

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN
SCIENZE E TECNOLOGIE AGRARIE, AMBIENTALI E
ALIMENTARI**

Ciclo XXXII

Settore Concorsuale: 07/B1 – Agronomia e Sistemi Colturali Erbacei ed Ortofloricoli

Settore Scientifico Disciplinare: AGR/02 – Agronomia e Coltivazioni Erbacee

**Assessment of molecular sequences and enzyme activity
of amylase trypsin inhibitors (ATIs)
from a selection of ancient and modern grains**

Presentata da: Dott.ssa Emanuela Simonetti

**Coordinatore Dottorato
Prof. Massimiliano Petracchi
Dinelli**

**Supervisore
Prof. Giovanni**

**Co-supervisore
Dott.ssa Sara Bosi**

Esame finale anno 2020

*A mio figlio Sirio:
non smettere mai di cercare
perché l'infinito si può scoprire
in ogni frammento dell'esistenza.*

“Se il punto in cui ti immergi in un fiume è il presente, pensai,
allora il passato è l’acqua che ti ha superato, quella che va verso il basso
e dove non c’è più niente per te,
mentre il futuro è l’acqua che scende dall’alto, portando pericoli e sorprese.
Il passato è a valle, il futuro a monte.”
Paolo Cognetti, *Le otto montagne*.

“È il tempo che tu hai perduto per la tua rosa
che ha fatto la tua rosa così importante.”
Antoine de Saint-Exupery, *Il Piccolo Principe*.

Abstract

Wheat amylase-trypsin inhibitors (ATIs) are a family of wheat proteins, which play an important role in plant defence against pest attacks. ATIs are also of great interest for their impact on human health and recently ATIs have been identified as major stimulators of innate immune cells.

In this study, ten selected wheat samples with different ploidy level and year of release were used for the agronomic trial, for *in vitro* enzymatic assays and for ATIs gene sequencing. Wheat samples were grown under organic farming management during three consecutive cropping years at two growing areas (Italy and USA).

The PCA analysis performed on the deduced amino acid sequences of four representative ATIs genes (WMAI, WDAI, WTAI-CM3, CMx) evidenced that the ten wheat varieties can be differentiated on the basis of their ploidy level, but not with respect to ancient or recently developed wheat genotypes.

The results from *in vitro* alpha-amylase and trypsin inhibitory activities showed high variability among the ten wheat genotypes and the contribution of the genotype and the cropping year was significant for both inhibitory activities. The hexaploid wheat genotypes showed the highest inhibitory activities. Einkorn showed a very low or even absent alpha-amylase inhibitory activity and the highest trypsin inhibitory activity. It was not possible to differentiate ancient and recently developed wheat genotypes on the basis of their ATIs activity.

The weather conditions differently affected the two inhibitory activities. In both cultivation areas, higher precipitation and lower high mean temperatures correlated with lower alpha-amylase inhibitory activities, while there were different correlations considering trypsin inhibitory activity for the two growing areas. The protein content negatively correlated with both inhibitory activities in USA and Italy.

This information can be important in the understanding of plant defence mechanisms in relation to the effect of both genotype and abiotic and biotic stress.

Index

1. INTRODUCTION	5
1.1 WHEAT AND HUMAN NUTRITION	5
1.2 WHEAT PROTEINS	6
1.3 WHEAT AMYLASE-TRYPSIN INHIBITORS (ATIs)	7
1.3.1 TYPES OF WHEAT AMYLASE-TRYPSIN INHIBITORS (ATIs)	7
1.3.2 BIOLOGICAL ACTIVITY OF ATIs IN WHEAT	8
1.3.3 ROLE OF ATIs IN HUMAN DISEASE	9
1.3.4 ATIs IN ANCIENT AND MODERN WHEAT	13
2. AIM	15
3. MATERIALS AND METHODS	18
3.1 PLANT MATERIAL	18
3.2 DNA ISOLATION AND PCR AMPLIFICATION	21
3.3 BIOINFORMATICS ANALYSIS	25
3.3.1 HAPLOTYPES IDENTIFICATION	25
3.3.2 PRIMARY STRUCTURAL ANALYSIS OF DEDUCED PROTEINS	25
3.3.3 SECONDARY STRUCTURAL PREDICTION OF DEDUCED PROTEINS	26
3.3.4 PHYLOGENETIC ANALYSIS	27
3.3.5 PRINCIPAL COMPONENT ANALYSIS (PCA)	27
3.4 AGRONOMIC TRIAL	28
3.5 ASSESSMENT OF MYCOTOXIN LEVELS	29
3.6 ENZYMATIC ASSAYS	29
3.6.1 ALPHA-AMYLASE INHIBITORY ACTIVITY ASSAY	29
3.6.2 TRYPSIN INHIBITORY ACTIVITY ASSAY	30
3.6.3 STATISTICAL ANALYSIS	31
4. RESULTS AND DISCUSSION	32
4.1 MONOMERIC ALPHA-AMYLASE INHIBITORS (WMAI)	32
4.1.1 SEQUENCE ANALYSIS OF MONOMERIC ALPHA-AMYLASE INHIBITORS	32
4.1.2 PRIMARY STRUCTURAL ANALYSIS OF MONOMERIC ALPHA-AMYLASE INHIBITORS	41
4.1.3 SECONDARY STRUCTURAL PREDICTION OF MONOMERIC ALPHA-AMYLASE INHIBITORS	42
4.1.4 PHYLOGENETIC ANALYSIS OF MONOMERIC ALPHA-AMYLASE INHIBITORS	43
4.1.5 PRINCIPAL COMPONENT ANALYSIS (PCA) OF WMAI SEQUENCES	44
4.2 DIMERIC ALPHA-AMYLASE INHIBITORS (WDAI)	47
4.2.1 SEQUENCE ANALYSIS OF DIMERIC ALPHA-AMYLASE INHIBITORS	47
4.2.2 PRIMARY STRUCTURAL ANALYSIS OF DIMERIC ALPHA-AMYLASE INHIBITORS	54
4.2.3 SECONDARY STRUCTURAL PREDICTION OF DIMERIC ALPHA-AMYLASE INHIBITORS	56

4.2.4 PHYLOGENETIC ANALYSIS OF DIMERIC ALPHA-AMYLASE INHIBITORS	58
4.2.5 PRINCIPAL COMPONENT ANALYSIS (PCA) OF WDAI SEQUENCES	60
4.3 TETRAMERIC ALPHA-AMYLASE INHIBITORS CM3 (WTAI-CM3)	62
4.3.1 SEQUENCE ANALYSIS OF TETRAMERIC ALPHA-AMYLASE INHIBITORS CM3	62
4.3.2 PRIMARY STRUCTURAL ANALYSIS OF TETRAMERIC ALPHA-AMYLASE INHIBITORS CM3	70
4.3.3 SECONDARY STRUCTURAL PREDICTION OF TETRAMERIC ALPHA-AMYLASE INHIBITORS CM3	71
4.3.4 PHYLOGENETIC ANALYSIS OF TETRAMERIC ALPHA-AMYLASE INHIBITORS CM3	73
4.3.5 PRINCIPAL COMPONENT ANALYSIS (PCA) OF WTAI-CM3 SEQUENCES	74
4.4 TRYPSIN INHIBITORS CMX	76
4.4.1 SEQUENCE ANALYSIS OF TRYPSIN INHIBITORS CMX	76
4.4.2 PRIMARY STRUCTURAL ANALYSIS OF TRYPSIN INHIBITORS CMX	81
4.4.3 SECONDARY STRUCTURAL PREDICTION OF TRYPSIN INHIBITORS CMX	82
4.4.4 PHYLOGENETIC ANALYSIS OF TRYPSIN INHIBITORS CMX	84
4.4.5 PRINCIPAL COMPONENT ANALYSIS (PCA) OF CMX SEQUENCES	85
4.5 PRINCIPAL COMPONENT ANALYSIS (PCA) OF WMAI, WDAI, WTAI-CM3 AND CMX SEQUENCES	86
4.6 AGRONOMIC TRIAL	90
4.7 MYCOTOXIN LEVELS	96
4.8 ENZYMATIC ASSAYS	97
4.8.1 ALPHA-AMYLASE INHIBITORY ACTIVITY ASSAY	97
4.8.2 TRYPSIN INHIBITORY ACTIVITY ASSAY	106
5. CONCLUSIONS	113
6. BIBLIOGRAPHY	115

1. Introduction

1.1 Wheat and human nutrition

The three main cereals of the world agriculture are wheat, rice and maize. Among them, wheat (*Triticum spp.*) is the most widely grown: originated from the Crescent Fertile region, at present the wheat cultivation area ranges from Scandinavia and Russia in the north to Argentina in the south and encloses also some regions of the tropics. Moreover, wheat is a very diverse crop which includes traditional landraces and population as well as more than 25000 different cultivars developed by plant breeders (Tatham et al., 2008). The FAO forecast for global wheat production in 2019 is 766 million tonnes (www.fao.org/faostat/) and more than 90% of this production is represented by *Triticum aestivum*, usually called “common” or “bread” wheat.

Wheat production is mainly addressed to human consumption after transformation in many types of food products (e.g. pasta, baked goods, noodles, couscous, bulgar). Wheat is consumed by billions of people and is the staple food of many diets, in particular the Mediterranean diet, which is recognized as one of the world’s healthiest diets, supplying a high portion of the energy daily intake.

Wheat has a high nutritional value and is considered a healthy source of many nutrients, including macronutrients such as starch and dietary fiber, proteins and fats, as well as essential minerals, vitamins and bioactive compounds, especially in whole-grain products. Many studies have associated a diet based on whole-grain products with reduced levels of cardio-metabolic risk factors such as total and LDL-cholesterol, triglycerides, blood glucose, blood pressure and body mass index and with a significant decreased risk for cardiovascular disease and cancer (Poutanen et al., 2008).

However, wheat is responsible for some adverse reactions which includes respiratory allergy (bakers’ asthma) and coeliac disease, a gluten intolerance in genetically predisposed individuals which affects about 1% of the industrialized Countries. Wheat dietary allergy is less spread and accounts for 0,1% of the general population.

1.2 Wheat proteins

The protein content in wheat is on average 10-15% of the dry weight and depends on the wheat genotype and on agronomic conditions (nitrogen availability).

Wheat proteins are classified in four groups according to their extractability in a series of solvents, as formalized by T. B. Osborne working in the late 19th/early 20th century. The four wheat protein fractions (often called 'Osborne fractions') are albumins, globulins, prolamins, glutenins and are extracted sequentially in water, dilute saline, alcohol/water mixtures and dilute acid respectively.

The gliadins are monomeric proteins and are classified into three groups (α/β -gliadins, γ -gliadins, ω -gliadins) on the basis of their electrophoretic mobility. Differently, glutenins consist on individual proteins linked together by disulphide bonds to form polymers and are classified into high molecular weight (HMW) and low molecular weight (LMW) groups based on separation by SDS-PAGE after reduction of disulphide bonds. Gliadins and glutenins account for approximately 80% of the total grain protein and are the major storage protein fractions providing nutrients to the growing seedling when digested during germination. When wheat flour is mixed with water to make dough, gliadins and glutenins form a continuous network called "gluten". Gluten is the main responsible for the cohesiveness and viscoelasticity of the wheat dough and is the main determinant of the high technological properties of wheat. Thanks to the properties of gluten, wheat can be transformed in a wide variety of products such as pasta and bread. The gluten proteins are also responsible for many gluten-related disorders like the celiac disease, wheat allergy and the non-celiac gluten sensitivity.

The nongluten proteins are albumins and globulins and include minor storage proteins, α -amylase/protease inhibitors and several other enzymes. These proteins are much more soluble in water or aqueous salt solutions than gluten. Several nongluten proteins, including α -amylase/protease inhibitor, thiol-reductase, serine protease inhibitor (serpin), and β -amylase have been identified as potent allergens in IgE-mediated wheat allergy and/or baker's asthma (Larre et al., 2011).

Recently, Huebener et al. (2015) showed that celiac people produce antibodies directed not only at gluten proteins, but also at a specific subset of the nongluten proteins of wheat, identified as serpins, purinins, α -amylase/protease inhibitors, globulins, and farinins.

1.3 Wheat amylase-trypsin inhibitors (ATIs)

1.3.1 Types of wheat amylase-trypsin inhibitors (ATIs)

Wheat amylase-trypsin inhibitors (ATIs) are an important family of low molecular weight, water soluble proteins and represent ~2–4% of total wheat protein (Dupont et al., 2011). ATIs are cysteine-rich proteins and have a 3D structure rich in α -helices that gives them a high stability to proteolysis and thermal processing. These proteins can inhibit amylases and/or trypsin or other proteases like chymotrypsin (Salcedo et al., 2004).

According to their aggregation state, three groups of alpha-amylase inhibitors have been identified that are active against insect, mite and mammalian alpha-amylases, but not against cereal enzymes (Carbonero et al., 1999). These three groups include the 12 kDa wheat monomeric amylase inhibitors (WMAI), often referred to as 0.28 proteins; proteins that form the 24 kDa wheat homodimeric amylase inhibitors (WDAI), sometimes referred to as the 0.19 and 0.53 proteins and proteins that make up the 60 kDa wheat heterotetrameric amylase inhibitors (WTAI). The WMAI inhibitors are encoded by genes on the short arm of chromosome 6 of *T. aestivum* (Gomez et al., 1990), the genes for WDAI are located in the short arm of chromosome 3 (Sanchez-Monge et al., 1986), in particular on 3BS and 3DS chromosomes, though not much is known about homologous common wheat chromosome 3AS (Pandey et al., 2016). Some studies using different methods like isoelectric focusing, two-dimensional gel electrophoresis and direct and clone sequencing highlighted the presence of multiple copies of the dimeric α -amylase inhibitor genes in polyploid wheat (Sánchez-Monge et al. 1986, Wang et al. 2005).

The tetrameric inhibitors are often called CM proteins because of their solubility in chloroform/methanol and in modern hexaploid wheat 5 sub-units have been identified: CM1, CM2, CM3, CM16 and CM17 (Carbonero et al., 1999). CM proteins are generally composed of one copy of either CM1 or CM2, encoded by genes on chromosomes 7D or 7B, plus one copy of either CM16 or CM17, encoded by genes on chromosomes 4B and 4D, plus two copies of CM3, also encoded on chromosomes 4B and 4D.

The putative wheat trypsin inhibitors are referred to as CMx proteins, are encoded by genes on the group 4 chromosomes and are monomeric.

Altenbach et al. (2011) identified 17 different ATIs species of ~120- to 150-amino in the modern wheat genome. The main ATIs species in wheat are 0.19 (WDAI) and CM3

(WTAI).

All ATIs are characterized by high homology in their amino acid sequence and by an analogous compact secondary structure showing 5 (or less often 4) intrachain disulphide bridges and a main body of the molecule built around four α -helices arranged in a “up and down” pattern linked together by loop segments (Oda et al., 1997). The disulphide bonds are essential for their functionality: Junker et al. (2012) showed that the reduction of the ATIs disulphide chains completely eliminated the stimulatory activity of the wheat fractions that contained them.

WMAI, WDAI and WTAI proteins mainly have amylase inhibitory activities, but some of them also show some proteolytic inhibitory activities for example against trypsin, chymotrypsin or subtilisin, even if this inhibitory activity is still not clear today. For example, the aggregation state and the combination of its CM subunits determine the proteolytic inhibitory activity of WTAI (Call et al., 2019).

1.3.2 Biological activity of ATIs in wheat

ATIs proteins are located in the endosperm of the grain kernel where they play an important role in plant defence against pest attacks. Moreover, ATIs were found to be induced by drought stress in grains (Yang et al. 2011).

ATIs are able to inhibit enzymes of common parasites such as mealworms and mealy bugs from digesting starch and protein in wheat. ATIs have diverse conformational structures that are specific to the enzymes of different animal species, leading some ATIs (mainly WMAI) to affect insect pests without acting strongly against human enzymes (Franco et al., 2000) while other ATIs (mainly fractions WDAI 0.19 and 0.38) are active against α -amylase in human saliva and pancreas. The WTAI proteins show a much greater inhibitory activity against *Tenebrio molitor* R-amylase than against the salivary enzyme (Gomez et al., 1989). CM proteins have no *in vitro* α -amylase inhibitory activity individually, but they increase *in vitro* inhibitory activity to insect α -amylases when CM2, CM16 and CM3 are combined to form a tetrameric protein (Wang et al., 2011).

The presence of several different expressed proteins, even in different combinations, with different specificity against exogenous enzymes, could be a general strategy adopted by

cereal seeds to better control the pest attacks (Capocchi et al., 2013).

ATIs may also regulate metabolic processes occurring during seed development and vary among cultivars. These variations affect their protein expression, accessibility and extractability (Finnie et al., 2002).

Besides playing a protective role, ATIs are accumulated at high levels in the grain to function as storage proteins. Unlike gluten proteins that contain large percentages of glutamine and proline, ATIs have more balanced amino acid compositions resulting in a partial compensation for deficiencies in essential amino acids in the gluten proteins. This is an important contribution for both seedling growth and human nutrition (Altenbach et al., 2011).

1.3.3 Role of ATIs in human disease

Beside being a staple food, wheat is associated with adverse reactions in predisposed subjects. Such reactions involve specific immunological mechanisms in the case of wheat allergies and celiac disease (CD), a permanent hypersensitivity to wheat gluten proteins with an autoimmune component. Intolerances, such as Non Celiac Wheat Sensitivity (NCWS), whose mechanism has not been clarified yet, are also described.

ATIs are of high interest because many of them are involved in wheat allergies such as baker's asthma, atopic dermatitis, urticaria, anaphylaxis and gastrointestinal hypersensitivity to wheat (Tatham et al., 2008).

Among the ATIs proteins reacting with patient's IgE, the major allergens are the monomeric 0.28 protein and the subunits CM16 and CM3. In particular CM3 and CM16 are involved in Wheat Dependent Exercise-Induced Anaphylaxis (WDEIA) (Salcedo et al., 2004). WTAI-CM3 is a recognized allergen (Tri a 30) involved in baker's asthma, one of the most important occupational allergies in many countries affecting about 10% of wheat flour workers. Moreover, WTAI-CM3 can bind IgE in sera of patients with atopic dermatitis (Salcedo et al. 2011) or with food allergy to wheat (Larrè et al., 2011). Recently, Sotkovsky et al. (2011) identified trypsin/AAl CMX1/CMX3 proteins as new potential wheat allergens.

The current literature lacks information on the peptide sequences and epitopes responsible

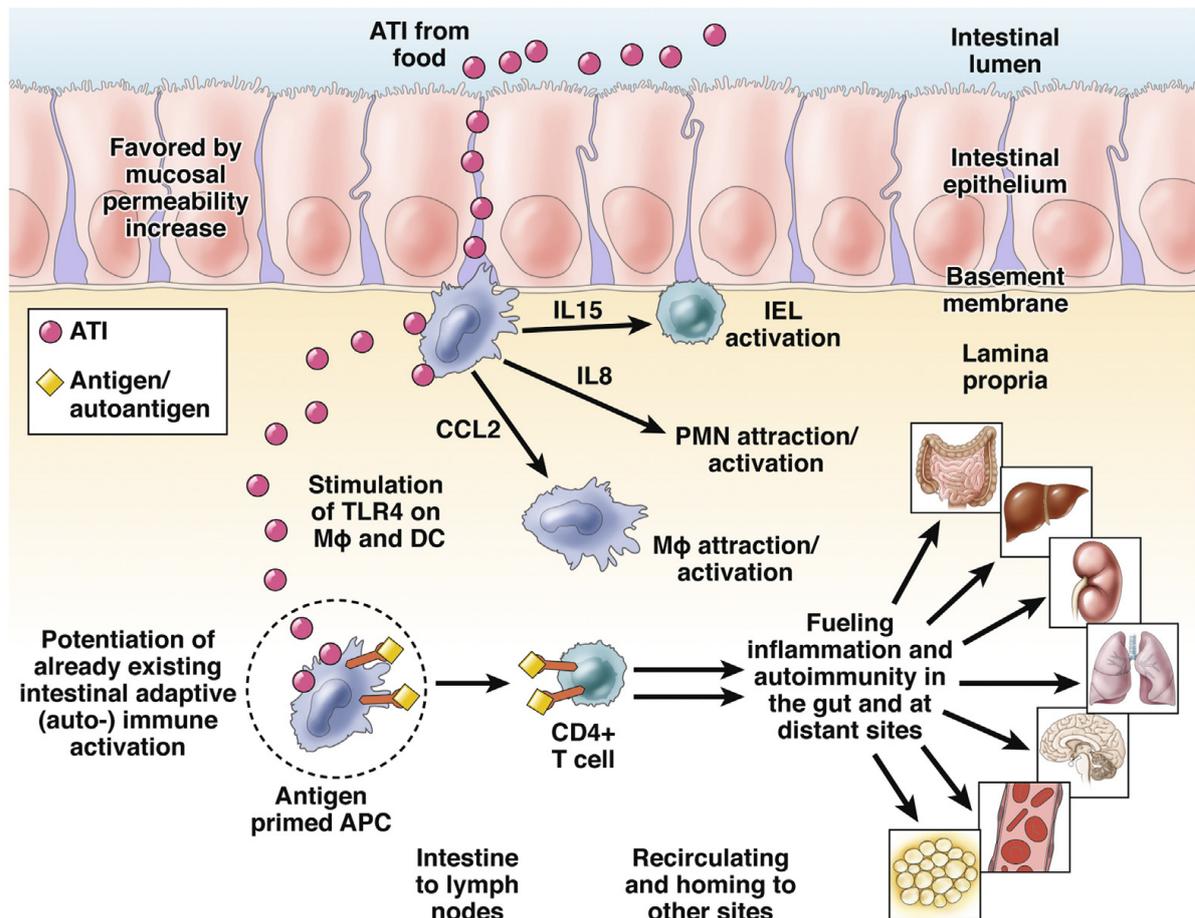
for the allergies triggered by wheat albumin and globulin proteins (Alves et al., 2018). There is only one study which identified a IgE-binding epitope in the primary structure of the inhibitor subunits (Walsh et al., 1989). Using synthetic hexapeptides covering the complete amino acid sequence of WMAI they identified residues 9 to 26 as a high IgE-binding region.

Coeliac disease is a serious autoimmune disease that occurs in genetically predisposed people where the ingestion of gluten leads to damage in the small intestine. It is estimated to affect 1 in 100 people worldwide. Coeliac disease is caused by a reaction to gliadins and glutenins. Besides gluten proteins, amylase-trypsin inhibitors (ATIs) have also been identified to play a role in the onset of a celiac disease. Antibodies directed to some members of ATIs family have been found in sera from patients with celiac disease (Huebener et al., 2015).

Recently, *in vitro* and *in vivo* studies have suggested ATIs (especially CM3 and 0.19), but not gluten, as a major nutritional trigger of human and murine innate immunity on wheat, which increase intestinal and systemic release of cytokines and chemokines. The induction of innate immune responses involves monocytes, macrophages, and dendritic cells activation of the toll-like receptor 4 (TLR4) complex (Junker et al., 2012). Toll-like receptors are ubiquitous in immune cells and their activation may lead to autoimmune, chronic inflammatory and infectious diseases. Additionally, ATIs potentiate antigen presentation on antigen-presenting cells (APCs) and this serves as adjuvant for an on-going adaptive gluten specific T-cell response, as occurs in chronic inflammatory and especially autoimmune diseases. Therefore, ATIs appear to be implicated in the pathogenesis of celiac disease and other autoimmune / inflammatory diseases such as non-celiac wheat sensitivity. However, human clinical trials are needed to confirm these *in vitro* reports.

As shown in Figure 1, the mechanisms by which ATIs from gluten-containing cereals are sensed via TLR4, by *lamina propria* monocytes, macrophages, and dendritic cells. TLR4 signaling leads to the release of inflammatory cytokines and chemokines. ATIs also are adjuvants for adaptive immune reactions in the intestine and possibly nearby lymph nodes, where they also might promote extraintestinal T-cell responses. This could cause adaptive inflammation in remote organs (Fasano et al., 2015).

Fig. 1. ATIs induce innate immune response by producing inflammatory cytokines and chemokines. Moreover, ATIs induce intestinal and extraintestinal adaptive immunity. This could cause adaptive inflammation in remote organs. APC, antigen presenting cell; CCS2, chemokine C-C motif ligand; DC, dendritic cell; IEL, intraepithelial lymphocyte (Fasano et al., 2015)



Plants other than wheat, rye, barley and their early ancestors also contain inhibitors of amylase and trypsin-like activities, but show only minimal or absent inflammatory activity in human cells. Therefore, a gluten-free diet is also essentially ATI-free.

On average, an adult person is exposed up to 1 g of ATIs *per* day, considering a daily wheat intake of 200 g, a wheat average protein content of 10% and a concentration of ATIs reaching the 4% of the total protein content. Moreover, ATIs are present and even enriched in commercial gluten and resist proteolytic digestion, e.g. gastric/ enteric proteases pepsin and trypsin, maintaining their biological activity of TLR4 activation throughout their intestinal passage after oral ingestion. This protease resistance is due to their compact

secondary structure, which is characterized by 5 intra-chain disulphide bonds (Cuccioloni et al., 2017).

There are only a few information about the immunogenic sequences of ATIs. There are short sequences in ATIs proteins which are similar to those in certain γ -gliadin and low molecular weight glutenin proteins, however it is not known if these shared sequences are actually immunogenic (Huebener et al., 2015).

Like other grains, wheat contains several anti-nutritional factors (FAO, 2018) that may interfere with the digestion of wheat products and thereby impact human health.

ATIs can act directly by triggering specific pro-inflammatory receptors (as described above), as well as indirectly by suppressing the activity of digestive enzymes, namely trypsin and alpha-amylase, with two independent binding sites (Cuccioloni et al., 2016; Altenbach et al., 2011).

Inhibition of α -amylase activity leads to accumulation of undigested and unabsorbed starchy constituents which are fermented in the distal gut by the microbial flora and can cause mild to moderate symptoms, including gas production, abdominal distension and flatulence (Gelinas, 2018). These symptoms are also reported by non-celiac wheat sensitivity after assuming wheat products (Bardella et al., 2016). However, the ability to delay carbohydrate digestion and absorption has recently attracted interest in wheat α -amylase inhibitors because they can reduce post-prandial glycemia, and consequently improve the glycemic control in type II diabetic patients. Consequently, α -amylase inhibitors may be a useful functional food, useful for controlling blood glucose in type II diabetic patients with mild hyperglycemia and for preventing diabetes in subjects with impaired glucose tolerance (Kodama et al., 2005)

Regarding the protease inhibitors, these proteins have been considered as “antinutritional compounds” so far, due to their inhibitory activity on digestive enzymes in humans and animals. Horton et al. (2006) reported that unhydrolyzed proteins due to trypsin inhibitors could cause gastric distress and lead to pancreatic hypertrophy or hyperplasia.

Cuccioloni et al. (2016) reported that there are albumin proteins that can increase the levels of gluten-like immunogenic peptides as the result of the inhibition of digestive enzymes and the consequent impaired degradation of dietary cereal proteins.

Besides showing few nutritional problems for healthy people, alpha-amylase inhibitors may have some toxicological effects in the diets of infants (who produce less pancreatic α -amylase than adults) and patients with impaired peptic or gastric function. Similar health concerns are associated with protease inhibitors, although the anti-nutritional status of proteinaceous inhibitors is of less concern as cooking denatures these inhibitors prior to ingestion (Reig-Otero et al., 2018).

