (2D-PAGE) e cromatografia liquida a fase inversa (RP-HPLC).

## Materiali e metodi

### *Materiale vegetale*

I semi delle seguenti varietà di frumento duro: Biensur (Apsovsementi), Aureo (Produttori Sementi Bologna), Ariosto (Apsovsementi), Liberdur (Società Italiana Sementi) e delle seguenti varietà di riso: Carnaroli, Karnak, Volano, Arborio (CRA-RIS, Vercelli) sono stati macinati per ottenere lo sfarinato da cui è stata effettuata l'estrazione proteica.

### *Estrazione delle proteine e separazione mediante SDS-PAGE*

Per le proteine di riserva da frumento duro è stato applicato un protocollo di estrazione sequenziale (Singh *et al.*, 1991) che ha consentito di separare le gliadine e le subunità gluteniniche ad alto e basso peso molecolare (rispettivamente HMW-GS e LMW-GS), che nel complesso costituiscono le proteine del glutine. L'estrazione delle proteine di riserva del seme di riso è stata condotta utilizzando un tampone contenente SDS-Urea (Khan *et al.*, 2008). Gli estratti sono stati separati in SDS-PAGE con l'apparato elettroforetico Mini-PROTEAN® Tetracell (Biorad) su gel 12% per le LMW-GS e 8% per le HMW-GS di frumento e su gel 15% per le proteine di riserva di riso.

### *2D-PAGE e software di analisi*

Le frazioni LMW-GS di frumento e le proteine di riserva di riso delle diverse varietà sono state separate in 2D-PAGE mediante apparato i12Protean (Biorad). L'IEF è stata condotta su IPG-strip con intervallo di punto isoelettrico pI 6-11 per le LMW-GS e su IPG-strip pI 3-10 per le proteine di riserva di riso. Per le varietà di frumento duro Aureo e Biensur, ritenute particolarmente interessanti dal punto di vista agronomico, l'analisi è stata condotta in due repliche biologiche e tre repliche tecniche al fine di poter condurre un'analisi statistica sui dati qualitativi e semi-quantitativi ottenuti mediante il software PD Quest (Biorad). Gli spot di proteine di riserva di riso sono stati excisi da gel, digeriti con tripsina e purificati con ZipTip C18 (Millipore) per l'identificazione in MS.

### *Cromatografia liquida RP-HPLC*

Le frazioni HMW-GS e LMW-GS di Aureo e Biensur sono state separate ed eluite in RP-HPLC secondo protocollo Pirondini *et al.* (2006) sfruttando l'apparato di seconda dimensione del sistema ProteomeLab™ PF2D (Beckman Coulter) che consente una quantificazione delle proteine.

***Spettrometria di massa MALDI-TOF/MS e identificazione in silico***

Le frazioni eluite da RP-HPLC sono state analizzate in MALDI-TOF/MS (Abi Sciex) in modalità lineare, per rilevare differenze in termini di massa delle subunità proteiche nelle diverse varietà. I digeriti triptici degli spot di proteine di riserva di riso sono stati analizzati in MALDI-TOF/MS con approccio Peptide Mass Fingerprinting (PMF) e l'identificazione *in silico* è stata ottenuta utilizzando il programma online Mascot (Matrix Science) con i seguenti parametri di ricerca: banca dati: Swiss-Prot; enzima: tripsina; tolleranza consentita nei siti di taglio: uno; tassonomia: Oryza sativa (rice); valore mono-isotopico ± 1,2Da; modificazioni fisse consentite: Cys carbamidomethyl.

## Risultati e discussione

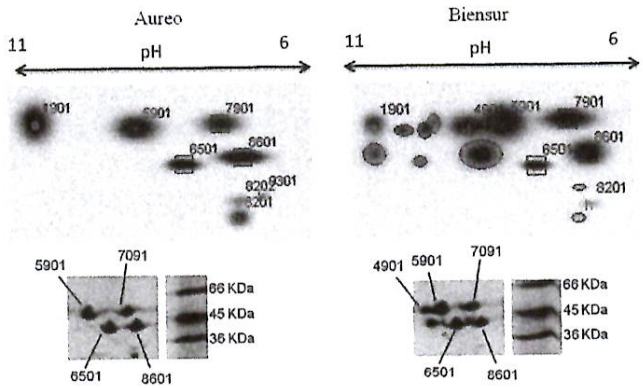
In una prima fase del lavoro le varietà di frumento duro Biensur, Ariosto, Aureo e Liberdur sono state caratterizzate mediante SDS-PAGE per la composizione delle GS, sfruttando come riferimento le varietà di frumento duro Dylan e Svevo, già caratterizzate per la composizione allelica di HMW-GS al locus GluB1 (per i geni associati Bx e By) e per la qualità tecnologica. Per Aureo e Biensur le GS del peso molecolare esatto atteso (Gianibelli *et al.*, 2001; D'Ovidio e Masci, 2004) sono state rilevate mediante MALDI-TOF/MS dalle frazioni eluite corrispondenti separate in RP-HPLC (Tab. 1).