1.3.4 ATIs in ancient and modern wheat

Over the last few years, there has been an increasing interest by consumers and farmers in the so-called “ancient” wheat varieties. There isn’t a precise definition of ancient wheat varieties, but generally these varieties have not been modified, while the modern wheat varieties were subjected to cross-breeding programs started during the “Green Revolution” of the last century. These intensive breeding programs were focused on improving the wheat technological properties and the yields. As a result, the modern varieties showed higher gluten content (to improve bread and pasta quality), a reduced susceptibility to diseases and insects, an increased tolerance to environmental stress and reduced heights to decrease both lodging and the maturation cycle. The improved varieties substituted the traditional ones with a consequent reduction of biodiversity. The most common ancient wheat species commercially available are einkorn (*Triticum monococcum*), emmer (*Triticum dicoccum*), khorasan (*Triticum turgidum* ssp. *turanicum*) and spelt (*Triticum spelta*). On top of that, other heritage cultivars of both *Triticum aestivum* and *Triticum durum* are also now available of the market. However, the most part of the ancient wheat varieties was lost during the introduction of the modern wheat varieties.

Over the last few years several studies showed that the introduction of modern wheat varieties had also an impact on human health. These studies highlighted that the ancient wheat varieties show higher nutritional value and higher antioxidant and anti-inflammatory properties compared to the modern wheat varieties (Dinu et al., 2018). The differences on health and nutritional properties between ancient and modern wheat are still under a hot debate.

Regarding the differences existing between the proteins from ancient and modern wheat varieties, recent data demonstrated that during the breeding programs wheat proteins

underwent considerable structural changes (Molberg O. et al., 2005; van den Broeck H.C. et al., 2010). Gliadin and glutenin proteins, which are widely responsible for allergy and celiac disease, were selected for improving the technological quality of wheat.

Junker et al. (2012) stated that selecting for high yielding and highly pest resistant wheat has led to a drastic increase of ATIs content: the old diploid wheat *Triticum monococcum* contains no or little ATIs protein and activity while the modern hexaploid high yielding varieties show the highest activities. However, Kissing Kucek et al. (2015) performed an extensive review of the values reported in the existing literature so far about the human alpha-amylase inhibitory activity in different wheat varieties and concluded that einkorn showed no or very low activity, but durum wheat varieties and emmer inhibited human saliva α -amylase activity at same or even higher levels than modern common wheat varieties. Moreover, they highlighted that there is a big difference in types and quantities of ATIs among genotypes within a species. They concluded that the literature lacked studies which directly compared heritage and modern wheat genotypes for inhibitory activity against human enzymes.

After this review, other papers which evaluated the ATIs activity in ancient and modern wheat varieties were published. A first paper assessed the inflammatory potential of ATIs (determined as interleukin IL-8 release after TLR4 stimulation) extracted from different wheat varieties and different types of wheat-related food products (Zevallos et al., 2017). They found that modern hexaploid wheat contains higher ATIs inflammatory activity than some ancient variants like diploid (einkorn) and tetraploid wheat (emmer, KAMUT[®] khorasan wheat) or older hexaploid variants like spelt. They found also differences among the modern hexaploid grains cultivated in different geographic locations (Europe, North America, and Asia), with the Iranian wheat displaying significantly lower bioactivity. So, they speculated that beside the wheat genotype, the geographic and environmental factors could influence the bioactivity of ATIs. The other major gluten-containing cereals (barley and rye), which express structurally related ATIs, showed bioactivities similar to those of modern wheat. Another recent paper (Gélinas et al., 2018) determined the alpha-amylase inhibitory activity against human α -amylase in several different wheat cultivars and showed that did not vary with respect to ancient or recently developed wheat cultivars.

Thus far, there is little information in literature about the diversity of ATIs between ancient and modern wheat genotypes.

2. AIM

Wheat amylase-trypsin inhibitors (ATIs) are an important family of low molecular weight, water-soluble proteins located in the endosperm of the grain kernel where they play an important role in plant defence against pest attacks (Carbonero et al., 1999). ATIs are able to inhibit enzymes of common parasites such as mealworms and mealy bugs from digesting starch and protein in wheat. Besides playing a protective role, ATIs are accumulated at high levels in the grain to function as storage proteins (Altenbach et al., 2011).

ATIs are of high interest also because these proteins have an important impact on human health. Some ATIs proteins are involved in wheat allergies such as baker's asthma, atopic dermatitis, urticaria, anaphylaxis and gastrointestinal hypersensitivity to wheat (Tatham et al., 2008). Recently, amylase-trypsin inhibitors (ATIs) have been identified to play a role in the onset of a celiac disease. Antibodies directed to some members of ATIs family have been found in sera from patients with celiac disease (Huebener et al., 2015). Other *in vitro* and *in vivo* studies identified ATIs (especially CM3 and 0.19), but not gluten, as a major nutritional trigger of human and murine innate immunity on wheat by the activation of the toll-like receptor 4 (TLR4) complex (Junker et al., 2012). Besides acting directly by triggering specific pro-inflammatory receptors, ATIs can suppress the activity of digestive enzymes, namely trypsin and alpha-amylase, causing intestinal and extra-intestinal symptoms (Cuccioloni et al., 2016; Altenbach et al., 2011).

In recent years, there has been an increasing interest by consumers and farmers in the so-called "ancient" wheat varieties and several studies highlighted the healthier nutritional profile and the higher antioxidant and anti-inflammatory properties of these ancient wheat compared to the modern wheat varieties (Dinu et al., 2018). Although the mechanisms responsible for these beneficial effects are not completely understood, a different protein composition, a higher mineral content and a better phytochemical profile of ancient wheat varieties seem to be the main determinants.

Regarding ATIs, there are few and in some cases contrasting data in literature about the differences existing between ancient and modern wheat varieties.

First, information about genetic sequences (especially of WTAI and CMx genes) is available only for a few types of wheat. Then, regarding ATIs activity, data concerning the inflammatory potential of ATIs (determined as interleukin IL-8 release after TLR4

stimulation) from different wheat varieties showed that modern hexaploid wheat contained higher ATIs inflammatory activity than some ancient variants like diploid (einkorn) and tetraploid wheat (emmer, KAMUT[®] khorasan wheat) or older hexaploid variants like spelt (Zevallos *et al.*, 2017). Another recent paper (Gélinas *et al.*, 2018) determined the alpha-amylase inhibitory activity against human α -amylase in several different wheat cultivars and showed that did not vary with respect to ancient or recently developed wheat cultivars.

However, these studies are missing information about the growing condition of the tested samples.

In the present study, ten selected samples belonging to different *Triticace* species i.e. *T. monococcum*, *T. turgidum* spp. *dicoccum*, *T. turgidum* spp. *turanicum*, *T. turgidum* spp. *durum* (cv. Peliss, cv. Alzada), *T. aestivum* L. (cv. Turkey Red, cv. Judee, cv. Marquis, cv. Vida), *T. aestivum* spp. *spelt* were used for the agronomic trial, for *in vitro* enzymatic assays and for ATIs gene sequencing. These genotypes were selected in order to cover different wheat species with a different genome composition (diploid, tetraploid, hexaploid) and with a different year of release. All wheat samples were grown under organic farming management during three consecutive cropping years at two different locations (Italy and USA).

The main purposes of the present study were:

- to determine the genetic sequences of four representative alpha-amylase/trypsin inhibitor genes (WMAI, WDAI, WTAI-CM3 and CMx) in ten selected ancient and modern wheat genotypes with a different genome composition in order to correlate sequence variants to differences in ATIs activities. The genetic sequences obtained will also be the basis for future studies aimed at identifying specific amino acid sequences in ATIs able to trigger innate and adaptive immune response. Moreover, the sequence variants of ATIs can be useful information for future breeding programs aimed at selecting less inflammatory and more pest resistant wheat varieties;
- to measure the inhibitory activities against both human salivary alpha-amylase and bovine trypsin enzymes to find possible differences among the wheat genotypes. Mammals enzymes have been used in order to evaluate the impact on human health;

- to evidence possible correlation between inhibitory activities and growing conditions. This information can be important in the understanding of the plant defence mechanisms in relation to the effect of both the wheat genotype and abiotic and biotic stress. To the best of our knowledge, this is the first study investigating ancient and modern wheat genotypes with a different ploidy level grown in two different environments (USA, Italy) for three consecutive years.

3. Materials and methods

3.1 Plant material

In the present study, ten selected samples belonging to different *Triticace* species i.e. *T. monococcum*, *T. turgidum* spp. *dicoccum*, *T. turgidum* spp. *turanicum*, *T. turgidum* spp. *durum* (cv. Peliss, cv. Alzada), *T. aestivum* L. (cv. Turkey Red, cv. Judee, cv. Marquis, cv. Vida), *T. aestivum* spp. *spelt* were used for the agronomic trials, *in vitro* enzymatic assays and the characterization of ATIs gene sequences (Table 1). These genotypes were selected in order to cover different wheat species with a different genome composition (diploid, tetraploid, hexaploid) and with a different year of release.

Tab. 1. List of wheat genotypes used in this study.

<i>Triticum</i> species	Genotypes	Genome formula	Year of release	Typology	Provenance	Name used in this study
<i>T. monococcum</i>	Local landrace	AA	/	ancient wheat	Montana State University (Montana, USA)	Einkorn
<i>T. turgidum</i> spp. <i>dicoccum</i>	Local landrace	AABB	/	ancient wheat	Montana State University (Montana, USA)	Emmer
<i>T. turgidum</i> L. ssp. <i>turanicum</i> *	QK-77	AABB	/	ancient wheat	Kamut International (Montana, USA)	Turanicum
<i>T. turgidum</i> ssp. <i>durum</i> Desf.	Peliss	AABB	1900	heritage durum wheat	University of Saskatchewan (Canada)	Peliss
<i>T. turgidum</i> ssp. <i>durum</i> Desf.	Alzada	AABB	2004	modern durum wheat	Montana State University (Montana, USA)	Alzada
<i>T. aestivum</i> ssp. <i>spelt</i>	Local landrace	AABBDD	/	ancient wheat	Montana State University (Montana, USA)	Spelt
<i>T. aestivum</i> L.	Turkey Red	AABBDD	1873	heritage hard red common wheat	Hartland Mills (Kansas, USA)	Turkey Red
<i>T. aestivum</i> L.	Judee	AABBDD	2011	modern hard red common wheat	Montana State University (Montana, USA)	Judee
<i>T. aestivum</i> L.	Marquis	AABBDD	1913	heritage hard red common wheat	Montana State University (Montana, USA)	Marquis
<i>T. aestivum</i> L.	Vida	AABBDD	2006	modern hard red common wheat	Montana State University (Montana, USA)	Vida

* sold under the KAMUT[®] brand. KAMUT[®] is a registered trademark of Kamut International, Ltd, and Kamut Enterprises of Europe, bvba.

The pedigrees of heritage and modern wheat varieties were retrieved in GRIS – Genetic Resources Information System for Wheat and Triticale (<http://wheatpedigree.net>) and are shown in Figure 2.

Figure 2. Pedigrees of heritage and modern wheat genotypes used in this study.

Peliss

PELISS ——— LV-Oran

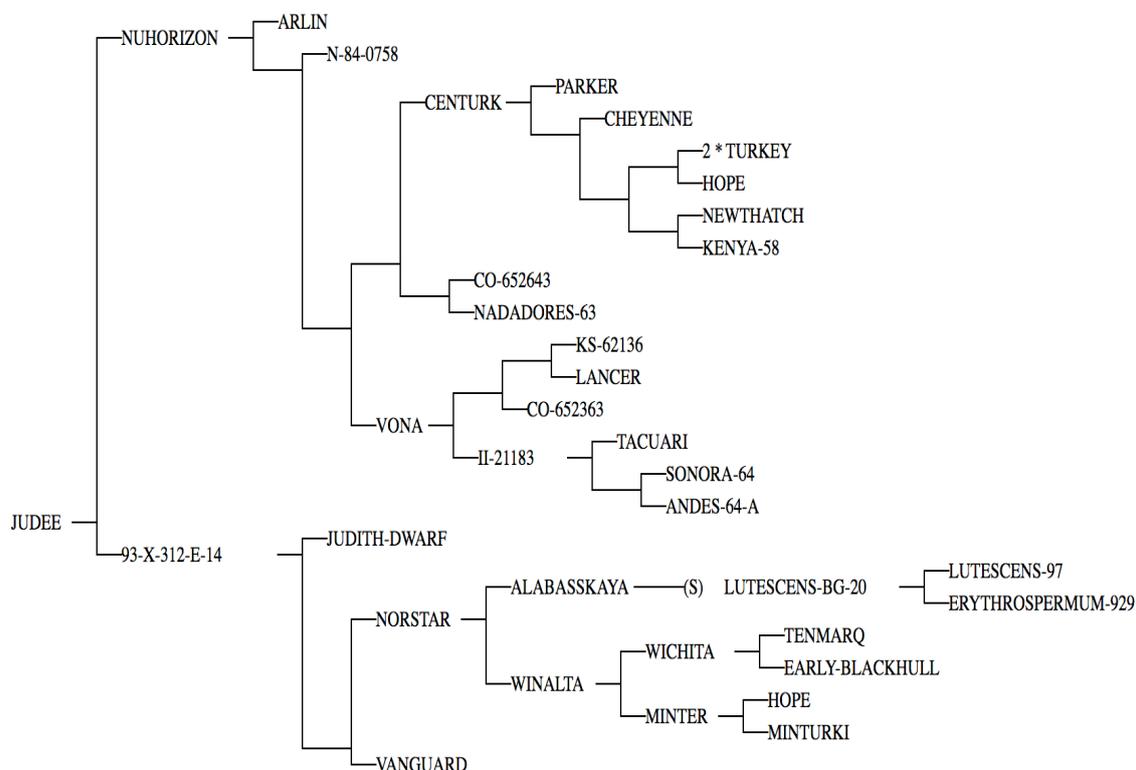
Alzada

ALZADA ——— KOFA ——— (S) Male_Sterile_Facilitated_Recurrent_Selection_Population-DICOCCUM-ALPHA-POP-85-S-1
 ——— MOHAWK ——— (S) Male_Sterile_Facilitated_Recurrent_Selection_Population-883-22-ALPHA-85-CHA

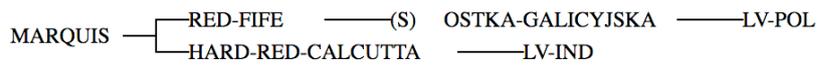
Turkey-Red

TURKEY-RED ——— (S) CRIMEAN ——— LV-RUS

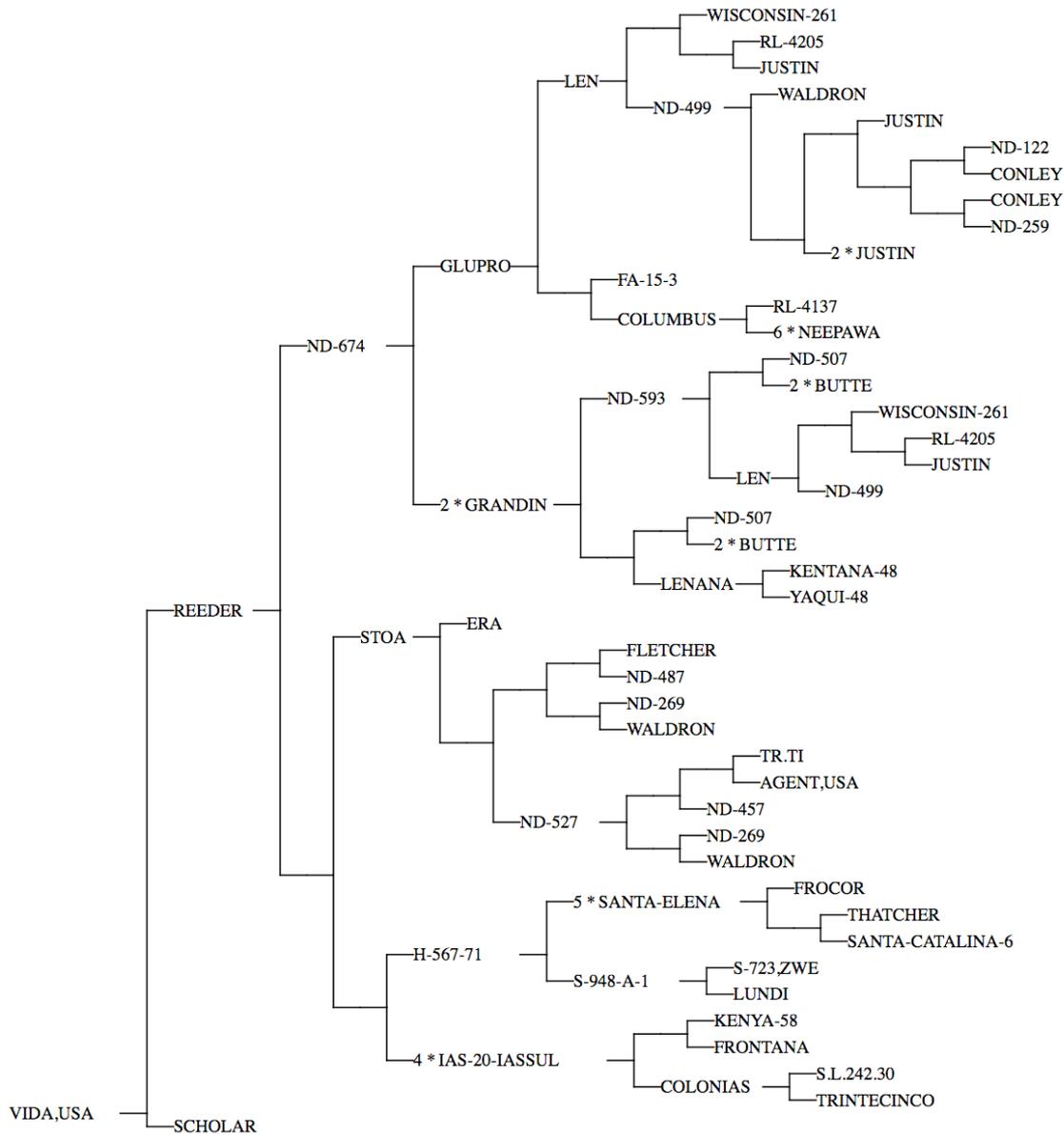
Judee



Marquis



Vida



3.2 DNA isolation and PCR amplification

Ten seeds of each wheat genotype were grown in sprouters for seven days. The leaf tissues were collected and immediately frozen in liquid nitrogen and ground in a mortar with a pestle. Total DNA extraction was performed with the Nucleospin® Plant II kit (Macherey

Nagel, Düren, Germany) following manufacturer's instructions. DNA quality and quantity was measured by NanoPhotometer[®] P-Class (Implen GmbH, München, Germany).

The primers used in this study to amplify WMAI, WDAI, WTAI-CM3 and CMX genes are listed in Table 2. Primers F1, F2, F3, F4, R1 and R2 were previously used by Wang et al. (2008), primers WDAIfor and WDAIrev were previously used by Wang et al. (2005), the other primers were designed *in house* with the software Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>) on the conserved sequences of the coding regions of genes obtained from the GenBank database.

The different combinations of primer sets used for the amplification of WMAI genes are listed in Table 3 with the corresponding annealing temperatures used and the expected amplicon sizes.

WDAI genes were amplified using WDAIfor and WDAIrev primers with an annealing temperature of 58°C. The expected amplicon size was 446 bp.

WTAI-CM3 genes were amplified using CM3for and CM3rev primers and an annealing temperature of 55°C. The expected amplicon size was 647 bp. In order to check for the possible presence of WTAI-CM3 gene sequences also in Einkorn, the internal primers CM3intfor, CM3intrev, CM3intfor2 and CM3intrev2 were designed and used according to the combinations and annealing temperatures listed in Table 4.

CMx genes were amplified using CMXfor and CMXrev primers and an annealing temperature of 60°C. The expected amplicon size was 498 bp.

All the PCR amplifications were performed with Biometra[®] ThermoCycler (Biosense srl, Milan, Italy) in 50 µl volume, consisted of 200-400 ng genomic DNA, 1 µM each primer, 1x HotStar HiFidelity PCR Buffer (containing 1,5 mM MgSO₄ and 0,3 mM dNTPs), 2,5 U of HotStar HiFidelity DNA Polymerase (QIAGEN, Hilden, Germany). For the amplification of WMAI genes in Emmer, Turanicum and Alzada and WTAI-CM3 genes in Einkorn, MgSO₄ was also tested at final concentration of 2 and 3 mM.

The cycling parameters were 95°C for 5 min to pre-denature, followed by 45 cycles of 94°C for 15 sec, annealing temperature (specific for each primer set) for 1 min, and 72°C for 90 sec, and a final extension at 72°C for 10 min.

Tab. 2. List of primers used in this study.

Gene	Name	Sequence
WMAI (NCBI GenBank: AJ223492)	F1	5' – CATAACAGTGGTCCTTGGAGT – 3'
	F2	5' – ATGTGGATGAAGACCGKGT – 3'
	F3	5' – ACAACAATGGCGGTTCGAGTA – 3'
	F4	5' – ATGCTCGTGGCGACAACAAT – 3'
	R1	5' – CACGCACCGCACCATTACTT – 3'
	R2	5' – GACTAGRYGTCCGKATACGC – 3'
	Mfor3	5' – ACTAAATTGAAACAATGTGGAT – 3'
	Mrev2	5' – CACCGCACCAATTAAGAT – 3'
	Mfor4	5' – GCAATGGTGAAGCTCCAGT – 3'
	Mrev5	5' – GATGGGCACCTTGCAGAC – 3'
	Mfor5	5' – AGCTGGCCGACATCAACA – 3'
	Mrev6	5' – GCACCTCCTCCCCTCAC – 3'
	Mrev3	5' – CGCACCAATTAAGATGCAGA – 3'
	Mfor1	5' – CACCACTTATATCCAAGGACCA – 3'
	Mrev4	5' – TTAAGATGCAGATTCGCTTGAC – 3'
	Mfor2	5' – AGCTTGCTTTGATTCTGCTGAT – 3'
Mrev1	5' – TCTCTGAGAGGACACATACACCA – 3'	
WDAI (NCBI GenBank: AK330823)	WDAIfor	5' – CTATGTATGCTCGTGGCGAC – 3'
	WDAIrev	5' – ACTCATTYGCTTGACTAGGC – 3'
WTAI-CM3 (NCBI GenBank: X17574)	CM3for	5' – CGAACCAGACTTGGCTAGAATA – 3'
	CM3rev	5' – ATTCATAGCAGATAGCCACAC – 3'
	CM3intfor	5' – GCGCTGCGCTACTTCATA – 3'
	CM3intrev	5' – GCCGCTCTCACCAACAT – 3'
	CM3intfor2	5' – ACAACAACTTGTGGCACCT – 3'
	CM3intrev2	5' – TGTGAATGGTCGCCAAGT – 3'
CMx (NCBI GenBank: X75608.1)	CMXfor	5' – AAGCACCAGCTCATCCTCTC – 3'
	CMXrev	5' – ATACACATATGCGATTCTGTTCA – 3'

Tab. 3. Combinations of primer sets used to amplify WMAI genes. For each primer set the expected amplicon size (bp) and annealing temperature used (°C) are listed.

Reverse Forward	R1	R2	Mrev2	Mrev3	Mrev5	Mrev6	Mrev4	Mrev1
F1	445 bp 50°C	374 bp 62°C	404 bp 50°C	401 bp 50°C				
F2	529 bp 58°C	458 bp 50°C	488 bp 50°C	485 bp 56°C				
F3	478 bp 50°C	407 bp 56°C	437 bp 52°C	434 bp 55°C				
F4	490 bp 50°C	419 bp 50°C	449 bp 50°C	446 bp 55°C				
Mfor3	543 bp 54°C	472 bp 50°C	502 bp 50°C	502 bp 50°C				
Mfor4				329 bp 54°C	240 bp 54°C	169 bp 54°C		
Mfor5				262 bp 58°C	173 bp 54°C	102 bp 54°C		
Mfor1							550 bp 50°C	
Mfor2								623 bp 50°C

Tab. 4. Combinations of primer sets used to amplify WTAI-CM3 gene in Einkorn. For each primer set the expected amplicon size (bp) and annealing temperature used (°C) are listed.

Reverse Forward	CM3rev	CM3intrev	CM3intrev2
CM3for	647 bp 50°C		
CM3intfor		78 bp 50°C	172 bp 50°C
CM3intfor2		226 bp 50°C	320 bp 50°C

The desired DNA fragment was recovered from 1,5 % agarose gel using QIAquick gel extraction kit (QIAGEN, Hilden, Germany), quantified by NanoPhotometer® P-Class, ligated to the pDrive Cloning Vector and used to transform *E. coli* competent cells by using QIAGEN® PCR Cloning Plus Kit (QIAGEN, Hilden, Germany). Five positive clones for

each sample were screened and sequenced with the Sanger method using the Brilliant Dye Terminator 1.1 kit (NimaGen BV, The Netherlands) and with ABI3730 automated sequencer. Sequences were obtained using the service provided by BMR genomics (Padova, Italy).

3.3 Bioinformatics analysis

3.3.1 Haplotypes identification

Open reading frames were obtained by using ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) and EMBL-EBI server (Madeira et al., 2019) (https://www.ebi.ac.uk/Tools/st/emboss_transeq/). A search for similarity was executed with the BLASTN and BLASTP programs available at the National Centre for Biotechnology Information (NCBI). Haplotypes identification was performed by MEGA software Version 7.0.26 (Kumar et al., 2016). Multiple sequence alignments were performed by Clustal omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), MEGA software Version 7.0.26 and Jalview version 2.11.0 (Waterhouse et al., 2009). Sequence identity analysis was carried out with SIAS server (<http://imed.med.ucm.es/Tools/sias.html>). Amino acid sequences of reference proteins were retrieved from the Universal Protein Resource (Uniprot) (<https://www.uniprot.org/>).

3.3.2 Primary structural analysis of deduced proteins

The signal peptides and the location of their cleavage sites in proteins were determined with SignalP 5.0 (<http://www.cbs.dtu.dk/services/SignalP/>) (Almagro Armenteros et al., 2019).

Expasy's ProtParam online server (<https://web.expasy.org/protparam/>) was used to determine the physiochemical properties of deduced proteins such as molecular weight, theoretical isoelectric point (pI), total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY) (Gasteiger et al., 2006).

The instability index provides the estimate of a protein's stability *in vitro* and is based on the fact that certain dipeptides occur differently in the unstable proteins compared with those in the stable ones.

The Instability index is calculated using the following equation:

$$\text{Instability index} = (10/L) * \text{Sum DIWV}(x[i]x[i + 1])$$

where $i = 1$, L is the length of the sequence and $\text{DIWV}(x[i] x[i + 1])$ is the instability weight value for the dipeptide starting in position i .

The value of extinction coefficient indicates the total of light absorbed by a protein at a specific wavelength. It is computed by the below equation:

$$E(\text{Prot}) = \text{Numb}(\text{Tyr}) * \text{Ext}(\text{Tyr}) + \text{Numb}(\text{Trp}) * \text{Ext}(\text{Trp}) + \text{Numb}(\text{Cys}) * \text{Ext}(\text{Cys})$$

where (for proteins in water measured at 280 nm): $\text{Ext}(\text{Tyr}) = 1490$; $\text{Ext}(\text{Trp}) = 5500$; $\text{Ext}(\text{Cys}) = 125$.

The aliphatic index value for globular proteins was calculated by calculating the volume occupied by aliphatic side chains (alanine, valine, leucine and isoleucine). It is determined using the equation:

$$\text{Aliphatic index} = X(\text{Ala}) + a * X(\text{Val}) + b * (X(\text{Ile}) + X(\text{Leu}))$$

where $X(\text{Ala})$, $X(\text{Val})$, $X(\text{Ile})$, and $X(\text{Leu})$ are mole percent ($100 \times$ mole fraction) of alanine, valine, isoleucine, and leucine. The coefficients a and b are the relative volume of valine side chain ($a = 2.9$) and of Leu/Ile side chains ($b = 3.9$) to the side chain of alanine.

The GRAVY score is the fraction of the sum of hydropathy values of all amino acids and the number of residues in the sequence.

3.3.3 Secondary structural prediction of deduced proteins

PredictProtein (<https://www.predictprotein.org/>), an open resource for online prediction of protein structural and functional features (Yachdav et al., 2014), was used to predict the location of disulfide bridges in proteins with DISULFID method (Ceroni A. et al., 2006) and to predict the secondary structure with the method PROFsec. PROFsec predicts secondary structure elements and solvent accessibility using evolutionary information from multiple sequence alignments and a multi-level system (Rost & Sander, 1994). PROFsec predicts three states of secondary structure: helix (H, includes alpha-, pi- and 3₁₀-helix), beta-strand (E, extended strand in beta-sheet conformation of at least two residues length)

and loop (L). Secondary structure is predicted by a system of neural networks with an expected average accuracy of more than 72%.

3.3.4 Phylogenetic analysis

The phylogenetic relationships between haplotypes were performed using MEGA software Version 7.0.26. The FASTA multiple sequence alignment was used to infer the Neighbor-Joining (NJ) phylogenetic tree with the following parameters, Poisson correction, pairwise deletion and bootstrap (1000 replicates).

3.3.5 Principal component analysis (PCA)

The multiple intercorrelation between all the haplotypes occurring in different wheat genotypes and their deduced amino acid sequence was determined by principal component analysis (PCA). PCA and factor analysis are the most widely used statistical methods to reduce the number of dimensions in data analysis and to investigate multiple intercorrelations between variables (Bilodeau et al., 2002).

PCA is an unsupervised clustering method, which is a powerful tool for analysis of multivariate data, without requiring any knowledge of the dataset (Jambu et al., 1991). PCA was used to transform a number of correlated variables into a smaller number of uncorrelated variables called principal components (Tabacknick et al., 2001). With this method the original space for variable measurements was projected down onto two low-dimensional subspaces. One of these was case-related (ten wheat genotypes), the other was variable-related. The variables were the amino acid substitutions at a specific position. The variable-related subspace was analysed (factor loading) to understand the correlation between the variables and wheat genotypes (principal component).

First, an interim PCA was run with all the variables included and the quality of representation (\cos^2) of each variable was checked. Second, a definitive PCA was run with those variable with a quality of representation higher than 0.5. The PCA analysis was performed using STATISTICA Software v. 7.1 (StatSoft, Tulsa, Oklahoma, USA).

3.4 Agronomic trial

The agronomic trial was carried out using a collection of ten selected samples belonging to different *Triticacae* species (listed in paragraph 3.1), including not-dwarf varieties, ecotypes and populations, cropped before the Green Revolution.

All wheat samples were from field trials conducted under organic farming management during three consecutive cropping years at two locations: at the Quinn Organic Farm located near Big Sandy, Montana, USA (48°02'19"N, 110°00'55"W, 921 m a.s.l.) and at the experimental farm Podere Santa Croce (Argelato, BO) of the Department of Agricultural and Food Science, University of Bologna (Italy). The soil in USA was sandy-loam and in Italy was clay-loam. The three cropping years were 2015-2016; 2016-2017; 2017-2018 in USA and 2016-2017; 2017-2018; 2018-2019 in Italy. In USA, the wheat samples Einkorn, Emmer, Turanicum, Peliss, Alzada, Marquis and Vida were cultivated on a spring cycle (planted at May and harvested in August), while Spelt, Turkey Red and Judee were cultivated on a winter cycle (planted in September and harvested at beginning of August). In Italy, all the ten wheat samples were planted in November and harvested in July.

In Italy several agronomic parameters were also recorded. In particular, in Italy the wheat accessions were grown in duplicate in small plots (1.1 x 6.5 m), cropped according to organic farming. The plots were sown at the seed density of 180 kg/ha. The following parameters were recorded in each single plot: yield, percentage of lodging, weed incidence, disease incidence and disease severity.

Assessment of disease incidence and disease severity in grain genotypes were assessed at BBCH GS73. Disease incidence was calculated as the percentage of ears that are visibly diseased in relation to the total number assessed. For disease severity evaluation, the scale rating of PuraHong et al. (2012) was applied.

The meteorological data (temperature and rain) were recorded at both sites from the National Center of Environmental Information (<https://www.ncdc.noaa.gov/cdo-web/datatools/>) for the American location and from Arpae Emilia Romagna (<https://www.arpae.it/smr/>) for the Italian location.

Grain from different genotypes was tested for their moisture and protein contents by employing the Foss Infratec 1229 NIT spectrophotometer (Global calibration No. WH000003).

3.5 Assessment of mycotoxin levels

All the analyses were performed at the Plant Physiology laboratory of the Department of Agricultural and Food Sciences (DISTAL) of the University of Bologna. The grain samples of each genotype were milled with the domestic stone mill (100% flour extraction) (Billy 200, Hawos Kornmülen gmbh, Germany) and the flours obtained were used for the assessment of the micotoxins content and for the enzymatic tests.

The content of deoxynivalenol was determined with an enzyme-linked immunosorbent assay using the AgraQuant[®] Deoxynivalenol ELISA test kit (Romer Labs GmbH, Austria) and following the manufacturer's instruction. The colour produced was read at 450 nm with the Labsystems Multiskan MS Microplate Reader (Labsystems, Canada).

3.6 Enzymatic assays

For each wheat genotype for each cropping year of two locations two independent replicates were extracted and two replicates from each extracted replicate were tested independently in the following enzymatic assays.

3.6.1 Alpha-amylase inhibitory activity assay

In duplicate, the wheat samples previously milled were extracted using cold water extraction procedure. In this procedure, 100 mg of each of the wheat samples were weighed and then homogenized with 500 µL of distilled water using an automatic shaker at 160 rpm for 30 minutes. The samples were then centrifuged at 7000 rpm for 5 minutes and 400 µL of supernatant was collected. The remaining pellet was again homogenized with 400 µL of distilled water at 160 rpm for 5 minutes and after a centrifugation at 7000 rpm for 5 minutes 350 µL of supernatant was collected and added to the first supernatant. Then, the supernatant was heated at 70 °C for 20 min to inactivate endogenous enzymes, diluted 1:100 in distilled water and used the same day for alpha-amylase inhibitory activity determination.

In duplicate, 1,7 µL of each extract was transferred in a 96 well flat-bottom clear plate and incubated for 10 min at 25°C with 48,3 µL of amylase assay buffer (MAK009A, Sigma-Aldrich Canada Co) containing 20 mU of alpha-amylase from human saliva (A1031; Sigma-Aldrich Canada Co). The control was the amylase assay buffer containing 20 mU of

alpha-amylase from human saliva without the addition of the sample extract. The amylase activity was tested with the Amylase Activity Assay Kit (MAK009, Sigma-Aldrich Canada Co) following the manufacturer instructions. This test is based on the production of a colorimetric product at 405 nm (*p*-nitrophenol) which is proportional to the amount of the substrate (ethylidene-pNP-G7) hydrolysed by the amylase. The absorbance was measured at 405 nm with the Labsystems Multiskan MS Microplate Reader (Labsystems, Canada). The inhibitory activity of the extracts were calculated as percentage with respect to the control according to the following formula:

$$\text{inhibition \%} = [(B_{\text{control}} - B_{\text{sample}})/B_{\text{control}} \times 100]$$

where B is the amount (nmol) of *p*-nitrophenol generated between T_{initial} and T_{final} and calculated using the standards provided by the kit; T_{initial} is the first absorbance measure taken 3 min after the addition of the substrate; T_{final} is the absorbance measured when the most active sample was close, but didn't exceed the end of the linear range of the standard curve.

3.6.2 Trypsin inhibitory activity assay

In duplicate, the wheat samples previously milled were extracted using a sodium acetate buffer solution with a pH 3.8 and ionic strength of 0.02N as reported by Chang and Tsen (1979). One g of each of the wheat samples was weighed and then homogenized with 15 ml of solvent using an automatic shaker at 160 rpm for 1 hour. The samples were then centrifuged at 10000 rpm for 30 minutes at 4°C and 800 µl of supernatant was collected and used the same day for trypsin inhibitory activity determination.

In duplicate, 2 µl of each extract was transferred in a 96 well flat-bottom clear plate and incubated for 20 min at 25°C with 48 µl of trypsin assay buffer (MAK290A, Sigma-Aldrich Canada Co) containing 8 U of trypsin from bovine pancreas (T8003; Sigma-Aldrich Canada Co). The control was the trypsin assay buffer containing 8 U of trypsin from bovine pancreas without the addition of the sample extract. The trypsin activity was tested with the Trypsin Activity Colorimetric Assay Kit (MAK290, Sigma-Aldrich Canada Co) following the manufacturer instructions. This test is based on the production of a colorimetric product at 405 nm (*p*-nitroaniline) which is proportional to the amount of the substrate hydrolysed

by the trypsin. The absorbance was measured at 405 nm with with the Labsystems Multiskan MS Microplate Reader (Labsystems, Canada).

The inhibitory activity of the extracts were calculated as percentage with respect to the control according to the following formula:

$$\text{inhibition \%} = [(B_{\text{control}} - B_{\text{sample}})/B_{\text{control}} \times 100]$$

where B is the amount (nmol) of *p*-nitroaniline generated between T_{initial} and T_{final} and calculated using the standards provided by the kit; T_{initial} is the first absorbance measure taken 3 min after the addition of the substrate; T_{final} is the absorbance measured when the most active sample was close, but didn't exceed the end of the linear range of the standard curve.

3.6.3 Statistical analysis

The general linear model (GLM) was used to assess the variance significance for the fixed (genotype) and the random (year) factors, as well as their interactions, for all measured variables. Tukey's (HSD) test was used to determine differences between means at $P < 0.05$. Analysis was performed using IBM SPSS Statistics 25 software.

The analysis of correlation was calculated using STATISTICA Software v. 7.1, (StatSoft, Tulsa, Oklahoma, USA).

4. Results and discussion

4.1 Monomeric alpha-amylase inhibitors (WMAI)

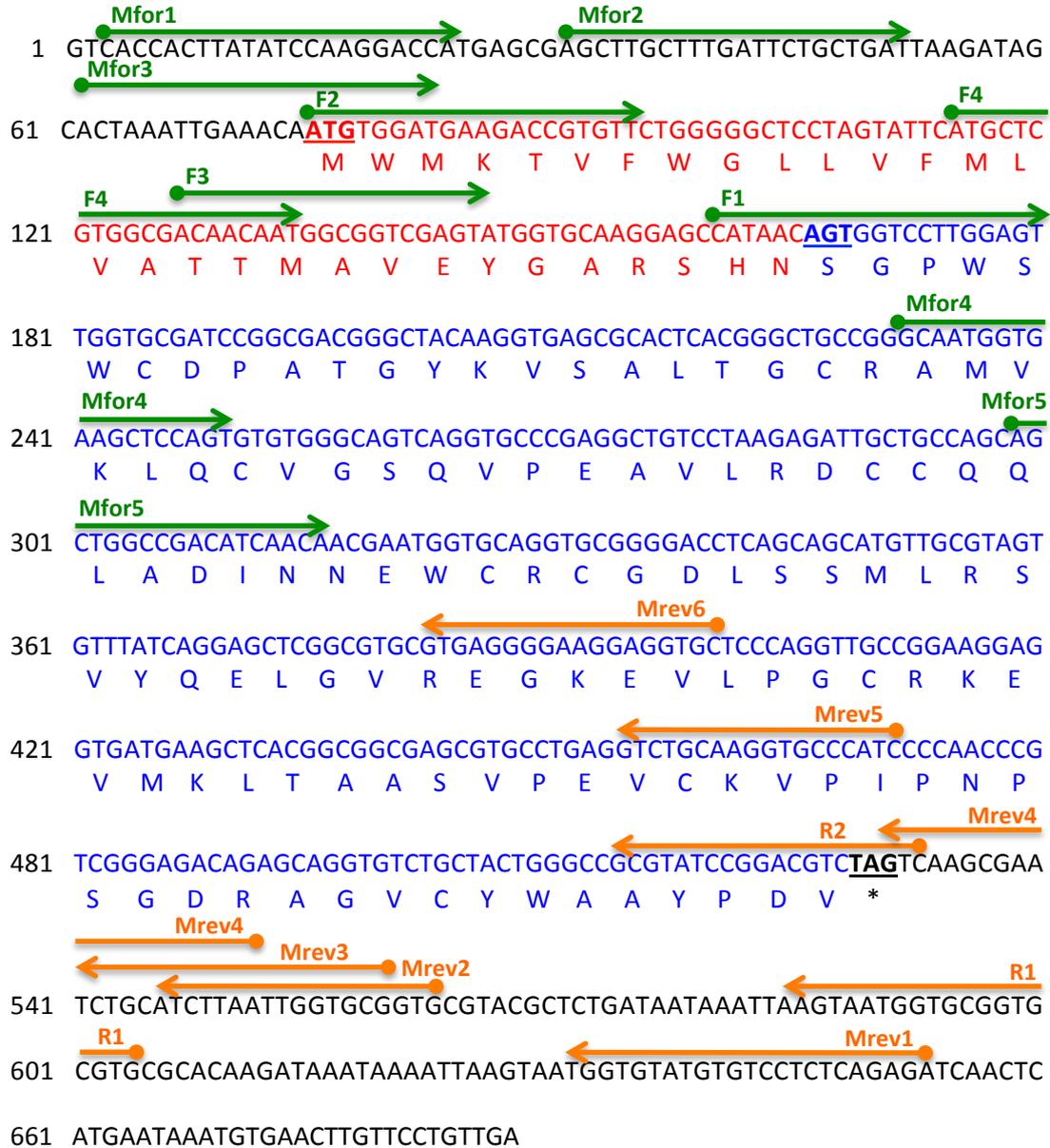
4.1.1 Sequence analysis of monomeric alpha-amylase inhibitors

The primers F2 and R1 (Wang et al., 2008) were used to amplify the ORFs of the monomeric alpha-amylase genes (WMAI) in common wheat samples (Judee, Turkey Red, Marquis, Vida). WMAI ORFs of Einkorn, Spelt and Peliss were amplified using the primer Mfor3 and Mrev2 designed *ex novo* on the sequence of the inhibitor 0.28 (GenBank: AJ223492). All these seven samples gave one desirable PCR product of the expected size (529 bp and 502 bp using F2 + R1 and Mfor3 + Mrev2 respectively).

Emmer, Turanicum and Alzada samples didn't produce any PCR product of the expected size with the two sets of primers cited above. In order to amplify WMAI ORFs from these samples, other primers were used: F1, F3, F4 and R2 (from Wang et al., 2008); Mfor1, Mfor2, Mfor4, Mfor5, Mrev1, Mrev3, Mrev4, Mrev5, Mrev6 (designed *ex novo* on the sequence of the inhibitor 0.28, GenBank: AJ223492). The primers were selected in order to cover the most part of the WMAI ORF region. Figure 3 shows the gene sequence of WMAI 0.28 and the all the primers used in this study to amplify the corresponding ORF.

The WMAI primers were used in many combinations with Emmer, Turanicum and Alzada samples (see Table 3), but it was possible to amplify a fair PCR band close to the expected length only with the primer sets F1–R2; F3-R2 and F2-Mrev2. The bands were cloned and sequenced, but the resulting sequences didn't show significant homology with WMAI genes. In conclusion, using all the primers described above it was not possible to amplify any sequence corresponding to the ORF of WMAI genes in the tetraploid Emmer, Turanicum and Alzada wheat samples. Interestingly, *T. durum* cv. Alzada was selected from a population of *T. dicoccum* (see Figure 2) and like the *T. dicoccum* analysed in this study, it didn't show any WMAI sequence. However, a previous study (Wang et al., 2010) determined the monomeric α -amylase inhibitor gene sequences of 14 populations of wild emmer wheat. Wild emmer wheat (*T. dicoccoides*) is the wild progenitor of modern tetraploid and hexaploid wheat, so in this study it was expected to amplify WMAI sequences in the tetraploid Emmer, Turanicum and Alzada wheat samples.

Fig. 3. The gene sequence of monomeric α -amylase inhibitor 0.28 (AJ223492) and the primers used to amplify WMAI genes. The sequences of signal peptide and mature protein are highlighted in red and blue respectively. The stop codon is marked with an asterisk.



In order to find any monomeric α -amylase inhibitor gene sequences of tetraploid grains from the available databases, the sequence of inhibitor 0.28 (NCBI: AJ223492) was used to perform a blast search against the GenBank non-redundant DNA database (access performed on October 2nd, 2019). It was possible to retrieve 16 monomeric α -amylase inhibitor gene sequences from *T. monococcum*; 53 from *T. aestivum* and 353 from *T. dicoccoides*, but no sequences from *T. durum*, *T. turanicum* and *T. dicoccum*.

Some studies determined the allergen relative abundance in wheat samples and highlighted clear differences in the amount of monomeric α -amylase inhibitors according to the wheat species. Rogniaux et al. (2015) using a targeted MS/MS approach was able to quantify WMAI protein in *T. aestivum*, *T. durum* and *T. monococcum*, but he couldn't detect WMAI protein in KAMUT[®] khorasan wheat sample. Geisslitz et al. (2018) used a new targeted LC-MS/MS method and detected WMAI proteins in three out of eight einkorn cultivars, but below the LOQ, and found that emmer had similar contents of ATI 0.28 as common wheat and spelt. However, they obtained mixed results within durum wheat: two out of eight cultivars had comparable contents of ATI 0.28 to common wheat, spelt and emmer, but the content was near or even below LOD in the other six cultivars. The ATIs contents of the eight durum wheat cultivars showed a higher variability compared to the other wheat species.

The lack of WMAI sequences in *T. durum*, *T. turanicum* and *T. dicoccum* from available databases together with the amplification of WMAI genes in *T. durum* cv. Peliss, but not in other tetraploid wheat in this study confirm that further studies are needed.

Regarding the other seven wheat genotypes (Einkorn, Peliss, Spelt, Turkey Red, Judee, Marquis and Vida) the DNA sequences from five clones per each wheat genotype were determined and used to deduce the full amino acid sequence of the proteins. As described before (Wang et al., 2008), there was no intron in the monomeric α -amylase inhibitor sequences. So, it was possible to isolate the complete coding sequences of monomeric α -amylase inhibitor gene by direct PCR amplification. In this study, no ins/del in the coding region of WMAI were found in any of the wheat samples.

From a total of 35 deduced amino acid sequences, 16 haplotypes have been identified. All the haplotypes were found in only one genotype. The only exception was haplotype HM7 which was shared by all the hexaploid wheat samples (it occurred in 13 out of 25 hexaploid wheat sequences) and consequently it turned out to be the most abundant WMAI haplotype. No pseudogenes were found (table 5 and 6). Each WMAI haplotype is indicated with the code HM followed by a number (Example: HM1). Previously, Wang et al. (2008) found 8 monomeric α -amylase inhibitor amino acid haplotypes in 14 selected common wheat and diploid putative progenitors of common wheat.

All the putatively functional genes encoded for a 151 amino acid protein (30 amino acid

signal peptide and a 121 amino acid mature protein).

Tab. 5. List of WMAI haplotypes.

SAMPLES	PUTATIVELY FUNCTIONAL GENES	PSEUDOGENES	TOTAL SEQUENCED CLONES
Einkorn	5 (HM1, HM2, HM3, HM3, HM3)	/	5
Peliss	5 (HM3, HM3, HM3, HM4, HM5)	/	5
Spelt	5 (HM6, HM7, HM8, HM9, HM10)	/	5
Turkey Red	5 (HM7, HM7, HM7, HM11, HM12)	/	5
Judee	5 (HM7, HM7, HM7, HM13, HM14)	/	5
Marquis	5 (HM7, HM7, HM7, HM7, HM15)	/	5
Vida	5 (HM7, HM7, HM15, HM15, HM16)	/	5

Tab. 6. Table showing in which genotype each WMAI haplotype occurred.

HAPLOTYPES	WHEAT GENOTYPES
HM1	Einkorn
HM2	Einkorn
HM3	Einkorn, Peliss
HM4	Peliss
HM5	Peliss
HM6	Spelt
HM7	Spelt, Turkey Red, Judee, Marquis, Vida
HM8	Spelt

Tab. 6. (continued)

HAPLOTYPES	WHEAT GENOTYPES
HM9	Spelt
HM10	Spelt
HM11	Turkey Red
HM12	Turkey Red
HM13	Judee
HM14	Judee
HM15	Vida, Marquis
HM16	Vida

In this study, after aligning the 16 monomeric α -amylase inhibitor haplotypes from 7 wheat genotypes, 18 nsSNPs were identified. The frequency of nsSNPs was 1 out of 25,2 bases. Wang et al. (2008) found 5 nsSNPs in 111 sequences of hexaploid wheat samples. Among these 5 nsSNPs, only R109G was found in this study (in haplotypes HM9 and HM10 which both occurred in Spelt). Interestingly, R109G is the only one among the nsSNP found by Wang et al. that seems to be able to provide structural changes by determining a different charge and consequently a different relative mobility of gel electrophoresis and differential inhibitory activities (Buonocore et al., 1977). Wang et al. (2010) found 11 nsSNPs in 114 accessions from 14 populations of wild emmer wheat (*T. dicoccoides*) collected in Israel and Golan. Among these, only 2 nsSNPs (amino acid positions 139 and 151) were also found in this study.

The alignment of the 16 monomeric α -amylase inhibitor haplotypes and the position of the nsSNPs are shown in Figure 4.

Sequence alignment shows that all the monomeric α -amylase inhibitors are highly homologous and are part of a monomeric α -amylase inhibitor family including WMAI 0.28. Haplotype HM7, which is the most abundant WMAI haplotype, has the same sequence as WMAI 0.28; the other haplotypes show a percentage of identity with WMAI 0.28 ranging from 96,68% to 99,33%.

The high homology among WMAI sequences confirms previous results and suggests that these inhibitors might have derived from a very limited number of ancestral genes (Wang et al., 2008).

All the deduced proteins of monomeric α -amylase inhibitors had 10 Cys residues which form five disulphide bonds (see Table 7). The disulphide bonds are essential for the inhibitory activity (Carbonero et al., 1999) and in fact the Cys residues are at conserved positions.

It is very important to understand the relationship between an amino acid change and the structure and function of the resulting protein. In particular, regarding WMAI inhibitors, these studies can help to develop new insecticides. It is known that the monomeric α -amylase inhibitors are highly active against α -amylase of *T. molitor* and show a low inhibitory activity against mammalian and some avians α -amylases. The 3-D structure of the complex between α -amylase from *T. molitor* and 0.28 α -amylase inhibitor has been determined (Payan et al., 2004) and three regions of contact have been identified: the N-terminal segment (residues 1-10); residue 53 (inside the second loop segment) and sequence including residues 103-119, which are part of the fourth loop. In particular, the last region, which corresponds to the C-terminal segment of 0.28, plays an important role in filling the central substrate-binding subsites of *T. molitor* α -amylase and in targeting its catalytic residues. Moreover, this region plays an important role in the specificity of 0.28 inhibitor. Figure 5 shows the alignment of the 16 WMAI haplotypes found in this study (after excluding the signal peptides) and WMAI 0.28 highlighting with yellow boxes the three regions of contact. Most of the haplotypes share the same sequence as WMAI 0.28 in these regions, with the following exceptions: HM1 (found in Einkorn) which shows Leu instead of Trp at amino acid position 6; HM9 and HM10 (both found in Spelt) which show Gly instead of Arg at 109; HM13 (found in Judee) which shows Thr instead of Ala at 117.

Another study (Garcia-Maroto et al., 1991) identified two regions critical for the inhibitory activity: the N-terminal sequence (positions 1-6) and the sequence after the CRC motif (positions 57-59). The first region is important for the kinetics of formation of enzyme-inhibitor complex, but is less critical for the inhibitory activity. Differently, the second region has a pivotal role, since any amino acid change occurring in this sequence completely destroys the inhibitory activity of the enzyme. The two regions are highlighted

with red boxes in Figure 5. Only haplotype HM1 (found in einkorn) displays a change at position 6, as described above.

The current literature lacks information on the peptide sequences and epitopes responsible for the allergies triggered by wheat albumin and globulin proteins (Alves et al., 2018). In fact, there is only one study which identified a IgE-binding epitope in the primary structure of the inhibitor subunits (Walsh et al., 1989). Using synthetic hexapeptides covering the complete amino acid sequence of WMAI they identified residues at the position 9 to 26 in the mature protein as a high IgE-binding region. As shown in Figure 5, only two haplotypes had an amino acid change at this level: HM5 (found in Peliss) which has Trp in place of Arg at position 22 and HM14 (found in Judee) which has Gly in place of Ser at position 16. Whether these changes are able to reduce or even delete their immunogenic properties is not possible to say from these data.

Figure 4. Multiple sequence alignment of 16 monomeric α -amylase inhibitor haplotypes obtained from different wheat genotypes and WMAI 0.28. The sequences related to the signal peptide have been included. The dots indicate conserved residues and the letters correspond to the substituted amino acid residues for each alignment gap. Light blue boxes highlight the positions of the 10 Cys residues.