**Tabella 2.** Caratterizzazione varietale delle varietà di frumento duro per la composizione delle GS.

| Frazione glutenine | Alleli | Svevo | Dylan | Ariosto | Liberdur | Aureo ([a]) | Biensur([a]) |
|---|---|---|---|---|---|---|---|
| HMW-GS | Bx | 7 | 6 | 7 | 20 | 6<br>(86.510 da; Abs 0,630±0,08) | 7<br>(82.927 da; Abs 0,542±0,09) |
| | By | 8 | 8 | 8 | silente | 8<br>(75.086 da; Abs 0,241±0,02 ) | 8<br>(74.862 da; Abs 0,131±0,02**) |
| LMW-GS | LMW-2 | LMW-2 | LMW-2 | LMW-2 | LMW-2 | LMW-2<br>(41.965 da; Abs 0,949±0,11)<br>(37.587da; Abs 0,587±0,2)<br>(38.406 da; Abs 0,490±0,03) | LMW-2<br>(41.997 da; Abs 0,969 ±0,03)<br>(37.671 da; Abs 0,656 ±0,02)<br>(38.500 da; Abs 0,584 ±0,008)<br>(39.064 da; Abs 0,355 ±0,01) |

([a]) massa esatta (da) e valore medio su tre repliche delle assorbanze (Abs) dei picchi cromatografici RP-HPLC ± deviazione standard. Significatività t-student **p <0,001.

Le subunità HMW-GS 7 + 8 indici di buona qualità pastificatoria, che caratterizzano Svevo, si ritrovano in Biensur e Ariosto; mentre Aureo presenta le subunità 6 + 8. Tutte le varietà hanno LMW-GS dell'assetto LMW-2 indicatore di buona qualità. Le varietà Ariosto e Biensur riportano, oltre ad una differente composizione e ammontare di HMW-GS, interessanti isoforme differenziali LMW-2 che sono state separate in 2D-PAGE per una caratterizzazione qualitativa e quantitativa mediante PD Quest (Fig. 1).

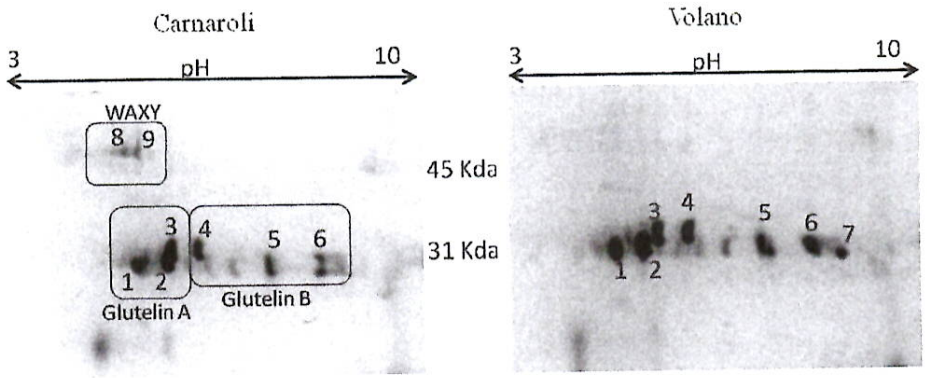



| | Aureo | Biensur |
|---|---|---|
| 4901 | - | ++ |
| 5901 | + | ++ |
| 6501 | ++ | ++ |
| 7901 | + | +++ (*) |
| 8601 | + | ++ |

**Figura 1.** Separazione 2D-PAGE delle frazioni LMW-GS di Aureo e Biensur. A sinistra sono riportati i Master gel ottenuti dall'analisi PD Quest di tre replicati per ogni campione e l'esempio di due gel reali. A destra sono riportati i dati di intensità degli spot interessanti e riproducibili espressi con i simboli + (aumento di espressione) e - (mancanza di espressione). Significatività in base a t-student *p <0,05.