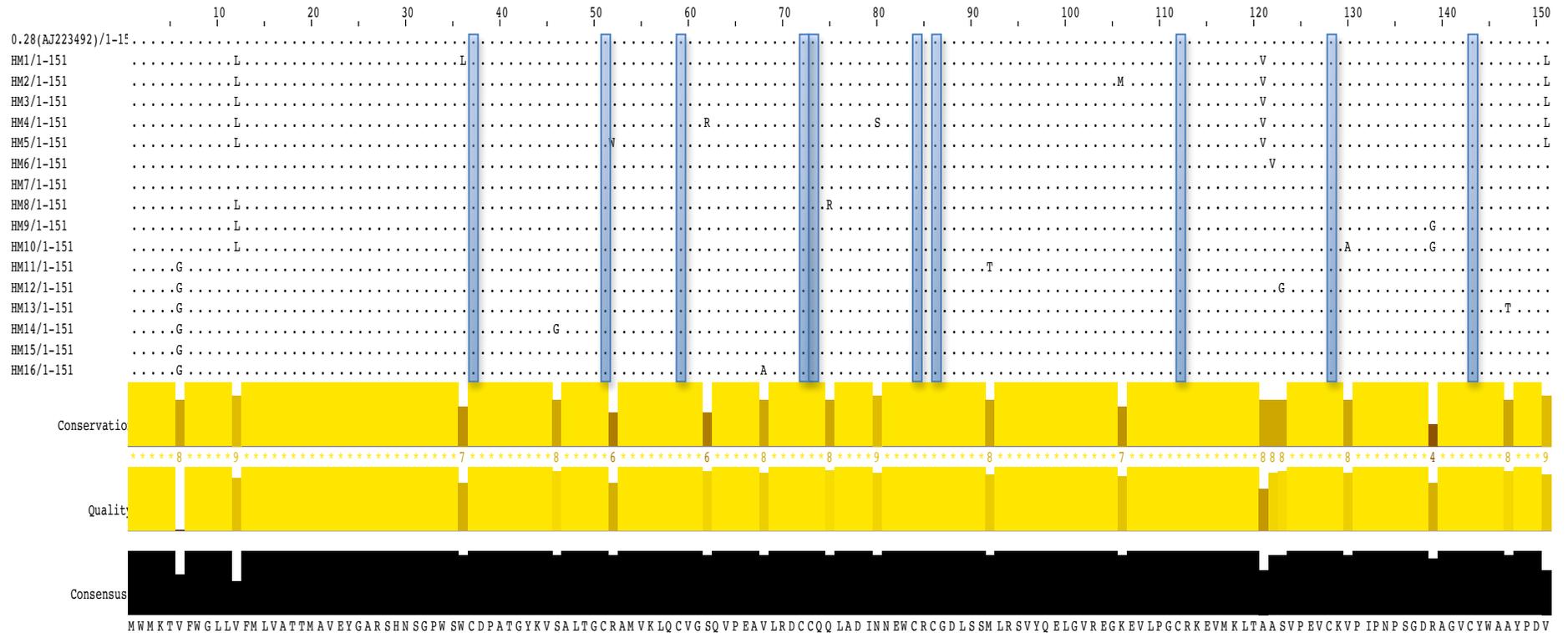
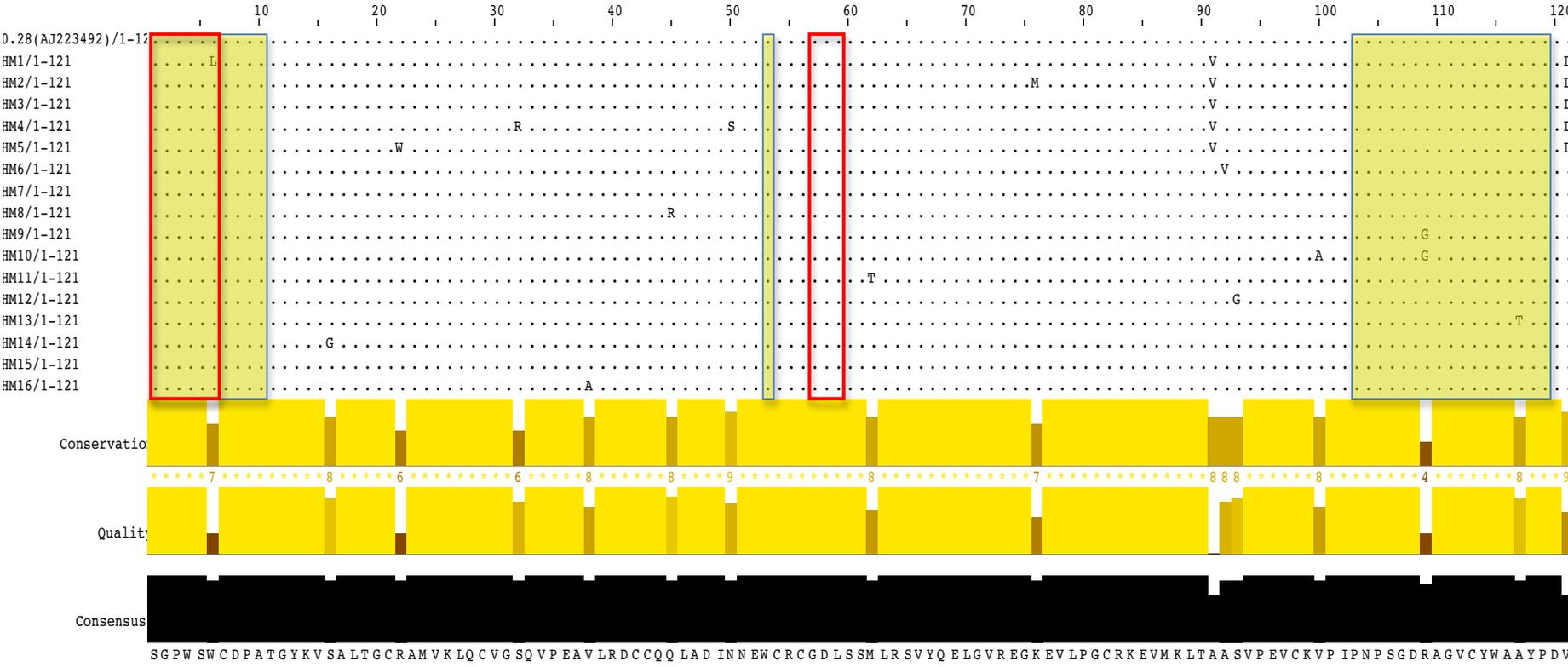


Fig. 5. Multiple sequence alignment of monomeric α -amylase inhibitor haplotypes. The amino acids at the three inhibitor spots are highlighted with yellow boxes. The sequences of the 30 amino acids of the signal peptide are not included in the alignment.



4.1.2 Primary structural analysis of monomeric alpha-amylase inhibitors

The primary structural analysis of the 16 monomeric α -amylase inhibitor haplotypes is shown in Table 7. Only the amino acid sequences related to the mature protein have been used for the analysis.

The calculated isoelectric points (pI) ranged from 5.37 (HM2, HM5, HM9, HM10) to 7.39 (HM4, HM8) and this indicates a higher presence of negatively charged residues. All the haplotypes showed a similar extinction coefficient and aliphatic index. All the haplotypes showed an instability index lower than 40 which suggests thermal stability. The GRAVY score is the sum of hydropathy values of all amino acids divided by the number of residues in the sequence. The GRAVY score is near to zero in all the haplotypes indicating hydrophilicity pattern and good interaction with water.

Tab. 7. Physiochemical properties of the monomeric α -amylase inhibitor haplotypes.

Haplotype	Molecular weight	p.I.	-R (neg residue)	+R (pos residue)	Extinction coefficient *	Instability index	Aliphatic index	GRAVY
HM1	13124.21	6.19	13	13	23085	25.16	82.89	-0.062
HM2	13200.29	5.37	13	12	28585	25.86	79.67	-0.053
HM3	13197.27	6.19	13	13	28585	25.16	79.67	-0.101
HM4	13239.35	7.39	13	14	28585	25.16	79.67	-0.109
HM5	13227.29	5.37	13	12	34085	25.87	79.67	-0.071
HM6	13183.24	6.19	13	13	28585	25.16	78.84	-0.098
HM7	13155.19	6.19	13	13	28585	25.16	77.27	-0.117
HM8	13183.24	7.39	13	14	28585	23.56	77.27	-0.126
HM9	13056.05	5.37	13	12	28585	25.08	77.27	-0.083
HM10	13028.00	5.37	13	12	28585	25.78	75.7	-0.103
HM11	13125.10	6.19	13	13	28585	25.16	77.27	-0.139
HM12	13125.16	6.19	13	13	28585	25.16	77.27	-0.114
HM13	13185.21	6.19	13	13	28585	25.16	76.45	-0.138
HM14	13125.16	6.19	13	13	28585	23.75	77.27	-0.114
HM15	13155.19	6.19	13	13	28585	25.16	77.27	-0.117
HM16	13127.13	6.19	13	13	28585	25.16	75.70	-0.137

* = Extinction coefficients are in units of $M^{-1} \text{ cm}^{-1}$, at 280 nm measured in water and assuming all pairs of Cys residues form cystines.

4.1.3 Secondary structural prediction of monomeric alpha-amylase inhibitors

Three types of secondary structure can be identified in proteins: alpha helices, beta strands and loops. Loops, which include turns, random coil and strands and which link the two main secondary structures (alpha helices and beta strands), are an irregular secondary structure in proteins.

The secondary structure prediction of each haplotype (after excluding the signal peptide sequence) has been performed using PredictProtein server (<https://www.predictprotein.org/>), which evaluates the tendency of each amino acid to be in one of the three conformational states (alpha helix, beta strand or loop) stabilized by hydrogen bonds. The secondary structural prediction of monomeric α -amylase inhibitors is shown in Table 8.

All the deduced proteins of monomeric α -amylase inhibitors had 10 Cys, and their positions were conserved, which indicates that the Cys are important for these inhibitors' three-dimensional structure (Liu et al., 2012). In agreement with the previous literature, all the haplotypes show five disulphide bonds which are covalent bonds between the sulphur atoms of the Cys residues and which influence the folding and the stability of the three-dimensional protein structure. However, the positions of the five disulphide bonds performed by PredictProtein server are slightly different from previous reports: in fact, according to Poerio et al., (1991) Cys29 binds to Cys82 and Cys43 to Cys98.

The percentages of the three types of secondary structure were very similar among the 16 haplotypes: alpha helix ranged from 35.54% to 37.19%, beta strand from 4.96% to 8.26% and loop from 55.37% to 57.85%. These results suggest that these proteins have a higher tendency to exist as random coil.

Tab. 8. Disulphide bridges and secondary structure analysis of monomeric α -amylase inhibitors from different wheat genotypes.

Haplotype	N° disulphide bridges	Disulphide bridges	Secondary structure analysis		
			Alpha helix (%)	Beta strand (%)	Loop (%)
HM1	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	7.44	55.37
HM2	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	4.96	57.85
HM3	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	7.44	55.37
HM4	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	4.96	57.85
HM5	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	4.96	57.85
HM6	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	7.44	55.37
HM7	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	4.96	57.85
HM8	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	7.44	55.37
HM9	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	5.79	57.02
HM10	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	5.79	57.02
HM11	5	7-54, 21-42, 29-43, 56-113, 82-98	35.54	8.26	56.20
HM12	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	4.96	57.85
HM13	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	4.96	57.85
HM14	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	4.96	57.85
HM15	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	4.96	57.85
HM16	5	7-54, 21-42, 29-43, 56-113, 82-98	37.19	7.44	55.37

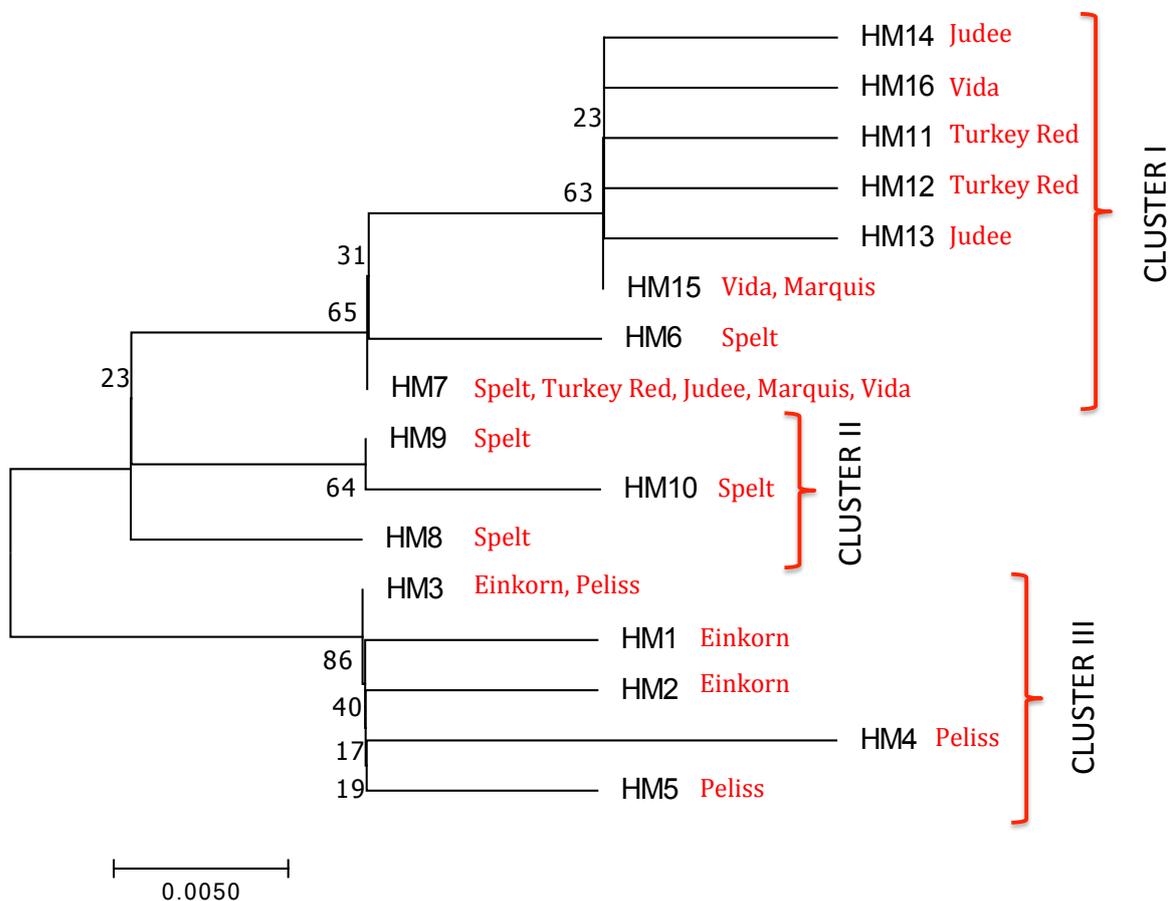
4.1.4 Phylogenetic analysis of monomeric alpha-amylase inhibitors

The Neighbor-Joining method was used to calculate the phylogenetic distances and to construct the phylogenetic tree (Saitou et al., 1987) among all the 16 monomeric α -amylase inhibitor haplotypes (Figure 6). The optimal tree with the sum of branch length = 0.12061355 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein et al., 1985).

The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkanndl et al., 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 151 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 with 1000 bootstrap replicates (Kumar et al., 2016).

Three clusters can be found: cluster I displays only hexaploid wheat genotypes (Turkey Red, Judee, Marquis, Vida and Spelt), cluster II displays all the haplotypes from Spelt, Cluster III displays haplotypes from only Einkorn and Peliss. So, the clusters separate clearly the wheat varieties according to their ploidy: the hexaploid genotypes are displayed only in clusters I and II while the tetraploid Peliss and the diploid Einkorn are together in Cluster III.

Figure 6. Phylogenetic analysis of monomeric α -amylase inhibitors. In red is indicated in which genotype each haplotype occurred.



4.1.5 Principal Component Analysis (PCA) of WMAI sequences

Figure 7 shows the principal component analysis based on deduced amino acid sequences of monomeric α -amylase inhibitors. PCA is an unsupervised clustering method, which is a

powerful tool for analysis of multivariate data, without requiring any specific knowledge of the data. PCA was used to transform a number of correlated variables into a smaller number of uncorrelated variables called principal components.

To visualise relationships between the tested variables, principal component analysis was run for each species in two steps. First, an interim PCA was run with all the variables included (18 nsSNPs), and the quality of representation (Cos2) of each variable was checked. Second, a definitive PCA was run with those variables with a quality of representation higher than 0.5 (8 nsSNPs).

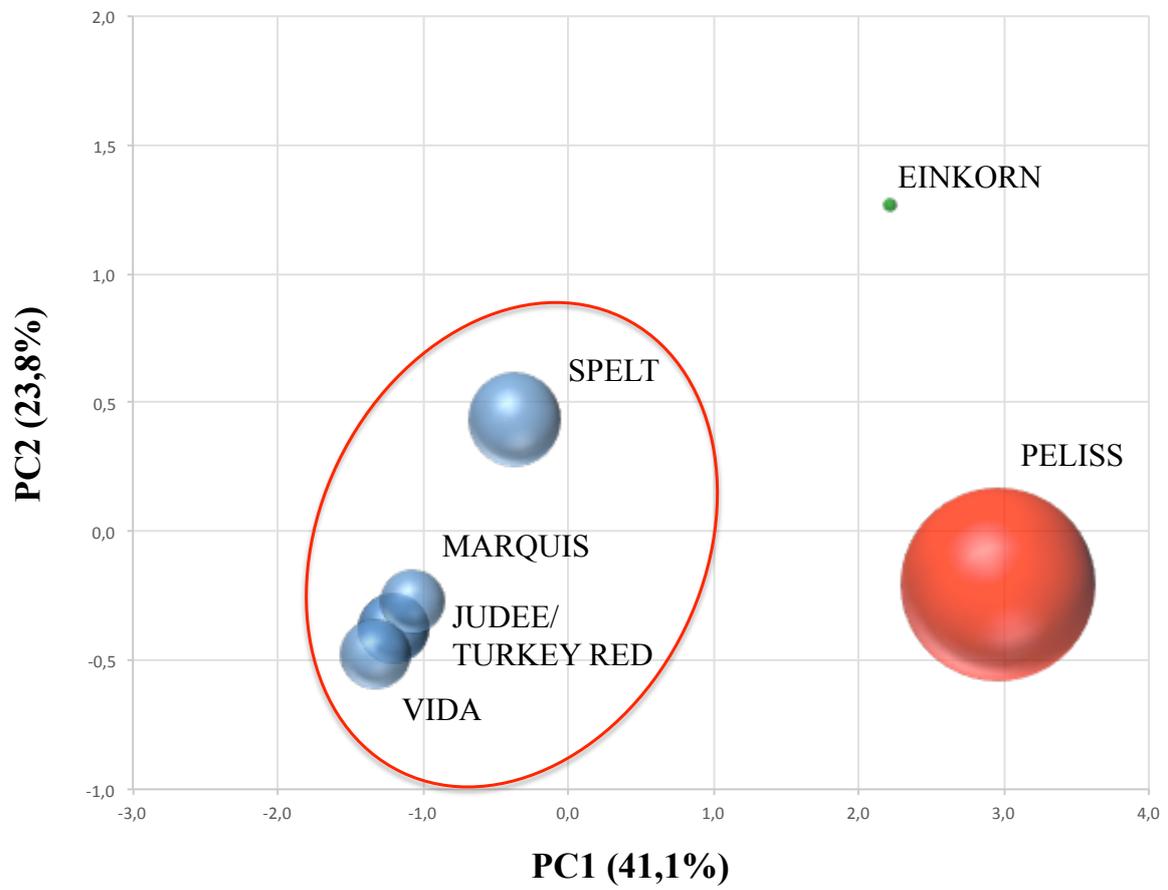
The scatter plot reports the projection of cases (7 wheat genotypes) on the first two components PC1 and PC2 and explained the 64,9% of total variance with the first PC accounting for 41,1% and the second PC for 23,8%. The diameter of each balloon is proportional to the variability observed for the wheat accession. The monomeric α -amylase inhibitors are highly aggregated and have a low variability. Regarding the projection on the first two components, a significant divergence of Einkorn and Peliss can be highlighted compared to the group of hexaploid wheat genotypes (evidenced with a red circle in the figure). This confirms what shown in the phylogenetic tree (Figure 6). More specifically, Peliss and Einkorn differ among them on the projection of the second component. Within the group of hexaploid wheat genotypes, Spelt differs from the others on the projection of the second component. The four common wheat genotypes are very close to each other, and in particular Turkey Red and Judee are overlapped.

Overall, the PCA analysis of monomeric α -amylase inhibitors is in line with the ploidy level of the wheat samples analysed.

The tetraploid Alzada, Emmer and Turanicum are not included in the analyses because it was not possible to amplify any sequence related to WMAI genes.

Fig. 7. Principal component analysis based on deduced amino acid sequences of monomeric α -amylase inhibitors. The scatter plot reports the projection of cases (7 wheat genotypes) on the first two components PC1 and PC2 (accounting for 64,9% of total variability): the diameter of each balloon is proportional to the variability observed for the wheat genotype. Each colour of the balloons corresponds to a different wheat ploidy: green for

diploid wheat; red for tetraploid wheat; blue for hexaploid wheat.



4.2 Dimeric alpha-amylase inhibitors (WDAI)

4.2.1 Sequence analysis of dimeric alpha-amylase inhibitors

The primers WDAIfor and WDAIrev were used to amplify the ORF of the dimeric alpha-amylase genes (WDAI). All the ten samples gave the PCR product of the expected size (~509 bp).

The DNA sequences from five clones per each wheat genotype were determined and used to deduce the full amino-acid sequence of the proteins.

From a total of 50 deduced amino-acid sequences, 35 haplotypes have been identified, among which 31 haplotypes were found in only one genotype and 6 haplotypes were identified as pseudogenes, due to the presence of one or more in-frame stop codons as a consequence of a single nucleotide polymorphism (substitution, insertion or deletion), although we cannot predict from the genomic data whether these sequences are being expressed (table 9 and 10). Each WDAI haplotype is indicated with the code HD followed by a number (Example: HD1); the pseudogenes are indicated with a p after the number (Example: HD19p). All the sequences obtained from the einkorn sample showed an insertion of C at position 160 (from the start codon ATG) which resulted in a premature stop codon and the impossibility to synthesize the correct mature protein. The same insertion of C was observed in the DNA sequences from all the *T. monococcum* accessions and from above all *T. boeoticum* accessions in a previous study (Wang et al., 2007b). The absence of gene sequences coding for functional WDAI genes in *T. monococcum* is in agreement with functional studies. Rogniaux et al. (2015) using a targeted MS/MS approach were not able to detect WDAI proteins in *T. monococcum* and he suggested that its WDAI gene is most likely silenced. Geisslitz et al. (2018) developed a new-targeted LC-MS/MS method and could detect WDAI proteins in only three out of eight einkorn cultivars, but with low contents.

All the putatively functional genes were 426 bp long, encoding for a 141 amino acids protein (17 amino acids signal peptide and a 124 amino acids mature protein). Haplotype HD18 was found in 6 gene samples from 5 different genotypes and it was the most abundant.

Previously, Pandey et al. (2016) found 61 dimeric α -amylase inhibitor haplotypes in 24 selected *Triticeae* species (Indian bread wheat, durum wheat and wild relatives); Wang et al.

(2008) found 74 dimeric α -amylase inhibitor haplotypes in 205 *T. dicoccoides* accessions and Wang et al. (2007) found 21 dimeric α -amylase inhibitor haplotypes in 68 diploid wheat accessions.

Tab. 9. List of WDAI haplotypes.

SAMPLES	PUTATIVELY FUNCTIONAL GENES	PSEUDOGENES	TOTAL SEQUENCED CLONES
Einkorn	/	5 (HD28p, HD28p, HD29p, HD34p, HD35p)	5
Emmer	5 (HD8, HD18, HD26, HD26, HD27)	/	5
Turanicum	5 (HD18, HD18, HD20, HD21, HD22)	/	5
Peliss	5 (HD1, HD2, HD4, HD18, HD32)	/	5
Alzada	5 (HD8, HD18, HD23, HD24, HD25)	/	5
Spelt	3 (HD15, HD15, HD31)	2 (HD30p, HD30p)	5
Turkey Red	5 (HD5, HD6, HD7, HD8, HD9)	/	5
Judee	5 (HD9, HD10, HD11, HD12, HD33)	/	5
Marquis	5 (HD13, HD14, HD15, HD15, HD16)	/	5
Vida	4 (HD3, HD9, HD17, HD18)	1 (HD19p)	5

Tab. 10. Table showing in which genotype each WDAI haplotype was found.

HAPLOTYPES	GENOTYPES
HD1	Peliss
HD2	Peliss
HD3	Vida
HD4	Peliss
HD5	Turkey Red
HD6	Turkey Red
HD7	Turkey Red
HD8	Turkey Red, Alzada, Emmer
HD9	Turkey Red, Judee, Vida
HD10	Judee
HD11	Judee
HD12	Judee
HD13	Marquis
HD14	Marquis
HD15	Spelt, Marquis
HD16	Marquis
HD17	Vida
HD18	Peliss , Vida, Turanicum, Alzada , Emmer
HD19p	Vida
HD20	Turanicum
HD21	Turanicum
HD22	Turanicum
HD23	Alzada
HD24	Alzada
HD25	Alzada
HD26	Emmer
HD27	Emmer
HD28p	Einkorn
HD29p	Einkorn
HD30p	Spelt
HD31	Spelt
HD32	Peliss
HD33	Judee
HD34p	Einkorn
HD35p	Einkorn

In this study, after aligning the 35 dimeric α -amylase inhibitor haplotypes from ten wheat genotypes, 69 nsSNPs were identified. The frequency of nsSNPs was 1 out of 6,2 bases, which is ten fold higher than what Pandey et al. (2016) previously found in different *Triticeae* genomes.

The alignment of the 35 dimeric α -amylase inhibitor haplotypes and the position of the nsSNPs are shown in Figure 8. The pseudogenes were also considered in this analysis and the genetic sequences used to deduce the amino acid sequences for the alignment were selected as follows: in case the internal stop codon was created by a substitution of a single nucleotide, then an asterisk was added in place of the stop codon and the sequence until the end of the ORF of the corresponding functional gene was used; in case the internal stop codon was created by a deletion or insertion of a single nucleotide, then the sequence without this nucleotide change was used.

The alignment in Figure 8 shows that the WDAI gene sequences had more divergence than the WMAI gene sequences of this study. Previously Liu et al. (2012) concluded that WMAI genes are more conserved than WDAI genes.

Fontanini et al. (2007) isolated in a emmer sample two dimeric alpha-amylase inhibitor proteins showing some of the amino acid changes highlighted in this study, in particular: His⁴⁷ --> Asp⁴⁷ or Asn⁴⁷; Asn²⁹ --> Asp²⁹; Glu⁴⁰ --> Asp⁴⁰; Val¹⁰⁵ --> Ile¹⁰⁵.

All the deduced proteins of WDAI but HD1 and ATI 0.53 had 10 Cys residues which form five disulphide bonds. The disulphide bonds are essential for the inhibitory activity (Poerio et al., 1991) and in fact the Cys residues are at conserved positions.

Haplotype 1 (found in Peliss) lacks one Cys residue at position 58 which was substituted by a Serine residue and consequently forms only four disulphide bonds (see Table 12).

Considering the structural features of the inhibitors, the Cys residues are not the only significant amino acids. Franco et al. (2000) and Payan et al. (2004) highlighted three inhibitor spots important for the inhibitory activity by studying the modelled complex of human salivary α -amylase (HSA) with 0.19 inhibitor. Considering this model, the important amino acids are His47, which is close to residue Glu349 of HSA; the second is Ser49, which is packed tight with Lys352 and Asp356 of the long loop of HSA; the third spot concerns the sequence Val104 – Val105 – Asp106 – Ala107 within the protruding long loop of 0.19,

which is believed to play an important role on the specificity of the inhibitory activity in both mammalian and insect α -amylases. Amino acids changes at the level of these three spots can result in a different ability to inhibit human α -amylases. As showed in Figure 9, the only haplotypes displaying all the above mentioned amino acids residues at the three inhibitor spots are HD5 found in Turkey Red, HD9 found in Turkey Red, Judee and Vida; HD11 found in Judee; HD15 found in Spelt and Marquis; HD17 found in Vida. Interestingly, all these haplotypes were found only in hexaploid wheat genotypes, which showed in fact the highest α -amylase inhibitory activity *in vitro* (see paragraph 4.8.1).

Fig. 8. Multiple sequence alignment of 35 WDAI haplotypes obtained from ten wheat genotypes with WDAI 0.19 (Uniprot: P01085) and 0.53 (Uniprot: P01084). The asterisks * show in-frame stop codons of pseudogenes. The blue and red triangles show the positions of insertions and deletions respectively found in the nucleotide sequences used to deduce the amino-acid sequences. The dots indicate conserved residues; the letters correspond to the substituted amino acid residues for each alignment gap. Light blue boxes highlight the positions of Cys residues.