Le differenze rilevate tra queste due varietà, già interessanti per resa agronomica e contenuto proteico, potrebbero essere associate alla qualità tecnologica e quindi costituire biomarcatori di selezione.

La frazione di proteine di riserva delle varietà di riso oggetto di studio è stata separata mediante 2D-PAGE per ottenere i profili proteici caratteristici delle singole varietà. Gli spot più abbondanti sono poi stati caratterizzati mediante spettrometria di massa. A titolo di esempio in Figura 2 è riportata l'analisi ottenuta per le varietà Carnaroli e Volano.




| | Massa (da) | N°accessione | Proteina |
|---|---|---|---|
| 1 | 31904 | gi 94730382 | Glutelin type A-3 |
| 2 | 31892 | gi 94730381 | Glutelin type A-1 |
| 3 | 31904 | gi 121475 | Glutelin type A-2 |
| 4 | 32037 | gi 121477/gi75290219 | Glutelin type B-4/B-5 |
| 5 | 31756/31265 | gi 94730383/gi94730384 | Glutelin type B-1 /B-2 |
| 6 | 31756 | gi 94730383 | Glutelin type B-1 |
| 7 | 31756 | gi 94730383 | Glutelin type B-1 |
| 8 | 66476 | gi 12216848 | WAXY 1 |
| 9 | 66476 | gi 12216848 | WAXY 1 |

**Figura 2.** Separazione 2D-PAGE delle proteine di riserva di Carnaroli e Volano. I numeri riportati indicano gli spot excisi da gel e identificati con spettrometria di massa PMF. L'identificazione delle proteine con i relativi valori di massa e numeri di accessione è riportata in tabella.

I risultati mostrati in Figura 2 evidenziano putative differenze nell'abbondanza e nella composizione di proteine Waxy e gluteline di tipo A e B. Le analisi così ottenute dal raccolto del 2011 verranno confrontate con il raccolto ottenuto nel 2012 delle stesse varietà cresciute anche in condizioni di assenza di acqua al fine di identificare biomarcatori correlati alla qualità tecnologica e alla capacità di adattamento a diverse condizioni di crescita (in particolare all'asciutta).
L'analisi proteomica differenziale tra le diverse varietà, intrapresa mediante tecniche separative e di MS accoppiate a software di analisi quantitativa avanzati, si è rivelata uno strumento utile per la caratterizzazione varietale di cereali con lo scopo di individuare biomarcatori associati a qualità tecnologica e nutrizionale e di risposta a diverse condizioni ambientali.

**Bibliografia**

D'Ovidio R., Masci S. 2004. The low-molecular-weight glutenin subunits of wheat gluten. *Journal of Cereal Science*, 39(3): 321-339.
Farinha A.P., Irar S., de Oliveira E., Oliveira M.M., Pagès M. 2011. Novel clues on abiotic stress tolerance emerge from embryo proteome analyses of rice varieties with contrasting stress adaptation. *Proteomics*, 11(12): 2389-2405.
Flagella Z. 2006. Qualità nutrizionale e tecnologica del frumento duro. *Italian Journal of Agronomy*, 1(1): 203-239.
Gianibelli M.C., Larroque O.R., MacRitchie F., Wringley W. 2001. Biochemical, Genetic, and Molecular Characterization of Wheat Glutenin and Its Component Subunits. *Cereal Chemistry*, 78(6): 635-646.
Khan N., Katsube-TanakaT., Lida S., Yamaguchi T., Nakano J., Tsujimoto H. 2008. Diversity of rice glutelin polypeptides in wild species assessed by the higher-temperature sodium dodecyl sulfate-polyacrylamide gel electrophoresis and subunit-specific antibodies. *Electrophoresis*, 29(6): 1308-1316.
Pirondini A., Visioli G., Malcevschi A., Marmiroli N. 2006. A 2-D liquid-phase chromatography for proteomic analysis in plant tissues. *Journal of Chromatography B*, 833(1): 91-100.
Singh N.K., Shepherd K.W., Cornish G.B. 1991. A simplified SDS-PAGE procedure for separating LMW subunits of glutenin. *Journal of cereal science*, 14(3): 203-208.