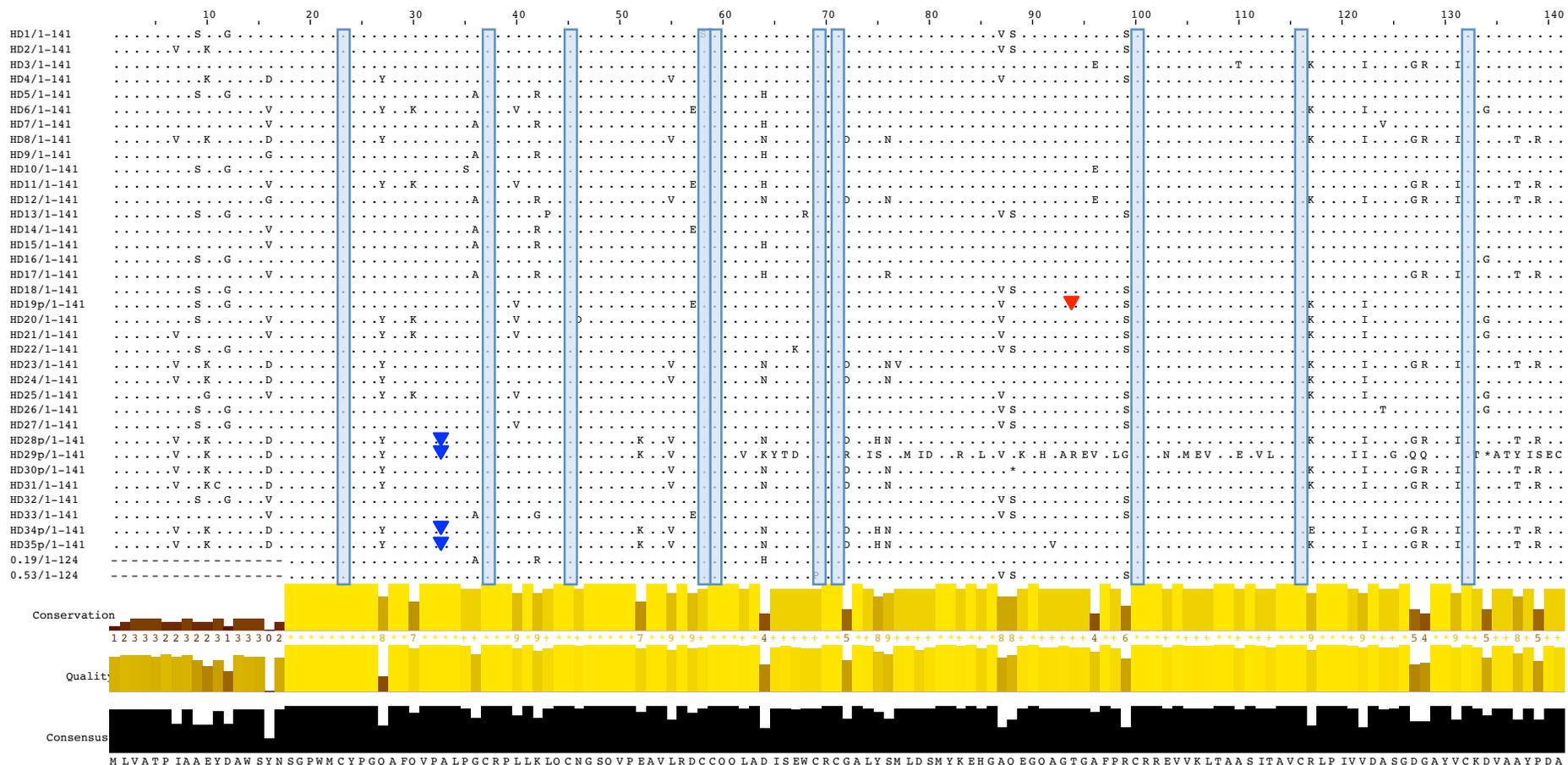
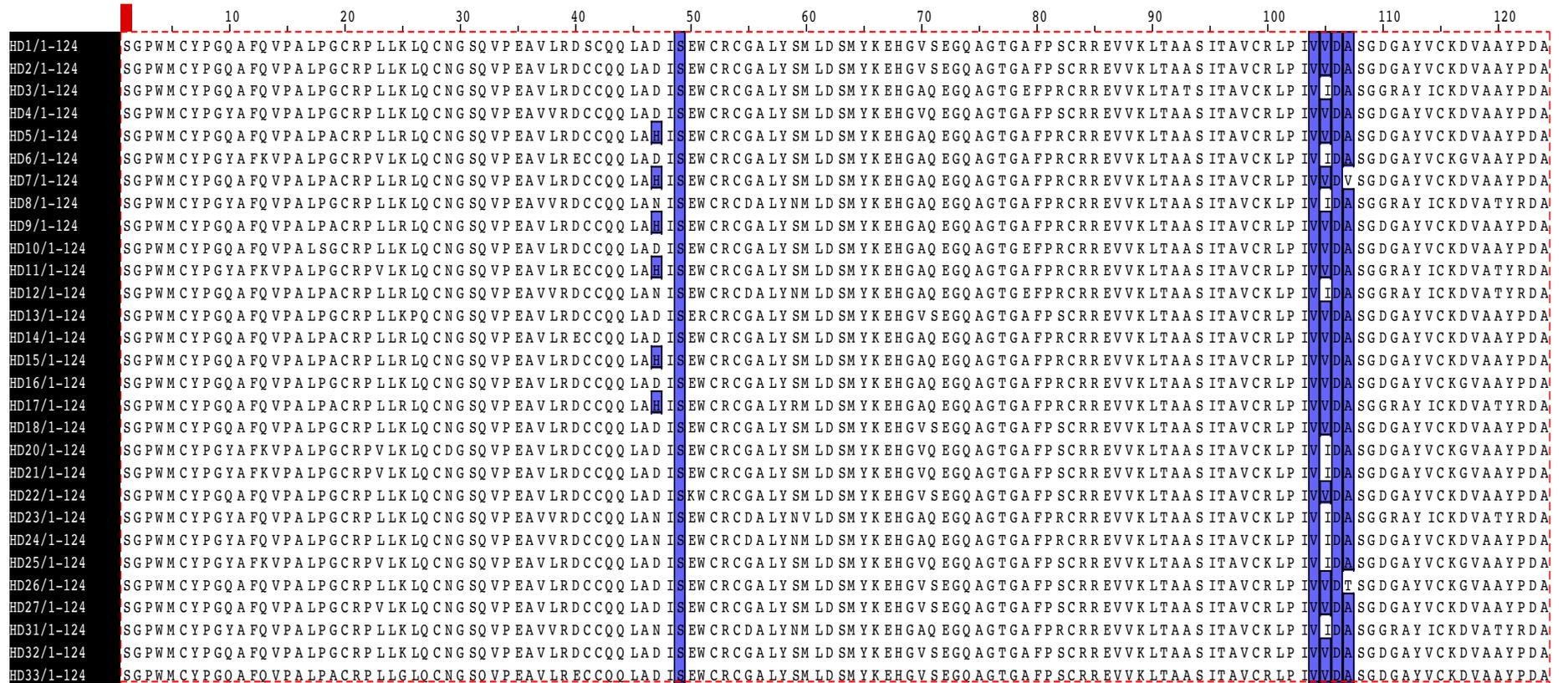
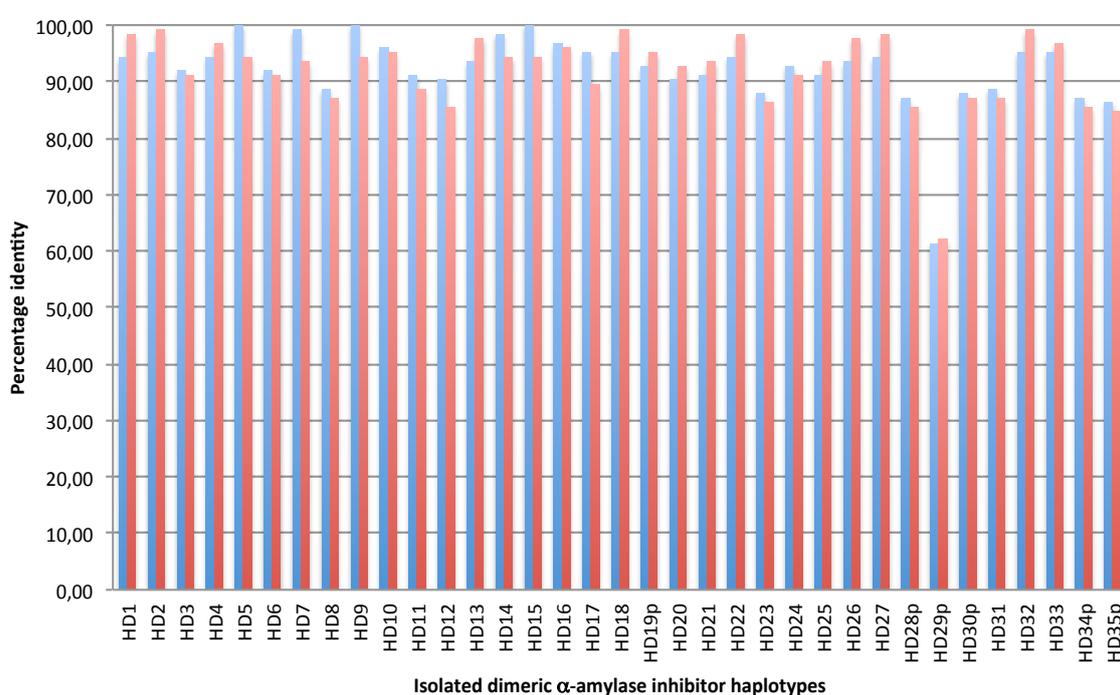


Fig. 9. Multiple sequence alignment of dimeric α -amylase inhibitor haplotypes (the pseudogenes are not included). The amino acids at the three inhibitor spots are highlighted with violet boxes. The sequences of the 17 amino acids of the signal peptide are not included in the alignment.



Considering the percentage of identity with the two amino acid sequences of the well-studied dimeric α -amylase inhibitors 0.19 and 0.53, only the haplotype HD9 had a 100% of identity with 0.19 (Fig 10). However, the level of homology is high with both 0.19 and 0.53 (ranged from 87,09% to 100% for 0.19 and from 84,67% to 99,19% for 0.53), with the exception of the pseudogene HD29p.

Fig. 10. Percentage of sequence identity of deduced amino acid sequence of 35 dimeric α -amylase haplotypes with 0.19 inhibitor (blue) and 0.53 inhibitor (red) amino acid sequences.



4.2.2 Primary structural analysis of dimeric alpha-amylase inhibitors

The primary structural analysis of the 29 dimeric α -amylase inhibitor haplotypes is shown in Table 11. The six pseudogenes have been excluded from this analysis. Only the amino acid sequences related to the mature protein have been used for the analysis.

The calculated isoelectric points (pI) ranged from 4,96 (HD33) to 8,58 (HD17) and this indicates a higher presence of negatively charged residues. All the haplotypes showed a similar extinction coefficient with the exception of HD13 (found in Marquis) which showed the lowest value (13575) suggesting a lower presence of Phe, Tyr and Trp residues.

Interestingly, HD13 showed also the lowest aliphatic index (75,56). The protein aliphatic index is defined as the relative volume occupied by aliphatic side chains (Ala, Val, Ile, and Leu). A high aliphatic index is considered a positive factor for the increase of thermo stability of globular proteins. Moreover, a larger presence of Phe, Tyr e Trp promotes the biophysical assays in solution (protein-protein and protein-ligand interactions). These factors taken together (lowest values of both extinction coefficient and aliphatic index) could have a negative effect on the functionality of HD13.

All the haplotypes showed an instability index lower than 40 which suggests thermal stability. The only two exceptions are HD12 and HD33 (both found in Judee) with instability indices of 41,31 and 43,33 respectively. The GRAVY score is the sum of hydropathy values of all amino acids divided by the number of residues in the sequence. The GRAVY score is near to zero in all the haplotypes indicating hydrophilicity pattern and better interaction with water.

Tab. 11. Physiochemical properties of the dimeric α -amylase inhibitor haplotypes.

Haplotype	Molecular weight	p.I.	-R (neg residue)	+R (pos residue)	Extinction coefficient *	Instability index	Aliphatic index	GRAVY
HD1	13175,11	5,23	12	10	18950	37,50	78,71	0,044
HD2	13191,17	5,23	12	10	19075	33,32	78,71	0,071
HD3	13402,48	6,5	12	12	19075	30,13	77,18	-0,061
HD4	13253,24	5,23	12	10	20565	32,99	77,90	0,070
HD5	13337,37	6,66	11	11	19075	39,41	77,98	0,015
HD6	13236,35	7,42	11	12	20565	32,72	77,18	0,050
HD7	13365,43	6,66	11	11	19075	39,41	79,52	0,035
HD8	13508,61	8,03	11	13	20565	34,44	77,18	-0,068
HD9	13337,37	6,66	11	11	19075	39,41	77,98	0,015
HD10	13321,28	5,28	13	11	19075	26,37	76,37	-0,036
HD11	13474,64	8,56	10	14	20565	36,35	76,37	-0,029
HD12	13573,64	7,43	12	13	19075	41,31	77,18	-0,115

Tab. 11. (continued)

Haplotype	Molecular weight	p.I.	-R (neg residue)	+R (pos residue)	Extinction coefficient *	Instability index	Aliphatic index	GRAVY
HD13	13145,10	5,71	12	11	13575	33,53	75,56	-0,002
HD14	13329,35	5,73	12	11	19075	39,41	77,98	0,013
HD15	13337,37	6,66	11	11	19075	39,41	77,98	0,015
HD16	13215,24	6,49	11	11	19075	27,16	77,18	0,025
HD17	13550,71	8,58	10	14	19075	37,97	77,98	-0,064
HD18	13191,17	5,23	12	10	19075	33,32	78,71	0,071
HD20	13182,25	5,71	12	11	20565	37,42	78,71	0,099
HD21	13181,26	6,49	11	11	20565	34,65	78,71	0,099
HD22	13190,23	6,49	11	11	19075	32,98	78,71	0,068
HD23	13476,55	8,03	11	13	20565	34,44	79,52	-0,049
HD24	13364,39	5,71	12	11	20565	34,51	77,18	-0,019
HD25	13181,26	6,49	11	11	20565	34,65	78,71	0,099
HD26	13163,16	5,70	11	10	19075	31,42	77,90	0,076
HD27	13177,15	5,23	12	10	19075	34,87	77,90	0,074
HD31	13508,61	8,03	11	13	20565	34,44	77,18	-0,068
HD32	13191,17	5,23	12	10	19075	33,32	78,71	0,071
HD33	13148,10	4,96	12	9	19075	43,33	79,52	0,117

* = Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water and assuming all pairs of Cys residues form cystines.

4.2.3 Secondary structural prediction of dimeric alpha-amylase inhibitors

The secondary structural prediction of dimeric α -amylase inhibitors (after excluding the signal peptide sequence) is shown in Table 12.

Tab. 12. Disulphide bridges and secondary structure analysis for dimeric α -amylase inhibitors from different wheat genotypes.

Haplotype	N° disulphide bridges	Disulphide bridges	Secondary structure analysis		
			Alpha helix (%)	Beta strand (%)	Loop (%)
HD1	4	6-52, 20-54, 42-99, 83-115	34,0	7,1	58,9
HD2	5	6-52, 20-41, 28-42, 54-115, 83-99	48,9	3,6	47,5
HD3	5	6-52, 20-41, 28-42, 54-115, 83-99	51,1	1,4	47,5
HD4	5	6-52, 20-41, 28-42, 54-115, 83-99	48,9	2,1	48,9
HD5	5	6-52, 20-41, 28-42, 54-115, 83-99	46,8	3,5	49,6
HD6	5	6-52, 20-41, 28-42, 54-115, 83-99	51,1	1,4	47,5
HD7	5	6-52, 20-41, 28-42, 54-115, 83-99	45,4	5,0	49,6
HD8	5	6-52, 20-41, 28-42, 54-115, 83-99	46,1	5,0	48,9
HD9	5	6-52, 20-41, 28-42, 54-115, 83-99	51,1	1,4	47,5
HD10	5	6-52, 20-41, 28-42, 54-115, 83-99	44,0	6,4	49,6
HD11	5	6-52, 20-41, 28-42, 54-115, 83-99	46,1	2,8	51,1
HD12	5	6-52, 20-41, 28-42, 54-115, 83-99	46,1	5,0	48,9
HD13	5	6-52, 20-41, 28-42, 54-115, 83-99	36,9	9,2	53,9
HD14	5	6-52, 20-41, 28-42, 54-115, 83-99	50,4	1,4	48,2
HD15	5	6-52, 20-41, 28-42, 54-115, 83-99	48,9	2,8	48,2
HD16	5	6-52, 20-41, 28-42, 54-115, 83-99	49,6	5,0	45,4
HD17	5	6-52, 20-41, 28-42, 54-115, 83-99	46,8	4,3	48,9
HD18	5	6-52, 20-41, 28-42, 54-115, 83-99	45,4	7,8	46,8
HD20	5	6-52, 20-41, 28-42, 54-115, 83-99	49,6	2,8	47,5
HD21	5	6-52, 20-41, 28-42, 54-115, 83-99	48,2	4,3	47,5
HD22	5	6-52, 20-41, 28-42, 54-115, 83-99	45,4	7,8	46,8
HD23	5	6-52, 20-41, 28-42, 54-115, 83-99	46,8	7,8	45,4
HD24	5	6-52, 20-41, 28-42, 54-115, 83-99	50,4	3,5	46,1
HD25	5	6-52, 20-41, 28-42, 54-115, 83-99	48,9	2,8	48,2
HD26	5	6-52, 20-41, 28-42, 54-115, 83-99	41,1	11,3	47,5
HD27	5	6-52, 20-41, 28-42, 54-115, 83-99	41,1	7,1	51,8
HD31	5	6-52, 20-41, 28-42, 54-115, 83-99	48,2	2,1	49,6
HD32	5	6-52, 20-41, 28-42, 54-115, 83-99	47,5	5,7	46,8
HD33	5	6-52, 20-41, 28-42, 54-115, 83-99	48,9	2,1	48,9

The deduced proteins of the dimeric α -amylase inhibitors had 10 Cys, and their positions were conserved, which indicated that the Cys were important for these inhibitors' three-dimensional structure (Liu et al., 2012). In agreement with the previous literature, all the haplotypes show five disulphide bonds which are covalent bonds between the sulphur atoms of the Cys residues and which influence the folding and the stability of the three-dimensional protein structure. However, the positions of the five disulphide bonds performed by PredictProtein server are slightly different from previous reports: in fact, according to Oda et al. (1997) Cys28 binds to Cys83 and Cys42 to Cys99.

Unlike all the other haplotypes, HD1 (found in Peliss) lacks one Cys residue at position 41 (substituted by a Serine residue) and consequently forms only four disulphide bonds.

The percentages of the three types of secondary structure were similar among the haplotypes, with a similar content of alpha helix and loop (with the exception of HD1, HD13, HD26, HD27 showing a slight higher percentage of loop) and with a beta strand ranged from 1,4% to 11,3%.

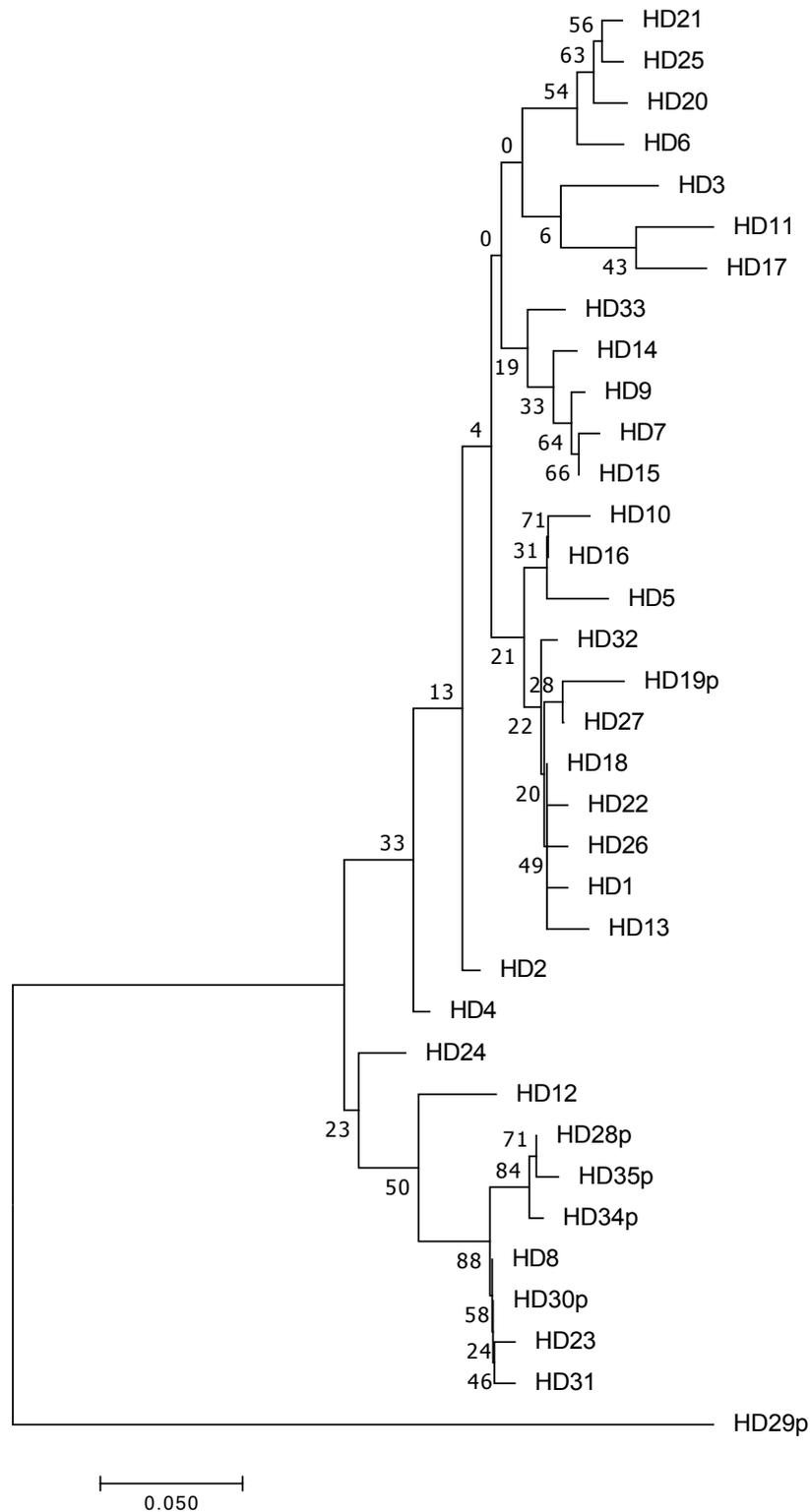
4.2.4 Phylogenetic analysis of dimeric alpha-amylase inhibitors

The Neighbor-Joining method was used to calculate the phylogenetic distances and to construct the phylogenetic tree (Saitou et al., 1987) among all the 35 dimeric α -amylase inhibitor haplotypes (Figure 11 a) and among only the 29 haplotypes after excluding the pseudogenes (Figure 11 b). The optimal trees with the sum of branch length = 0.96668030 (a) and 0.58559645 (b) are shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein et al., 1985).

The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerandl et al., 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 139 (a) and 141 (b) positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

Figure 11. Phylogenetic analysis of the 35 dimeric α -amylase inhibitor haplotypes (a) and of the 29 haplotypes after excluding the pseudogenes (b). In red is indicated in which genotype each haplotype occurred.

a. with pseudogenes



b. without pseudogenes

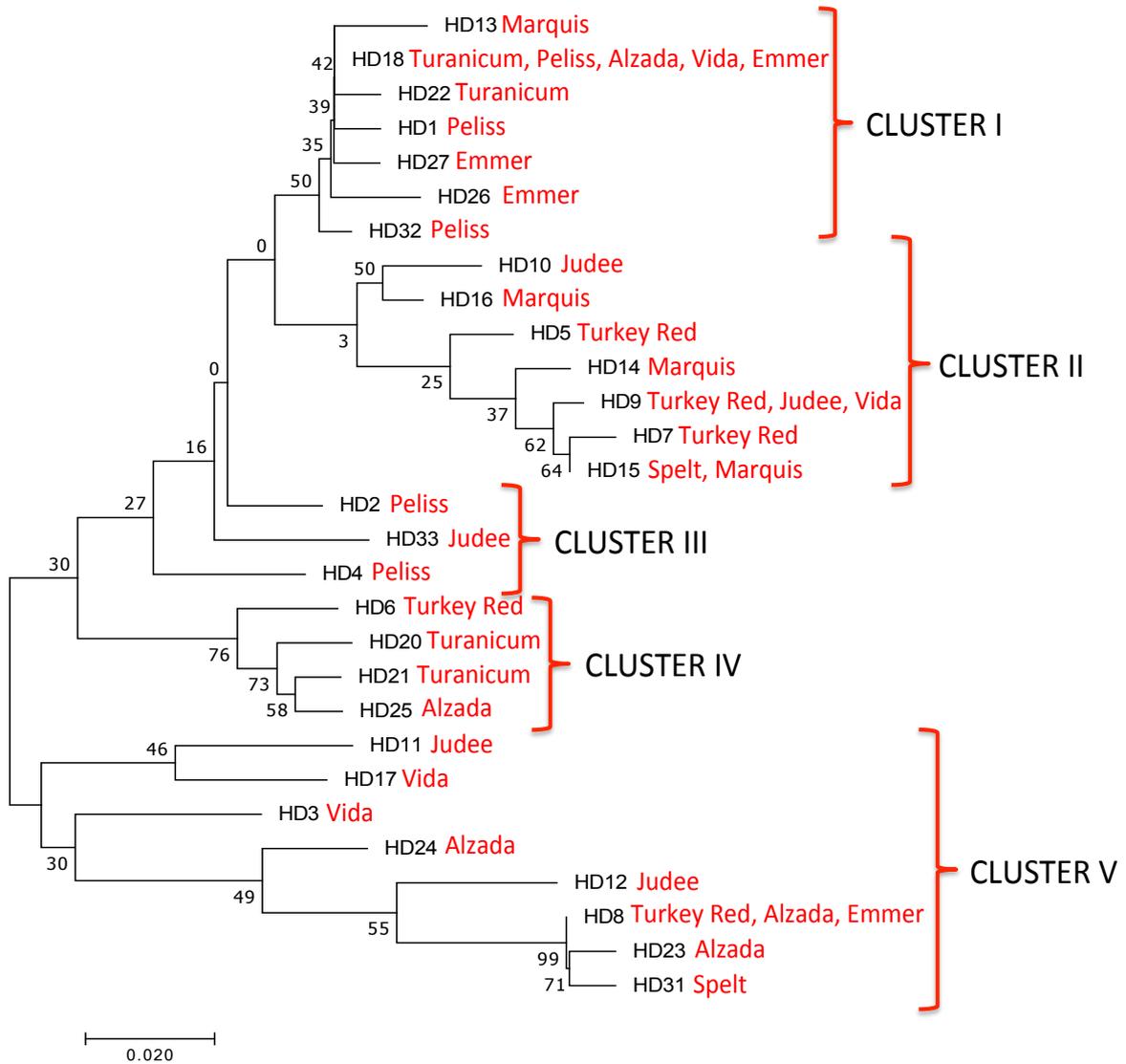


Fig. 11 a shows the results of all the haplotypes, including the pseudogenes. HD29p was the most divergent and was highly separated from the other haplotypes. The other pseudogenes from Einkorn (HD28p, HD34p, HD35p) diverged together in a separated group. Figure 11 b shows the results considering only the putatively functional genes and four clusters can be found.

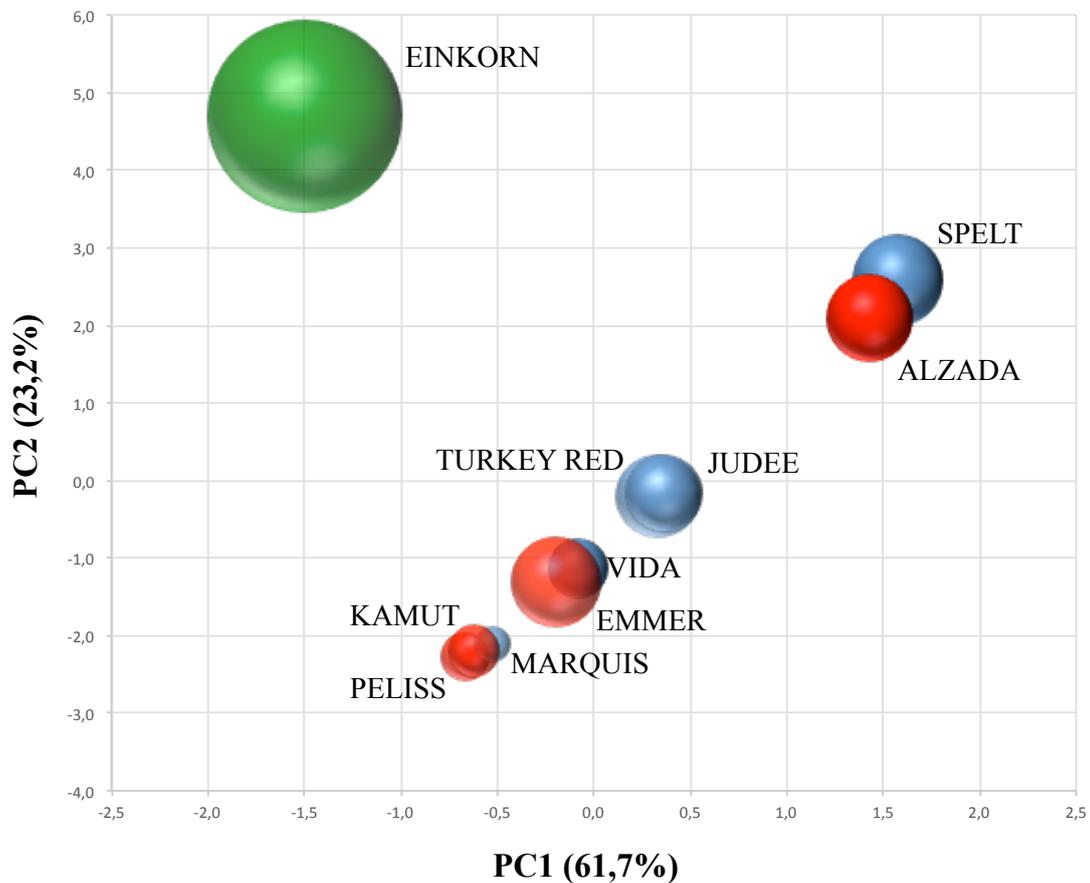
4.2.5 Principal Component Analysis (PCA) of WDAI sequences

To visualise relationships between the tested variables, principal component analysis was run for each species in two steps. First, an interim PCA was run with all the variables

included (69 nsSNPs), and the quality of representation (Cos2) of each variable was checked. Second, a definitive PCA was run with those variables with a quality of representation higher than 0.5 (48 nsSNPs).

The principal component analysis of deduced amino acid sequences of dimeric α -amylase inhibitors explained the 84,9% of the total variance with the first PC accounting for 61,7% and the second PC for 23,2%. Figure 12 shows that Einkorn highly diverges from all the other wheat genotypes. A second group which includes Spelt and Alzada can be identified. All the remaining wheat genotypes can be assigned to a third group.

Fig 12. Principal component analysis based on deduced amino acid sequences of dimeric α -amylase inhibitors. The scatter plot reports the projection of cases (10 wheat genotypes) on the first two components PC1 and PC2 (accounting for 84,9% of total variability): the diameter of each balloon is proportional to the variability observed for the wheat accession. Each colour of the balloons corresponds to a different wheat ploidy: green for diploid wheat; red for tetraploid wheat; blue for hexaploid wheat.



4.3 Tetrameric alpha-amylase inhibitors CM3 (WTAI-CM3)

4.3.1 Sequence analysis of tetrameric alpha-amylase inhibitors CM3

Wheat tetrameric α -amylase inhibitor (WTAI) is a mixture of 4 subunits: one WTAI-CM2 plus two WTAI-CM3 plus one WTAI-CM16, where none of these is active on its own.

WTAI-CM3 has been selected as representative of WTAI genes in this study because of its extended interaction with α -amylases compared to the other CM subunits (Capocchi et al., 2013) and because it was the most abundant expressed WTAI proteins in cv. Butte 86 (Altenbach et al., 2011).

The sequence of CM3 encoding gene reveals that there is only one ORF encoding the entire protein (Singh et al., 2012).

There is little information about the diversity of proteins within the WTAI group because the various subunits of the tetrameric inhibitor have been sequenced from only a few wheat genotypes.

The primers CM3for and CM3rev were designed *in house* and were used to amplify the ORFs of the tetrameric α -amylase gene CM3 and all the samples gave the PCR product of the expected size (647 bp). The only exception was Einkorn which didn't produce any PCR product of the expected size with the primers cited above. In order to amplify WTAI-CM3 ORF from this sample, two sets of internal primers were designed *in house* and were used in all the possible combinations with the Einkorn sample, but no band of the expected size was amplified. In conclusion, using all the primers described above it was not possible to amplify any sequence corresponding to the ORF of WTAI-CM3 genes in the diploid Einkorn sample.

In order to check if WTAI-CM3 sequences of diploid grains are present in the available database, the sequence of inhibitor WTAI-CM3 (NCBI: X17574) was used to perform a blast search against the GenBank non-redundant DNA database (access performed on October 2nd, 2019). It was not possible to retrieve any WTAI-CM3 sequence from *T. monococcum*.

Some studies were previously performed to quantify allergens abundance in different wheat genotypes and raised doubts on the presence of WTAI-CM3 proteins in diploid wheats. Rogniaux et al. (2015) used a targeted MS/MS approach and detected WTAI-CM3 proteins

at high levels in tetraploid and hexaploid grains, but at very low levels in two *T. monococcum* (cv. Engrain, cv. DV92). Geisslitz et al. (2018) used a new targeted LC-MS/MS method and detected WTAI-CM3 in only four out of eight einkorn cultivars, but in very low concentrations near the LOQ or even below.

Regarding the other nine wheat genotypes (Emmer, Peliss, Turanicum, Alzada, Spelt, Turkey Red, Judee, Marquis and Vida), the DNA sequences from five clones per each wheat genotype were determined and used to deduce the full amino acid sequence of the proteins. As described before (Singh et al., 2012) there was no intron in the WTAI-CM3 sequences. In this study, no ins/del in the coding region of WTAI-CM3 was found in any of the wheat samples.

From a total of 45 deduced amino-acid sequences, 25 haplotypes have been obtained, among which one haplotype was identified as a pseudogene due to the presence of one in-frame stop codon (table 13 and 14). Each WTAI-CM3 haplotype is indicated with the code HC followed by a number (Example: HC1); the pseudogene is indicated with a p after the number (HC20p, found in Turkey Red). All the haplotypes were found in only one genotype, with two exceptions: HC6 (shared by Emmer, Turanicum, Peliss, Alzada, Spelt, Turkey Red, Marquis, Vida) and HC7 (shared by Spelt, Marquis, Vida). HC6 occurred in 18 out of 45 WTAI-CM3 sequences and consequently it turned out to be the most abundant WTAI-CM3 haplotype.

All the putatively functional genes encoded for a 168 amino acid protein (25 amino acid signal peptide and 143 amino acid mature protein).

In this study, after aligning the 25 WTAI-CM3 haplotypes from 9 wheat genotypes, 34 nsSNPs were identified. The frequency of nsSNPs was 1 out of 14,8 bases. Previously, Liu et al. (2012) stated that WTAI genes (CM2, CM3, CM16) are more conserved than WMAI and WDAI genes probably because these subunits are combined to form the tetrameric inhibitor and an amino acid change in one of them can affect the structure and consequently the resulting inhibitory activity. However, in this study results showed that WTAI-CM3 was more conserved than WDAI genes where it was found a snSNPs frequency of 1 out of 6,2 bases, but less conserved than WMAI genes where the snSNP frequency was 1 out of 25,2 bases.

Previously, hundreds of WTAI-CM3 sequences were obtained in emmer samples and they found only 5 cSNP in the coding sequence (Liu et al., 2012). Wang et al. (2011) found only 3 snSNPs in 14 populations of wild emmer wheat, but none of these snSNPs was found in this study.

Tab. 13. List of WTAI-CM3 haplotypes.

GENOTYPES	PUTATIVELY FUNCTIONAL GENES	PSEUDOGENES	TOTAL SEQUENCED CLONES
Emmer	5 (HC6, HC6, HC6, HC16, HC17)	/	5
Turanicum	5 (HC6, HC9, HC10, HC11, HC12)	/	5
Peliss	5 (HC6, HC6, HC6, HC6, HC18)	/	5
Alzada	5 (HC6, HC6, HC13, HC14, HC15)	/	5
Spelt	5 (HC6, HC6, HC7, HC22, HC23)	/	5
Turkey Red	4 (HC6, HC6, HC19, HC25)	1 (HC20p)	5
Judee	5 (HC1, HC2, HC3, HC4, HC5)	/	5
Marquis	5 (HC6, HC6, HC7, HC21, HC24)	/	5
Vida	5 (HC6, HC6, HC7, HC8, HC8)	/	5

Tab. 14. Table showing in which genotype each WTAI-CM3 haplotype occurred.

HAPLOTYPES	WHEAT GENOTYPES
HC1	Judee
HC2	Judee
HC3	Judee
HC4	Judee
HC5	Judee
HC6	Peliss, Spelt, Turkey Red , Marquis, Vida, Turanicum, Alzada , Emmer
HC7	Spelt , Marquis , Vida
HC8	Vida
HC9	Turanicum
HC10	Turanicum
HC11	Turanicum
HC12	Turanicum
HC13	Alzada
HC14	Alzada
HC15	Alzada
HC16	Emmer
HC17	Emmer
HC18	Peliss
HC19	Turkey Red
HC20p	Turkey Red
HC21	Marquis
HC22	Spelt
HC23	Spelt
HC24	Marquis
HC25	Turkey Red

The alignment of the 25 WTAI-CM3 haplotypes and the position of the nsSNPs are shown in Figure 13.

Haplotype HC6, which is the most abundant WTAI-CM3 haplotype, has the same sequence as WTAI-CM3 (X17574) studied by Garcia-Maroto et al. (1990), the other haplotypes show a high percentage of identity with WTAI-CM3 (X17574) ranging from 92,85% to 99,4%.

The most part of the deduced proteins of WTAI-CM3 had 10 Cys residues which form five disulphide bonds (see Table 16). The disulphide bonds are essential for the inhibitory activity (Liu et al., 2012) and in fact the Cys residues are at conserved positions. These conserved cysteine motifs are in particular the double Cys-Cys (residues 83-84) followed by a consensus sequence of Cys-Arg-Cys (CRC, residues 94-96), both found also in WMAI and WTAI proteins (Singh et al., 2012). The exception was HC3 (found in Judee) with Tyr instead of the second Cys of the CRC motif. Moreover the following haplotypes showed a change at one of the other Cys residues: HC9 (Tyr at position 148) and HC11 (Arg at position 29) both haplotypes found in Turanicum; HC18 (Tyr at position 52) found in Peliss.

The sequence DLPGCPRE (amino acid positions 101-108, highlighted with a red box in Fig. 14) is the most conserved: it is also conserved in alpha-amylase /trypsin inhibitor of *Eleusine coracana* (RBI) and Hageman factor inhibitor, but it was not found in WMAI and WDAI proteins. It was previously hypothesized that this sequence can be a distinct sequence of alpha amylase/trypsin-bifunctional inhibitors (Singh et al., 2012). Most of the haplotypes share this conserved sequence, with the following exceptions: HC14 (found in Alzada) which shows Pro instead of Leu at amino acid position 102; HC3 and HC4 (both found in Judee), HC7 (found in Spelt, Marquis, Vida), HC8 (found in Vida), HC21 (found in Marquis), HC25 (found in Turkey Red) which show Gln instead of Glu at amino acid position 108.

Capocchi et al. (2013) proposed a structural model for the emmer tetrameric α -amylase inhibitor and studied the interaction between this model and the *T. molitor* α -amylase. In this complex, the first CM3 subunit allows the positioning of the tetrameric inhibitor inside the active site depression of the α -amylase. The residues which are in contact with the active site of α -amylase and which are important for the inhibitory activity are located within the sequences Phe30-Lys55 and Lys116-Gln122 (highlighted with yellow boxes in

Fig. 14) which reach the substrate binding site and target the catalytic residues. Four amino acid changes occurred in some haplotypes at level of the first sequence: Thr instead of Met at amino acid position 40 occurred in HC16 (found in Emmer); Val instead of Ala at amino acid position 43 occurred in HC1 (found in Judee); Phe instead of Tyr at amino acid position 46 and Met instead of Gly at position 49 both occurred in HC3 (found in Judee), HC7 (found in Spelt, Marquis, Vida), HC8 (found in Vida), HC21 (found in Marquis), HC25 (found in Turkey Red). Interestingly, the haplotypes which had the last two amino acid changes also had changes at the conserved sequence DLPGCPRE and all these were found only in hexaploid wheat genotypes.

ATIs proteins from wheat have been demonstrated to trigger inflammation by eliciting strong innate immune effects *in vitro* and *in vivo* with the activation of the TLR4–MD2–CD14 complex (Junker et al., 2012). Cuccioloni et al. (2017) using surface plasmon resonance binding analyses and computational studies investigated the interaction between WTAI-CM3, as representative of ATIs proteins, and human TLR4. They showed the ATIs ability to directly target the receptor and predicted the ATI-TLR4 binding interface regions. These sequences (residues 33-44 and 90-100 in the mature protein sequence) are evidenced with blue boxes in Figure 14. The haplotypes which showed one or two amino acid changes within these regions are: HC1 and HC5 (both found in Judee); HC8 (found in Vida); HC15 (found in Alzada); HC16 (found in Emmer). From the existing evidences it is not possible to say if these changes are able to decrease or delete the ability of CM3-WTAI to bind human TLR4.

Fig. 13. Multiple sequence alignment of 25 WTAI-CM3 haplotypes obtained from different wheat genotypes and WTAI-CM3 (GenBank: X17574). The sequences related to the signal peptide have been included. The dots indicate conserved residues and the letters correspond to the substituted amino acid residues for each alignment gap. Light blue boxes highlight the positions of the 10 Cys residues within the mature protein sequence.

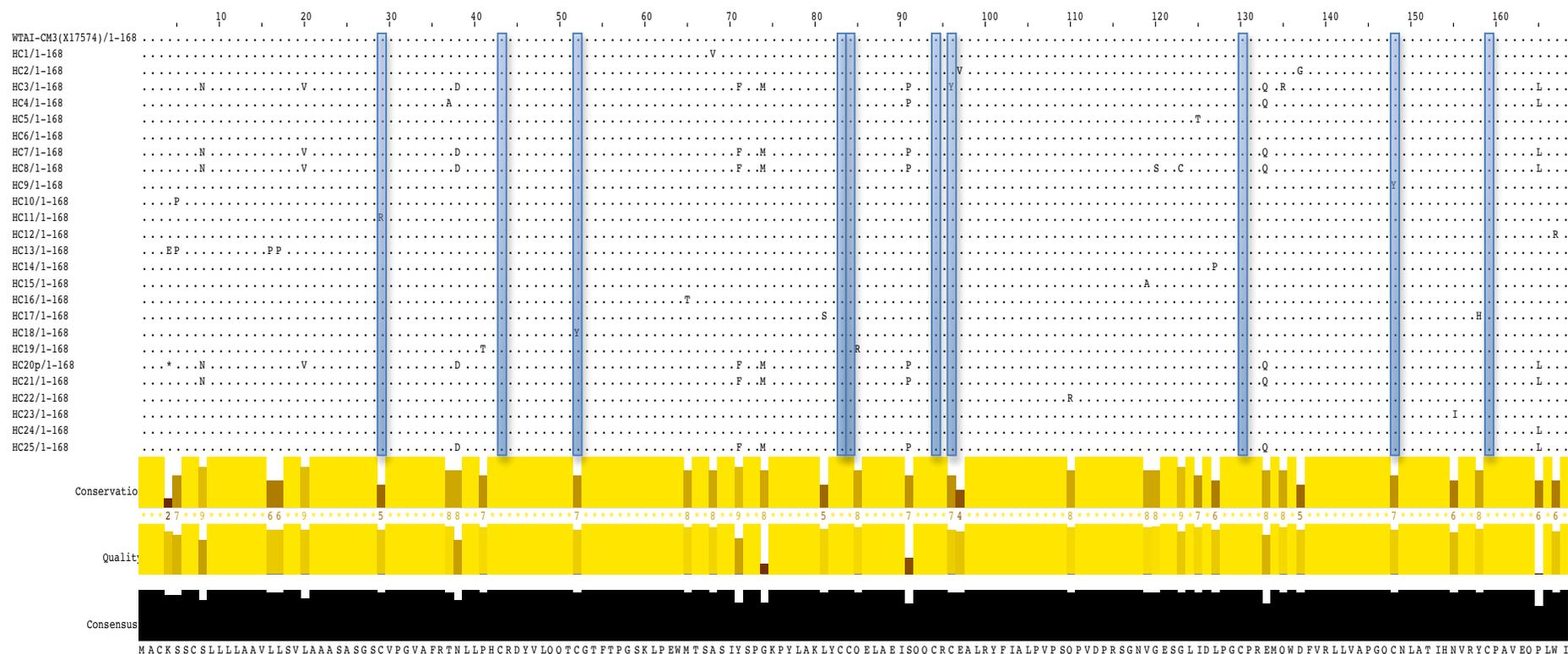
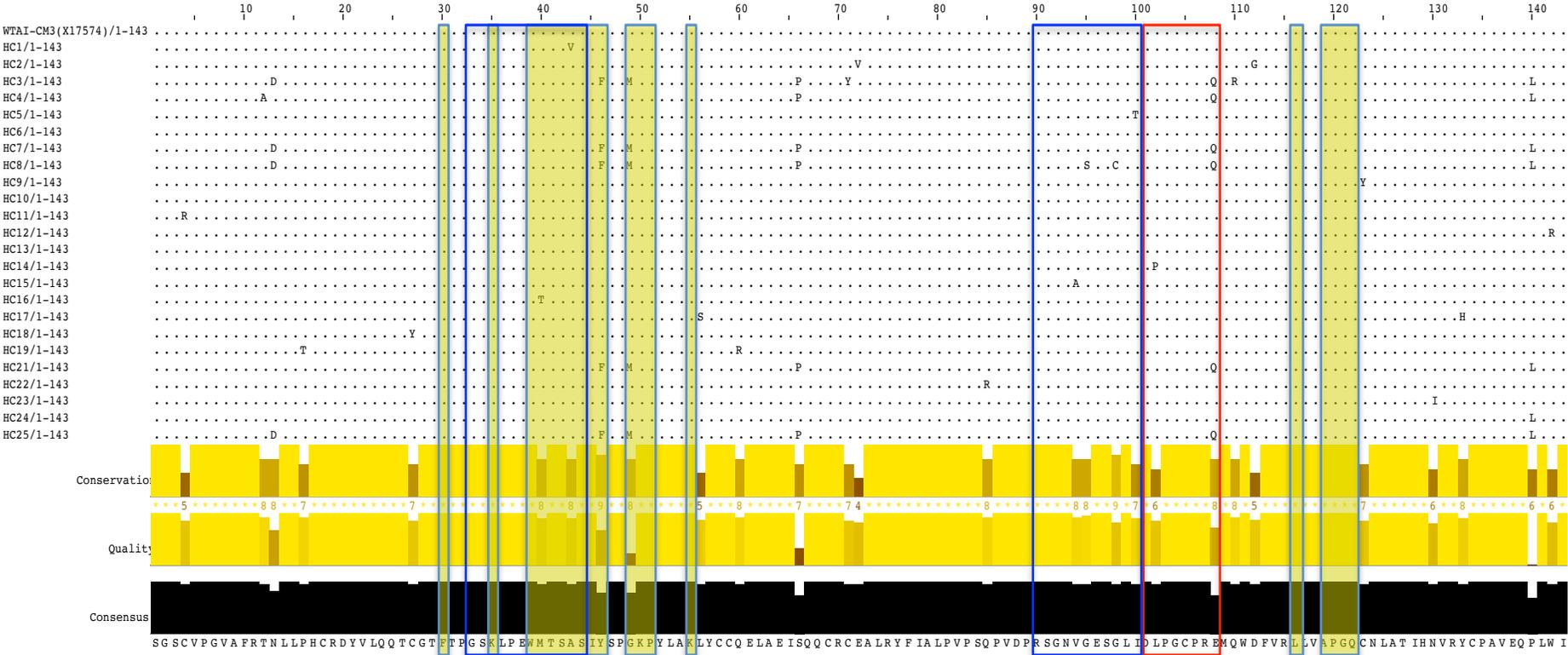


Fig. 14. Multiple sequence alignment of WTAI-CM3 haplotypes with WTAI-CM3 X17574 showing the conserved sequence DLPGCPRE with the red box and the amino acid residues of contact with TMA with yellow boxes. The sequences of the 25 amino acids of the signal peptide are not included in the alignment.



4.3.2 Primary structural analysis of tetrameric alpha-amylase inhibitors CM3

The primary structural analysis of the 24 WTAI-CM3 haplotypes is shown in Table 15. The pseudogene HC20p has been excluded and only the amino acid sequences related to the mature protein have been used for the analysis.

The calculated isoelectric points (pI) ranged from 6.66 to 8.04 (HC2) and this indicates a higher presence of negatively charged residues. All the haplotypes showed a similar extinction coefficient and aliphatic index and showed an instability index higher than 40 which suggests thermal instability. This can be explained by the fact that none of these proteins is active on its own, but they are stable when they form the quaternary structure in combination with the other CM proteins. The GRAVY score is near to zero in all the haplotypes indicating hydrophilicity pattern and good interaction with water.

Tab. 15. Physiochemical properties of the WTAI-CM3 haplotypes.

Haplotype	Molecular weight	p.I.	-R (neg residue)	+R (pos residue)	Extinction coefficient *	Instability index	Aliphatic index	GRAVY
HC1	15860.37	6.66	11	11	26065	47.62	85.17	-0.029
HC2	15744.29	8.04	9	11	26065	46.90	85.87	0.029
HC3	16004.63	7.48	11	12	25940	51.49	86.57	-0.003
HC4	15827.38	7.45	10	11	26065	48.47	87.27	0.003
HC5	15820.26	6.66	11	11	26065	47.62	81.12	-0.083
HC6	15832.31	6.66	11	11	26065	47.62	83.85	-0.046
HC7	15916.54	6.66	11	11	24575	48.54	86.57	0.031
HC8	15992.65	6.65	11	11	24575	54.63	86.57	0.048
HC9	15892.35	6.66	11	11	27430	47.62	83.85	-0.073
HC10	15832.31	6.66	11	11	26065	47.62	83.85	-0.046
HC11	15885.36	7.48	11	12	25940	47.22	83.85	-0.095
HC12	15802.29	7.46	11	12	20565	47.32	83.85	-0.071
HC13	15832.31	6.66	11	11	26065	47.62	83.85	-0.046

Tab. 15. (continued)

Haplotype	Molecular weight	p.I.	-R (neg residue)	+R (pos residue)	Extinction coefficient *	Instability index	Aliphatic index	GRAVY
HC14	15816.27	6.66	11	11	26065	47.62	81.12	-0.084
HC15	15804.26	6.66	11	11	26065	48.22	82.52	-0.063
HC16	15802.23	6.66	11	11	26065	45.12	83.85	-0.064
HC17	15780.20	6.76	11	11	24575	50.09	81.12	-0.092
HC18	15892.35	6.66	11	11	27430	47.03	83.85	-0.073
HC19	15864.36	7.46	11	12	26065	45.46	83.85	-0.047
HC21	15915.55	7.46	10	11	24575	47.49	86.57	0.031
HC22	15860.37	7.46	11	12	26065	47.62	83.85	-0.053
HC23	15831.37	6.66	11	11	26065	48.45	86.57	0.010
HC24	15848.36	6.66	11	11	26065	46.28	86.57	-0.008
HC25	15916.54	6.66	11	11	24575	48.54	86.57	0.031

* = Extinction coefficients are in units of $M^{-1} \text{ cm}^{-1}$, at 280 nm measured in water and assuming all pairs of Cys residues form cystines.

4.3.3 Secondary structural prediction of tetrameric alpha-amylase inhibitors CM3

The secondary structure prediction of each haplotype (after excluding the signal peptide sequence) has been performed using PredictProtein server (<https://www.predictprotein.org/>), which evaluates the tendency of each amino acid to be in one of the three conformational states (alpha helix, beta strand or loop) stabilized by hydrogen bonds. The secondary structural prediction of WTAI-CM3 is shown in Table 16.

Tab. 16. Disulphide bridges and secondary structure analysis for WTAI-CM3 from different wheat genotypes.

Haplotype	N° disulphide bridges	Disulphide bridges	Secondary structure analysis		
			Alpha helix (%)	Beta strand (%)	Loop (%)
HC1	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC2	5	4-69, 18-58, 27-59, 71-134, 105-123	33.57	0	66.43
HC3	4	4-69, 18-105, 27-58, 59-123	34.27	0	65.73
HC4	5	4-69, 18-58, 27-59, 71-134, 105-123	34.27	0	65.73
HC5	5	4-69, 18-58, 27-59, 71-134, 105-123	34.27	0	65.73
HC6	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC7	5	4-69, 18-58, 27-59, 71-134, 105-123	34.27	0	65.73
HC8	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC9	4	4-69, 18-71, 27-58, 59-105	34.97	0	65.03
HC10	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC11	4	27-69, 58-71, 59-134, 105-123	34.27	0	65.73
HC12	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC13	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC14	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC15	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC16	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC17	5	4-69, 18-58, 27-59, 71-134, 105-123	34.27	0	65.73
HC18	4	4-71, 18-105, 58-69, 59-123	34.97	0	65.03
HC19	5	4-69, 18-58, 27-59, 71-134, 105-123	34.27	0	65.73
HC21	5	4-69, 18-58, 27-59, 71-134, 105-123	34.27	0	65.73
HC22	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC23	5	4-69, 18-58, 27-59, 71-134, 105-123	34.97	0	65.03
HC24	5	4-69, 18-58, 27-59, 71-134, 105-123	34.27	0	65.73
HC25	5	4-69, 18-58, 27-59, 71-134, 105-123	34.27	0	65.73

Almost all the deduced proteins of the WTAI-CM3 had 10 Cys, and their positions were conserved, which indicated that the Cys were important for these inhibitors' three-dimensional structure (Liu et al., 2012). In agreement with the previous literature, all the haplotypes show five disulphide bonds which are covalent bonds between the sulphur atoms

of the Cys residues and which influence the folding and the stability of the three-dimensional protein structure.

As shown previously in Fig. 13, the haplotypes HC3 (found in Judee); HC9 and HC11 (both found in Turanicum); HC18 (found in Peliss) lack one Cys residue and consequently form only 4 disulphide bridges.

The percentages of the three types of secondary structure were very similar among the 24 haplotypes: alpha helix ranged from 33.57% to 34.97%, loop from 65.03% to 66.43% and beta strand was not present according to this prediction. These results suggest that these proteins have a higher tendency to exist as random coil.

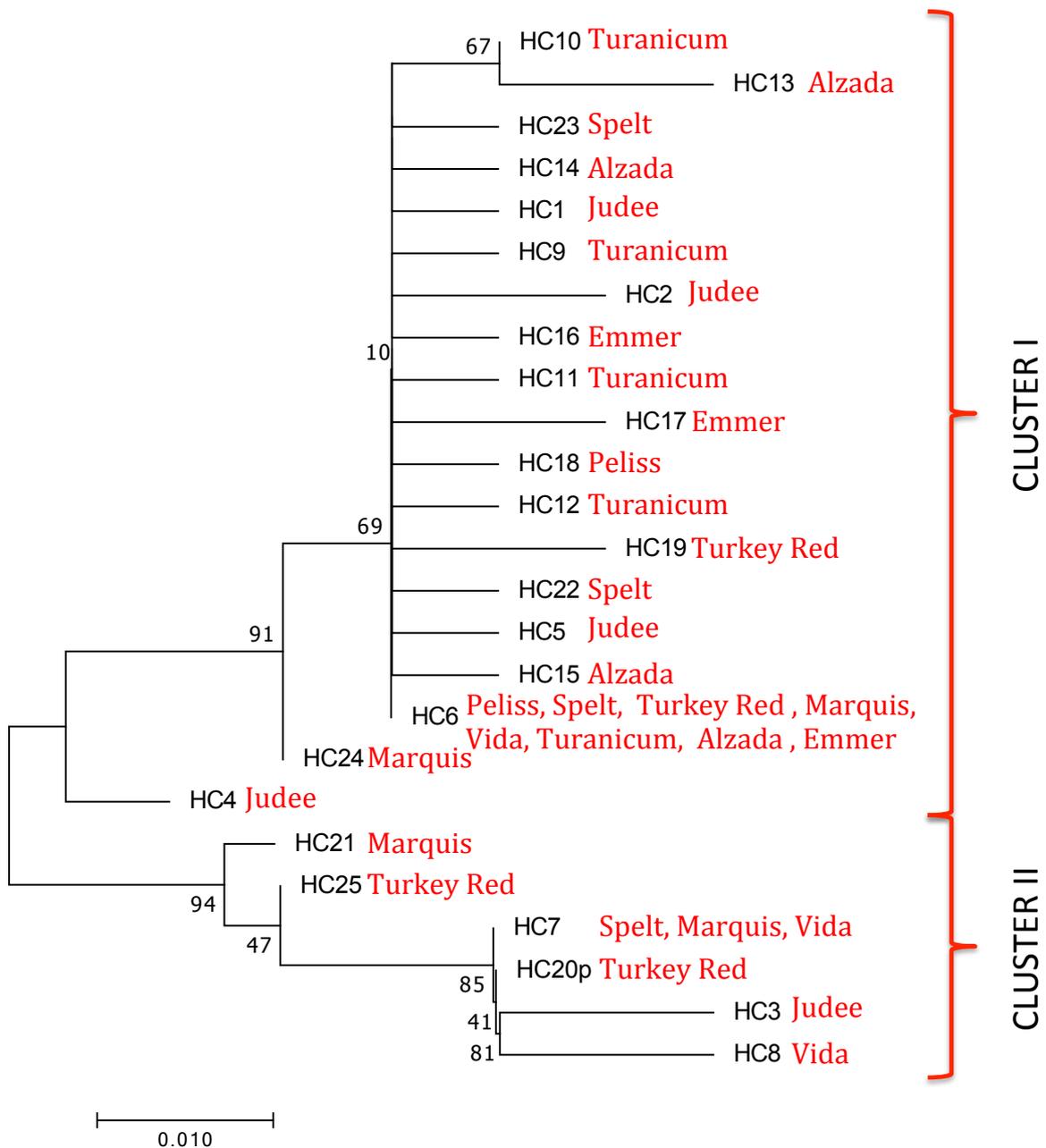
4.3.4 Phylogenetic analysis of tetrameric alpha-amylase inhibitors CM3

The Neighbor-Joining method was used to calculate the phylogenetic distances and to construct the phylogenetic tree (Saitou et al., 1987) among all the 25 WTAI-CM3 haplotypes (Figure 15). The optimal tree with the sum of branch length = 0.20342301 is shown. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein et al., 1985).

The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckermandl et al., 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 167 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 with 1000 bootstrap replicates (Kumar et al., 2016).

Two clusters can be identified with cluster II displaying only haplotypes from hexaploid wheat genotypes.

Fig. 15. Phylogenetic analysis of WTAI-CM3 inhibitors. In red is indicated in which genotype each haplotype occurred.



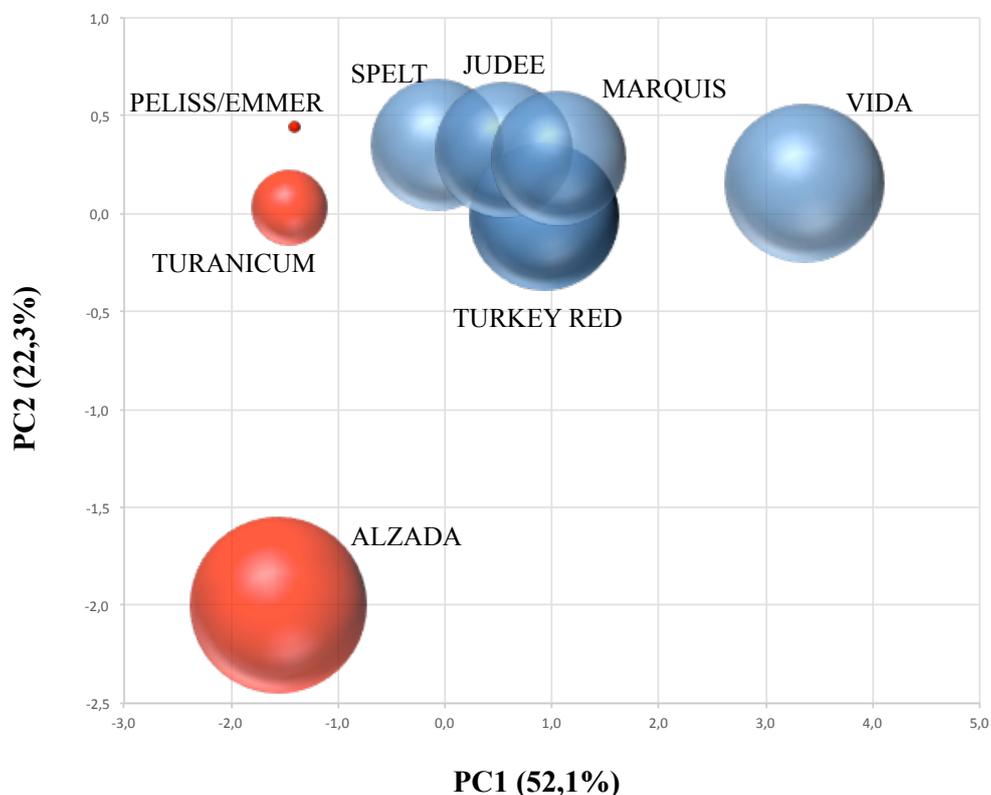
4.3.5 Principal Component Analysis (PCA) of WTAI-CM3 sequences

Figure 16 shows the principal component analysis based on deduced amino acid sequences of WTAI-CM3. To visualise relationships between the tested variables, principal component analysis was run for each species in two steps. First, an interim PCA was run with all the variables included (34 nsSNPs), and the quality of representation (Cos2) of each variable was checked. Second, a definitive PCA was run with those variables with a quality of representation higher than 0.5 (15 nsSNPs).

The scatter plot reports the projection of cases (9 wheat genotypes) on the first two components PC1 and PC2 and explained the 74.4% of total variance with the first PC accounting for 52.1% and the second PC for 22.3%. The diameter of each balloon is proportional to the variability observed for the wheat accession. The samples can be divided in two groups according to the projection of the first component: a first group includes all the tetraploid grains (with Alzada differing among the others on the projection of the second component) and a second group includes the four common wheat genotypes. Spelt is in the middle of the two groups.

Overall, the PCA analysis of WTAI-CM3 inhibitors is in line with the ploidy level of the wheat samples analysed. The diploid Einkorn is not included in the analyses because it was not possible to amplify any sequence related to WMAI genes.

Fig. 16. Principal component analysis based on deduced amino acid sequences of WTAI-CM3. The scatter plot reports the projection of cases (9 wheat genotypes) on the first two components PC1 and PC2 (accounting for 74,4% of total variability): the diameter of each balloon is proportional to the variability observed for the wheat genotype. Each colour of the balloons corresponds to a different wheat ploidy: red for tetraploid wheat; blue for hexaploid wheat.



4.4 Trypsin inhibitors CMx

4.4.1 Sequence analysis of trypsin inhibitors CMx

The trypsin inhibitor from barley, named BTI-CMe, is the best characterised among the trypsin inhibitors of the *Triticeae* family. Homologous proteins have been identified in maize (MTI) and ragi (RBI) as well as rye (Salcedo et al., 2004). The cDNA (pCMx) of the wheat homologue has been first characterized by Sánchez de la Hoz et al. (1994). Moreover, Altenbach et al. (2011) identified the expressed forms of four putative trypsin inhibitors from flour extract of the US wheat cv. Butte 86. Trypsin inhibitors are encoded by genes on the group 4 chromosomes and belong to the monomeric type (Salcedo et al., 2004).

The primers CMXfor and CMXrev were designed *in house* and were used to amplify the ORFs of the CMx genes and all the ten wheat genotypes gave the PCR product of the expected size (498 bp). The DNA sequences from five clones per each wheat genotype were determined and used to deduce the full amino-acid sequence of the proteins. Since CMXfor was designed 9 nt downstream the A of the start codon, these first 9 nucleotides were taken from the reference sequence (GenBank: X75608) and added to each DNA sequence in order to deduce complete protein sequences.

From a total of 50 deduced amino-acid sequences, 22 haplotypes have been obtained, among which two haplotypes were identified as pseudogenes due to the presence of one in-frame stop codon at amino acid residue 61 in HX21p (as a consequence of a single nucleotide substitution) and at amino acid residue 89 in HX22p (as a consequence of a single nucleotide deletion) (table 17 and 18).

Each CMx haplotype is indicated with the code HX followed by a number (Example: HX1); the two pseudogenes are indicated with a p after the number (HX21p found in Einkorn and HX22p found in Marquis). All the haplotypes were found in only one genotype, with three exceptions: HX8 (shared by Spelt, Turkey Red, Marquis, Vida), HX11 (shared by Emmer and Spelt) and HX20 (shared by Turanicum, Peliss, Alzada, Judee, Marquis). HX20 occurred in 15 out of 50 CMx sequences and consequently it turned out to be the most abundant CMx haplotype.

The deduced proteins from haplotypes HX1-HX10 were 145 amino acids long and were found only in hexaploid wheat genotypes and the deduced proteins from haplotypes HX13-HX20 were 146 amino acids long. Differently, two nucleotide sequences showed a

premature stop codon after nucleotide 363 as observed by Sanchez de la Hoz et al. (1994), which resulted in a deduced protein of 121 amino acids (haplotypes HX11 found in Spelt and Emmer and HX12 found in Emmer). Interestingly, Emmer wheat showed only the two haplotypes with the deduced protein of 121 amino acids. The homology between the deduced HX11 and HX12 amino acid sequences and those of the other CMx haplotypes was maintained beyond the premature stop codon, up to the second stop codon which appeared in the same position as the others haplotypes. This premature stop codon found in HX11 and HX12 is not present in the trypsin inhibitor from barley (BTI-CMe) and was not found by Altenbach et al. (2011) in the coding sequences of the four putative trypsin inhibitors expressed in grain from the US wheat Butte 86.

Tab. 17. List of CMx haplotypes.

SAMPLES	PUTATIVELY FUNCTIONAL GENES	PSEUDOGENES	TOTAL SEQUENCED CLONES
Einkorn	4 (HX13, HX13, HX13, HX14)	1 (HX21p)	5
Emmer	5 (HX11, HX11, HX11, HX11, HX12)	/	5
Turanicum	5 (HX17, HX18, HX20, HX20, HX20)	/	5
Peliss	5 (HX20, HX20, HX20, HX20, HX20)	/	5
Alzada	5 (HX19, HX20, HX20, HX20, HX20)	/	5
Spelt	5 (HX3, HX8, HX11, HX11, HX11)	/	5
Turkey Red	5 (HX1, HX5, HX8, HX8, HX9)	/	5
Judee	5 (HX6, HX15, HX16, HX20, HX20)	/	5
Marquis	4 (HX7, HX8, HX10, HX20)	1 (HX22p)	5
Vida	5 (HX2, HX4, HX8, HX8, HX8)	/	5

Tab. 18. Table showing in which genotype each CMx haplotype occurred.

HAPLOTYPES	WHEAT GENOTYPES
HX1	Turkey Red
HX2	Vida
HX3	Spelt
HX4	Vida
HX5	Turkey Red
HX6	Judee
HX7	Marquis
HX8	Spelt, Turkey Red, Marquis, Vida
HX9	Turkey Red
HX10	Marquis
HX11	Emmer, Spelt
HX12	Emmer
HX13	Einkorn
HX14	Einkorn
HX15	Judee
HX16	Judee
HX17	Turanicum
HX18	Turanicum
HX19	Alzada
HX20	Turanicum, Peliss, Alzada, Judee, Marquis
HX21p	Einkorn
HX22p	Marquis

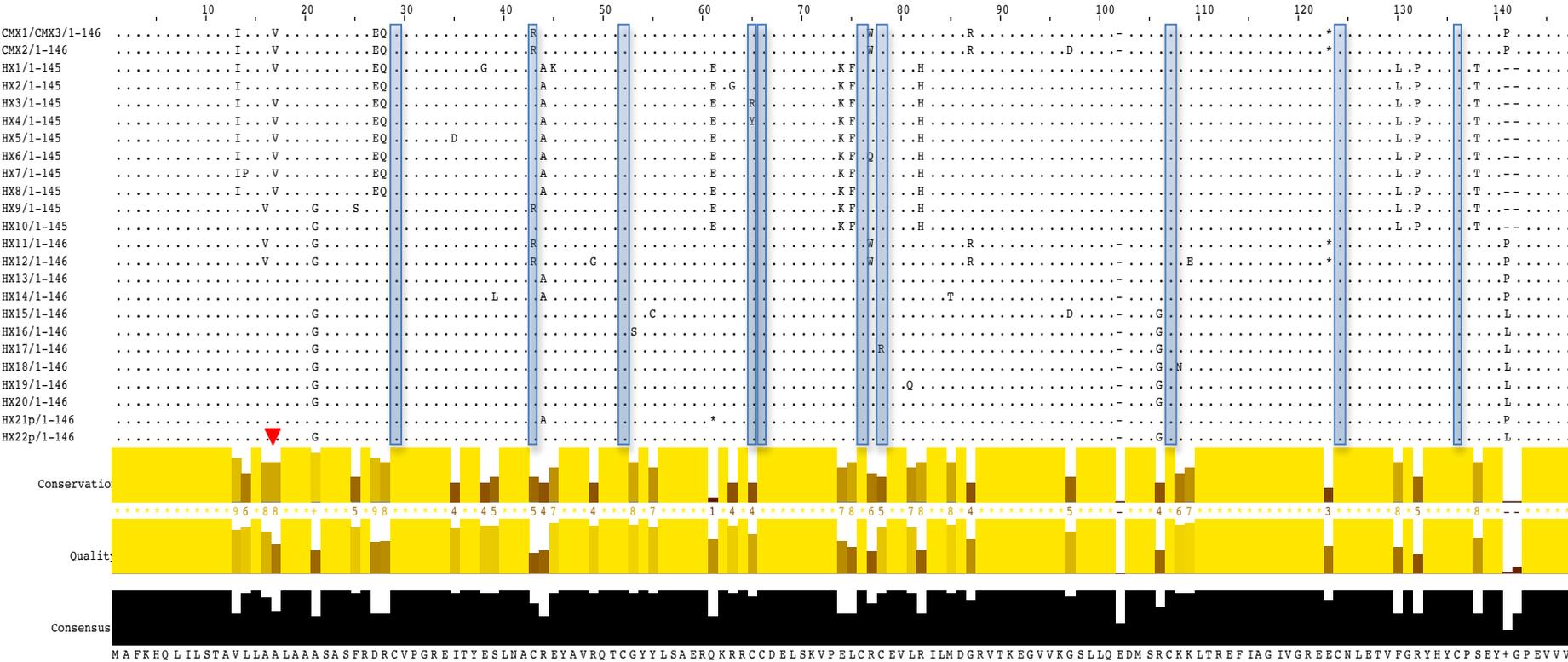
In this study, after aligning the 22 CMx haplotypes from 10 wheat genotypes, 39 nsSNPs were identified. The frequency of nsSNPs was 1 out of 11,2 bases, similar to what has been found for CM3 genes in this study.

The alignment of the 22 CMx haplotypes and the position of the nsSNPs are shown in Figure 17. The pseudogenes were also considered in this analysis and the genetic sequences used to deduce the amino acid sequences for the alignment were selected as follows: in case the internal stop codon was created by a substitution of a single nucleotide, then an asterisk was added in place of the stop codon and the sequence until the end of the ORF of the corresponding functional gene was used; in case the internal stop codon was created by a deletion or insertion of a single nucleotide, then the sequence without this nucleotide change was used.

The most part of the CMx deduced proteins had 10 Cys residues which form up to five disulphide bonds (see Table 20). The disulphide bonds are essential for the inhibitory activity (Liu et al., 2012) and in fact the Cys residues are at conserved positions and show an arrangement similar to the other alpha-amylase/protease inhibitors.

The exceptions were HX9 (found in Turkey Red), HX11 (found in Emmer and Spelt) and HX12 (found in Emmer) with Arg at position 43; HX3 (found in Spelt) with Arg at position 65; HX4 (found in Vida) with Tyr at position 65; HX17 (found in Turanicum) with Arg at position 78. Interestingly, all the haplotypes found in Emmer in this study doesn't show all the 10 Cys residues.

Fig. 17. Multiple sequence alignment of 22 CMx haplotypes obtained from different wheat genotypes with CMX1/CMX3 (deduced from the sequence GenBank: X75608) and with CMX2 (deduced from the sequence GenBank: X75609). The sequences related to the signal peptide have been included. The asterisks * show in-frame stop codons of pseudogenes. The red triangle shows the positions of the deletion found in the nucleotide sequence used to deduce the amino-acid sequence. The dots indicate conserved residues and the letters correspond to the substituted amino acid residues for each alignment gap. Light blue boxes highlight the positions of the 10 Cys residues.



4.4.2 Primary structural analysis of trypsin inhibitors CMx

The primary structural analysis of the 20 CMx haplotypes is shown in Table 19. The pseudogenes HX21p and HX22p have been excluded and only the amino acid sequences related to the mature protein have been used for the analysis.

Tab. 19. Physiochemical properties of the CMx haplotypes.

Haplotype	Molecular weight	p.I.	-R (neg residue)	+R (pos residue)	Extinction coefficient *	Instability index	Aliphatic index	GRAVY
HX1	13849.06	8.07	17	19	11055	58.65	78.02	-0.334
HX2	13822.93	5.77	19	17	11055	52.19	78.02	-0.322
HX3	13975.11	6.89	19	19	10930	65.05	78.02	-0.414
HX4	13982.10	6.20	19	18	12420	59.69	78.02	-0.388
HX5	13923.99	5.78	20	18	11055	59.07	74.79	-0.422
HX6	13894.01	5.77	19	17	11055	59.07	78.02	-0.348
HX7	13922.06	6.19	19	18	11055	60.32	78.02	-0.356
HX8	13922.06	6.19	19	18	11055	60.32	78.02	-0.356
HX9	13770.89	7.66	19	20	10930	60.66	78.49	-0.468
HX10	14021.20	7.64	19	20	11055	58.72	77.19	-0.417
HX11	11422.31	9.25	14	21	11835	58.24	80.31	-0.541
HX12	11324.12	8.76	15	19	11835	56.26	80.31	-0.495
HX13	14025.19	8.08	18	20	11055	58.69	77.38	-0.381
HX14	14021.19	8.08	18	20	11055	57.11	80.57	-0.365
HX15	14025.21	7.60	19	20	9565	58.94	79.75	-0.349
HX16	14057.24	8.08	18	20	11055	56.79	79.75	-0.358
HX17	14080.26	8.40	18	21	10930	60.79	79.75	-0.412
HX18	14013.14	7.63	18	19	11055	58.04	79.75	-0.352
HX19	14042.18	8.08	18	20	11055	54.52	76.56	-0.415
HX20	14027.21	8.08	18	20	11055	56.10	79.75	-0.355

* = Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water and assuming all pairs of Cys residues form cystines.

The predicted molecular weights of the mature CMx proteins are similar to WMAI, WDAI, and WTAI subunits CM1, CM2, CM16 and CM17 determined in *T. aestivum* cv. Butte 86 by Altenbach et al. (2011). The molecular weights ranged from 13770,89 to 14080,26. HX11 and HX12 showed the lowest molecular weights (11422,31 and 11324,12 respectively) because they have the shortest amino acid sequence due to the presence of a premature stop codon, as described previously. The calculated isoelectric points (pI) ranged from 5.77 to 9.25. HX11 and HX12 showed the highest pI due to a lower presence of negatively charged residues. All the haplotypes showed a similar extinction coefficient and aliphatic index and showed an instability index higher than 40 which suggests thermal instability and a short *in vivo* half-life (Gasteiger et al., 2005). The GRAVY score on average is lower than WMAI, WDAI and WTAI-CM3 proteins examined in this study and this indicates higher hydrophilicity of CMx proteins.

4.4.3 Secondary structural prediction of trypsin inhibitors CMx

The secondary structure prediction of each haplotype (after excluding the signal peptide sequence) has been performed using PredictProtein server (<https://www.predictprotein.org/>), which evaluates the tendency of each amino acid to be in one of the three conformational states (alpha helix, beta strand or loop) stabilized by hydrogen bonds. The secondary structural prediction of CMx proteins is shown in Table 20.

Almost all the CMx deduced proteins had 10 Cys, and show five disulphide bonds which are covalent bonds between the sulphur atoms of the Cys residues and which influence the folding and the stability of the three-dimensional protein structure. As shown previously in Fig. 17, the haplotypes HX3 (found in Spelt); HX4 (found in Vida); HX9 (found in Turkey Red) and HX17 (found in Turanicum) lack one Cys residue and consequently form only 4 disulphide bridges. Differently, HX15 (found in Judee) had all the 10 Cys residues, but ProteinPredict server predicted only 3 disulphide bridges. HX11 and HX12 showed only 3 disulphide bridges because of their shorter amino acid sequence and because they lack a Cys residue at position 19 of the mature protein.

Regarding the percentages of the three types of secondary structure, the 20 CMx haplotypes can be divided in three groups: a first group (HX1-HX10) characterized by the absence of beta strand and with a ratio of alpha helix/loop of about 2:3; a second group (HX13-HX20)

characterized by a low percentage of beta strand and with a ratio of alpha helix/loop similar to what seen in the previous group; a third group (HX11 and HX12) characterized by the absence of beta strand and with a similar content of alpha helix and loop.

Tab. 20. Disulphide bridges and secondary structure analysis for CMx proteins from different wheat genotypes.

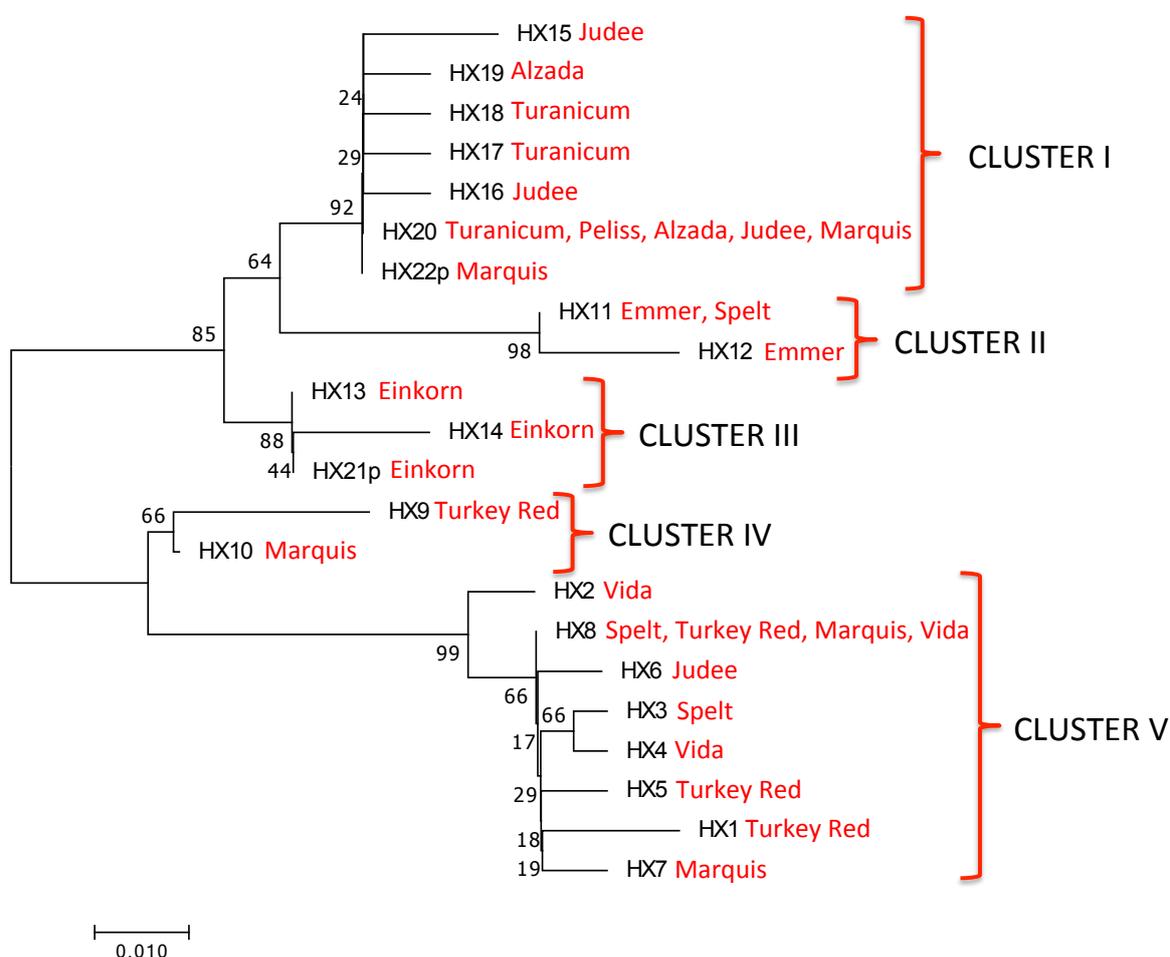
Haplotype	N° disulphide bridges	Disulphide bridges	Secondary structure analysis		
			Alpha helix (%)	Beta strand (%)	Loop (%)
HX1	5	5-52, 19-112, 28-100, 41-42, 54-83	40.50	0	59.50
HX2	5	5-52, 19-112, 28-100, 41-42, 54-83	41.32	0	58.68
HX3	4	5-54, 19-83, 28-52, 42-100	39.67	0	60.33
HX4	4	5-54, 19-83, 28-42, 100-112	39.67	0	60.33
HX5	5	5-52, 19-41, 28-42, 54-112, 83-100	41.32	0	58.68
HX6	5	5-52, 19-41, 28-42, 54-112, 83-100	39.67	0	60.33
HX7	5	5-52, 19-112, 28-100, 41-42, 54-83	40.50	0	59.50
HX8	5	5-52, 19-112, 28-100, 41-42, 54-83	40.50	0	59.50
HX9	4	3-52, 26-40, 50-98, 81-110	40.34	0	59.66
HX10	5	5-52, 19-112, 28-100, 41-42, 54-83	39.67	0	60.33
HX11	3	5-42, 28-52, 54-82	49.48	0	50.52
HX12	3	5-42, 28-52, 54-82	50.52	0	49.48
HX13	5	5-52, 19-111, 28-99, 41-42, 54-82	40.16	4.10	55.74
HX14	5	5-52, 19-111, 28-99, 41-42, 54-82	39.34	4.92	55.74
HX15	3	5-54, 19-42, 28-99	40.98	3.28	55.74
HX16	5	5-52, 19-111, 28-99, 41-42, 54-82	39.34	1.64	59.02
HX17	4	5-52, 19-82, 28-99, 42-111	39.34	3.28	57.38
HX18	5	5-52, 19-41, 28-42, 54-111, 82-99	40.16	1.64	58.20
HX19	5	5-52, 19-111, 28-99, 41-42, 54-82	40.16	1.64	58.20
HX20	5	5-52, 19-41, 28-42, 54-111, 82-99	39.34	1.64	59.02

4.4.4 Phylogenetic analysis of trypsin inhibitors CMx

The Neighbor-Joining method was used to calculate the phylogenetic distances and to construct the phylogenetic tree (Saitou et al., 1987) among all the 22 CMX haplotypes (Figure 18). The optimal tree with the sum of branch length = 0.27715519 is shown. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein et al., 1985).

The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkanndl et al., 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 142 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 with 1000 bootstrap replicates (Kumar et al., 2016).

Fig. 18. Phylogenetic analysis of CMX inhibitors. In red is indicated in which genotype each haplotype occurred.



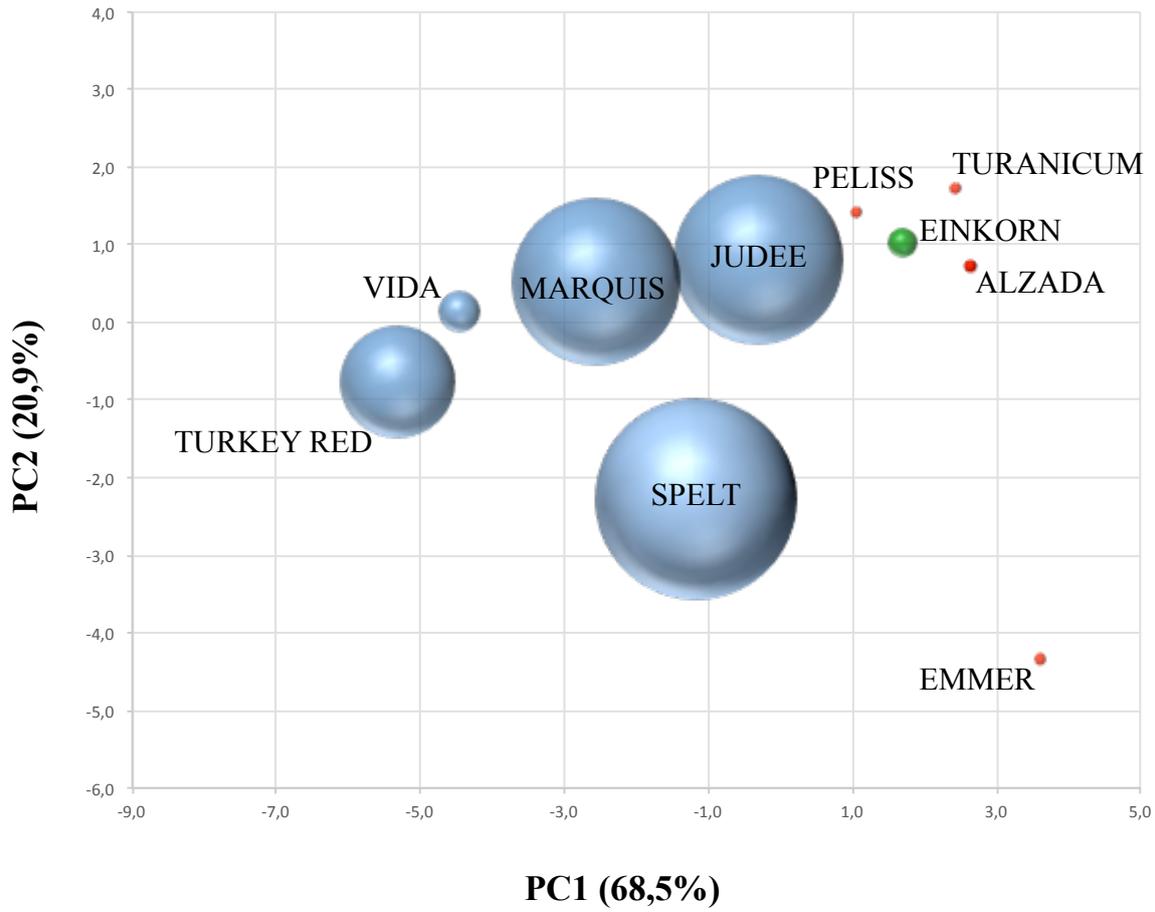
As shown in Figure 18, a first subdivision is between haplotypes HX1-HX10 (which are all the 145 amino acids long CMx proteins and which were found only in hexaploid wheat genotypes) and HX11-HX22p. Interestingly, this subdivision follows what seen for the secondary structure analysis. Then a further subdivision distinguishes HX9 and HX10 (cluster IV) from the other haplotypes of the first group (cluster V), while in the second group three clusters can be identified: cluster I with haplotypes found in both tetraploid and hexaploid genotypes, cluster II showing the two 121 amino acids long CMx proteins, cluster III showing only haplotypes found in Einkorn.

4.4.5 Principal Component Analysis (PCA) of CMx sequences

Figure 19 shows the principal component analysis based on deduced amino acid sequences of CMx. To visualise relationships between the tested variables, principal component analysis was run for each species in two steps. First, an interim PCA was run with all the variables included (39 nsSNPs), and the quality of representation (Cos2) of each variable was checked. Second, a definitive PCA was run with those variables with a quality of representation higher than 0.5 (22 nsSNPs).

The scatter plot reports the projection of cases (10 wheat genotypes) on the first two components PC1 and PC2 and explained the 89,4% of total variance with the first PC accounting for 68,5% and the second PC for 20,9%. The diameter of each balloon is proportional to the variability observed for the wheat accession. The samples can be divided in two groups according to the projection of the first component: a first group includes all the diploid and tetraploid grains (with Emmer differing among the others on the projection of the second component) and a second group includes the hexaploid wheat genotypes.

Fig. 19. Principal component analysis based on deduced CMx amino acid sequences. The scatter plot reports the projection of cases (10 wheat genotypes) on the first two components PC1 and PC2 (accounting for 89,4% of total variability): the diameter of each balloon is proportional to the variability observed for the wheat genotype. Each colour of the balloons corresponds to a different wheat ploidy: green for diploid wheat; red for tetraploid wheat; blue for hexaploid wheat.



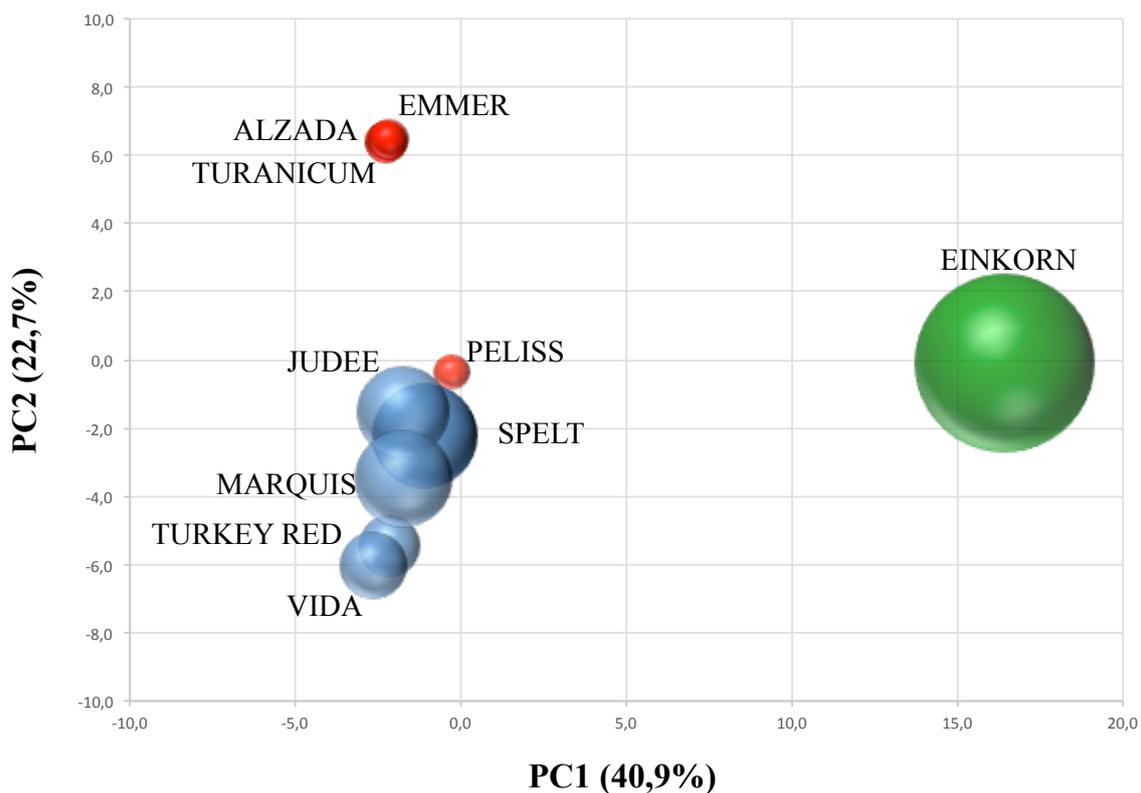
4.5 Principal Component Analysis (PCA) of WMAI, WDAI, WTAI-CM3 and CMx sequences

A further PCA analysis has been performed based on deduced amino acid sequences of all the four genes sequenced in this study (WMAI, WDAI, WTAI-CM3 and CMx). To visualise relationships between the tested variables, principal component analysis was run for each species in two steps. First, an interim PCA was run with all the variables included (160 nsSNPs), and the quality of representation (Cos2) of each variable was checked. Second, a definitive PCA was run with those variables with a quality of representation higher than 0.5 (103 nsSNPs).

The scatter plot reports the projection of cases (10 wheat genotypes) on the first two components PC1 and PC2 and explained the 63,6% of total variance with the first PC accounting for 40,9% and the second PC for 22,7%. The diameter of each balloon is proportional to the variability observed for the wheat accession. Einkorn diverges from all

the other genotypes according to the projection of the first component. The tetraploid wheat can be divided from the hexaploid wheat genotypes according to the second projection with the tetraploid Peliss which is closer to the hexaploid group. This fact can be explained in part by the fact that Peliss was the only tetraploid wheat which was amplified with primers for WMAI genes.

Fig. 20. Principal component analysis based on deduced WMAI, WDAI, WTAI-CM3 and CMx amino acid sequences. The scatter plot reports the projection of cases (10 wheat genotypes) on the first two components PC1 and PC2 (accounting for 63,6% of total variability): the diameter of each balloon is proportional to the variability observed for the wheat genotype. Each colour of the balloons corresponds to a different wheat ploidy: green for diploid wheat; red for tetraploid wheat; blue for hexaploid wheat.



Since genetically divergent wheat accessions have been used in the analysis which could have flattened the results, we decide to perform further analyses using the sequences of all the four ATIs genes, but taking the tetraploid and the hexaploid genotypes separately.

All the hexaploid wheat accessions showed high diversity among the five genetic sequences obtained in each genotype (see Table 21). This strong data dispersion affected the final result of the PCA analysis (data not shown).

Tab. 21. Canonical coordinates of the five genetic replicates for each hexaploid genotypes considering the deduced amino acid sequences from all the four ATIs genes sequenced.

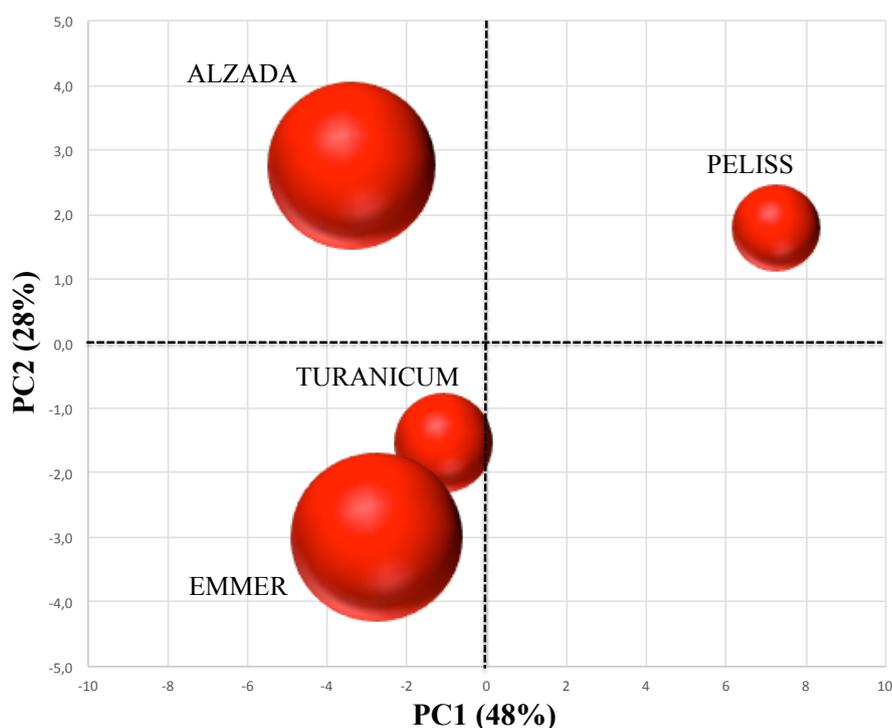
Genotype	PC1	PC2
Turkey Red	-4.18172	2.4431
Turkey Red	-4.30383	1.8560
Turkey Red	-4.18172	2.4431
Turkey Red	-3.99539	-11.3919
Turkey Red	-0.19391	0.8606
Judee	-4.18172	2.4431
Judee	7.78882	3.0921
Judee	8.60309	-4.2548
Judee	8.73933	-5.0041
Judee	7.78882	3.0921
Marquis	-4.18172	2.4431
Marquis	-4.18172	2.4431
Marquis	-0.24303	0.1813
Marquis	8.03387	1.1950
Marquis	7.89904	2.7852
Vida	-3.44576	-0.5921
Vida	-4.18172	2.4431
Vida	-3.37157	-5.6143
Vida	-4.04352	-0.1384
Vida	-4.16563	-0.7255

Differently, the tetraploid wheat accessions were characterized by a higher homology of the genetic replicates in each accession and the principal component analysis based on deduced amino acid sequences of all the four ATIs genes sequenced was performed (Figure 21). The scatter plot reports the projection of cases (4 wheat genotypes) on the first two components PC1 and PC2 and explained the 76% of total variance with the first PC accounting for 48%

and the second PC for 28%. The diameter of each balloon is proportional to the variability observed for the wheat accession. From this analysis, Turanicum and Emmer are close in the South-West quadrant. Alzada and Peliss diverge from Turanicum and Emmer and are placed in the North-West and North-East quadrants respectively. Peliss, which is a heritage durum genotype, diverges from the others for the projection of the first component. This could be in part explained by the fact that Peliss was the only tetraploid wheat genotype which showed WMAI sequences in this study.

Overall, we observed a higher variability intra genotype in hexaploid compared to tetraploid genotypes. The fact that hexaploid wheat genotypes have a further chromosomal set compared to the tetraploid ones can be one of the reasons. However, additional studies considering more genotypes and more genetic sequences per genotype are needed to confirm the present results.

Fig. 21. Principal component analysis based on deduced WMAI, WDAI, WTAI-CM3 and CMx amino acid sequences. The scatter plot reports the projection of cases (4 tetraploid wheat genotypes) on the first two components PC1 and PC2 (accounting for 76% of total variability): the diameter of each balloon is proportional to the variability observed for the wheat genotype.



4.6 Agronomic trial

The agronomic trial was carried out using a collection of ten selected samples belonging to different *Triticace* species (listed in paragraph 3.1) and conducted under organic farming management for three consecutive cropping years at the experimental farm Podere Santa Croce (Argelato, BO) of the Department of Agricultural and Food Science, University of Bologna (Italy). The three cropping years were 2016-2017; 2017-2018; 2018-2019. The following agronomic parameters were recorded: yield, percentage of lodging, weed incidence, disease incidence, disease severity.

Moreover, the ten selected genotypes were also grown at the Quinn Organic Farm located near Big Sandy, Montana, USA under organic farming management during three consecutive cropping years (2015-2016; 2016-2017; 2017-2018).

The meteorological data (temperature and rain) were recorded at both sites and are shown in Table 22.

As described in Table 23, considering the genotype, the yield and the percentage of lodging were significant. Common wheat genotypes showed the highest yields and the hulled wheat genotypes (Spelt, Emmer and Einkorn) the lowest ones. The data confirmed that not-dwarf genotypes and with higher yields are more subjected to lodging. Among the ancient genotypes not subjected to lodging there are Emmer and Einkorn which in fact showed the lowest yields. Considering the cropping year, the weed incidence and the disease incidence were significant. In the third cropping year the weed and disease incidence were higher than the previous two years because there was much rainfall during the spring.

At the American location it was not possible to harvest Peliss wheat in 2016 and 2017; moreover, Emmer harvested in USA in 2016 was not included in the analyses because badly stored.

All the wheat genotypes grown in the two areas during the three cropping years were tested for their moisture and protein contents and results are shown in Table 24. The protein content of samples grown in USA was higher than that of the same samples cultivated in Italy. This is in agreement with the fact that compared to Italy, in Montana the averaged wheat yields are typically lower and the protein quality is higher (Pastaria International 6/2015).

Tab. 22. Meteorological data recorded during the three cropping years in USA (a) and in Italy (b).

a. USA

		Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug
2015-2016 cropping year	Precipitation (mm)	40,1	25,7	12,7	7,1	10,7	10,7	23,6	99,8	104,4	43,2	71,1	38,4
	n° rainy days	5	5	4	2	5	5	5	10	13	14	12	5
	Mean minimum temperature (°C)	6,3	2,4	-5,4	-9,1	-11,2	-3,4	-2,2	0,8	4,9	9,4	12,2	11,7
	Mean maximum temperature (°C)	23,2	16,6	6,8	1,0	-0,3	9,3	12,9	15,3	18,1	25,1	28,7	28,1
	Mean temperature (°C)	14,8	9,5	0,7	-4,0	-5,8	2,9	5,4	8,1	11,5	17,2	20,4	19,9
2016-2017 cropping year	Precipitation (mm)	67,1	99,1	0,0	12,7	10,2	22,9	5,1	38,1	25,4	45,7	5,2	2,3
	n° rainy days	4	12	0	4	4	4	3	11	7	7	3	2
	Mean minimum temperature (°C)	7,2	2,2	-1,4	-14,3	-14,2	-8,3	-3,6	1,6	5,3	10,1	13,9	11,1
	Mean maximum temperature (°C)	21,2	13,6	12,9	-4,4	-4,4	1,6	9,2	15,2	21,9	26,2	33,5	29,5
	Mean temperature (°C)	14,2	7,8	5,7	-9,4	-9,3	-3,3	2,8	8,4	13,7	18,2	23,7	20,3
2017-2018 cropping year	Precipitation (mm)	19,8	53,3	7,62	22,9	12,7	35,6	7,6	5,1	27,9	63,5	10,2	12,7
	n° rainy days	3	7	1	2	3	2	1	6	9	14	3	5
	Mean minimum temperature (°C)	6,9	0,1	-6,1	-11,4	-12,7	-20,5	-9,9	-4,1	7,4	10,6	11,6	10,4
	Mean maximum temperature (°C)	22,9	13,4	5,8	-1,1	-0,7	-7,7	0,5	8,7	22,4	24,3	29,8	28,6
	Mean temperature (°C)	14,9	6,7	-0,2	-6,3	-6,7	-14,1	-4,7	2,3	14,9	17,4	20,7	19,6

b. Italy

		Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July
2016-2017 cropping year	Precipitation (mm)	89,8	74,0	23,6	1,4	52,4	6,2	32,0	100,0	51,6	19,6
	number of rainy days	8	8	4	1	5	3	3	9	4	1
	Mean minimum temperature (°C)	8,8	5,3	0,0	-3,3	2,1	4,9	7,4	11,7	17,0	17,7
	Mean maximum temperature (°C)	17,4	11,8	6,4	5,3	10,3	18,2	20,1	24,0	30,6	31,8
	Mean temperature (°C)	12,8	8,3	2,8	0,5	6,0	11,1	13,5	17,9	24,0	25,0
2017-2018 cropping year	Precipitation (mm)	8,8	126,6	8,8	3,8	116,2	67,6	18,8	37,0	67,0	64,6
	number of rainy days	6	10	6	3	11	13	7	12	9	10
	Mean minimum temperature (°C)	8,9	3,8	-0,9	1,4	-0,1	2,8	9,5	13,5	15,8	18,8
	Mean maximum temperature (°C)	20,7	11,8	7,8	9,1	6,3	11,3	21,4	24,4	28,7	31,4
	Mean temperature (°C)	14,2	7,4	2,8	5,0	3,0	7,0	15,4	18,7	22,4	25,0
2018-2019 cropping year	Precipitation (mm)	66,8	60,6	13,4	29,4	14,6	7,4	54,2	150,2	60,8	36,6
	number of rainy days	11	18	7	7	5	5	13	21	1	7
	Mean minimum temperature (°C)	10,8	8,1	-0,2	-2,1	0,7	3,1	7,0	9,7	17,0	18,2
	Mean maximum temperature (°C)	21,1	13,6	6,3	6,6	12,0	17,0	18,3	19,3	30,6	31,8
	Mean temperature (°C)	15,5	10,5	2,7	1,6	5,6	9,9	12,5	14,4	24,0	25,1

Tab. 23. Agronomic data of the Italian agronomic trial for wheat genotypes and years.

Wheat genotype	Yield (t/ha)	**	Lodging (%)	*	Weed incidence (%)	ns	Disease incidence (0-10)	ns	Disease severity (0-10)	ns
Vida	4,09	a	0	b	10		3,33		1,67	
Judee	3,72	ab	0	b	33,3		3		1,67	
Turkey	2,88	abc	53,3	a	30		2		1	
Marquis	2,81	abc	15	ab	30		1,33		0,33	
Turanicum	2,4	abc	5	b	50		2		0,67	
Peliss	2,31	abc	30	ab	40		2,3		1	
Alzada	1,74	abc	0	b	50		4		3	
Spelt	1,66	bc	33,33	ab	23,3		2,3		1	
Emmer	1,34	bc	6,67	b	43,3		1		0	
Einkorn	1	c	0	b	53,3		0,33		0	
Year	Yield (t/ha)	ns	Lodging (%)	ns	Weed incidence (%)	***	Disease incidence (0-10)	***	Disease severity (0-10)	ns
2016/2017	1,767		15		27	b	0,8	b	0,5	
2017/2018	2,795		14		20	b	1,9	b	0,7	
2018/2019	2,627		14		62	a	3,8	a	1,9	

Tab. 24. Protein and moisture content of the wheat samples.

a. samples grown in USA

Wheat genotype	First cropping year		Second cropping year		Third cropping year	
	Protein (g/100g dm)	Moisture (g/100g)	Protein (g/100g dm)	Moisture (g/100g)	Protein (g/100g dm)	Moisture (g/100g)
Einkorn	26,7	11,7	20,8	8,8	20,8	10,1
Emmer	Sample not available		19,7	9,3	19,1	9,8
Turanicum	19,4	11,8	15,8	9,8	17,1	10,6
Peliss	Sample not available		Sample not available		18,5	10,5
Alzada	24,7	11,2	17,3	9,5	19,7	10,0
Spelt	14,8	11,0	18,6	10,3	19,5	10,4
Turkey Red	14,6	11,2	17,0	11,1	16,8	10,8
Judee	13,7	11,4	17,0	10,0	16,4	10,5
Marquis	22,3	11,5	18,5	9,7	19,2	10,4
Vida	18,6	11,9	17,8	9,9	17,9	10,7

b. samples grown in Italy

Wheat genotype	First cropping year		Second cropping year		Third cropping year	
	Protein (g/100g dm)	Moisture (g/100g)	Protein (g/100g dm)	Moisture (g/100g)	Protein (g/100g dm)	Moisture (g/100g)
Einkorn	15,2	11,2	14,0	12,2	17,3	12,2
Emmer	12,3	11,7	12,5	12,6	11,1	12,4
Turanicum	12,7	11,9	12,8	12,9	10,3	12,4
Peliss	12,3	10,8	13,2	11,6	10,1	12,4
Alzada	12,6	11,2	14,2	12,2	11,5	12,3
Spelt	12,0	11,7	13,8	12,4	14,0	12,3
Turkey Red	11,3	11,4	11,4	12,4	10,1	12,9
Judee	11,4	11,6	11,3	12,6	10,1	12,8
Marquis	12,5	11,2	12,9	12,2	12,5	12,7
Vida	12,7	11,4	13,3	12,4	13,1	12,7

4.7 Mycotoxin levels

All the levels of micotoxins (DON) measured in this study were below the maximum limit allowed both for raw material durum wheat and oat (1750 ppb) and for the other raw material cereals (1250 ppb). However, in USA during the first cropping year (2015-2016) there were positive values of DON content in some samples and this can be explained by the fact that in 2016 there was more rainfall in spring and summer compared to 2017 and 2018. Similarly, in Italy the precipitation in spring time was most abundant in 2019 and in this year there were positive DON levels in 8 out of 10 wheat genotypes, even if below the limit.

Tab. 25. Mycotoxin levels (DON)

a. samples grown in USA

Wheat genotype	First cropping year (ppb)	Second cropping year (ppb)	Third cropping year (ppb)
Einkorn	< 200	< 200	< 200
Emmer	< 200	< 200	< 200
Turanicum	253	< 200	< 200
Peliss	< 200	< 200	< 200
Alzada	376	< 200	< 200
Spelt	204	< 200	< 200
Turkey Red	< 200	< 200	< 200
Judee	< 200	< 200	< 200
Marquis	< 200	< 200	< 200
Vida	296	< 200	< 200

b. samples grown in Italy

Wheat genotype	First cropping year (ppb)	Second cropping year (ppb)	Third cropping year (ppb)
Einkorn	< 200	< 200	< 200
Emmer	< 200	< 200	219
Turanicum	< 200	593	969
Peliss	< 200	383	342
Alzada	299	547	1122
Spelt	< 200	< 200	260
Turkey Red	< 200	< 200	271
Judee	< 200	< 200	920
Marquis	< 200	< 200	< 200
Vida	< 200	< 200	373

4.8 Enzymatic assays

A colorimetric test was used to determine the inhibitory activity against human salivary α -amylase and against bovine trypsin of ten selected wheat genotypes cultivated for three consecutive years in two growing areas. The inhibitory activity was calculated as percentage with respect to the control.

4.8.1 Alpha-amylase inhibitory activity assay

The results of the alpha-amylase inhibitory activity showed that there was high variability among the 10 wheat genotypes analysed (Table 26 and Figure 22). The general linear model (GLM) highlighted that the contribution of the fixed (genotype) and the random (year) factors, as well as their interactions, to the variability of the alpha-amylase inhibitory activity were significant in both growing areas.

The mean values for the alpha-amylase inhibitory activity ranged between 1,12% and 66,63%, with the lowest activity observed in Einkorn sample in both growing areas. In general, with the exception of Emmer, the hexaploid wheat genotypes showed higher inhibitory activities than tetraploid and diploid wheat genotypes. Moreover, it is not possible to highlight differences between ancient and modern wheat genotypes.

Considering the sequence results of this study discussed in paragraphs 4.1 - 4.5, some considerations can be made. Choudhury et al. (1996) showed that WDAI 0.19, even if consisted of 24.9% of total wheat albumin, it contributed approximately 80% of the inhibiting activities against human α -amylase. Interestingly, analysing the WDAI sequences obtained in this study, the only haplotypes which showed a perfect homology with the three inhibitor spots important for the human alpha-amylase inhibitory activity (Franco et al., 2000) were present in hexaploid wheat genotypes which showed also the highest alpha-amylase inhibitory activity in this study.

Gelinas et al. (2018) measured the alpha-amylase inhibitory activity against human α -amylase in different ancient and modern tetraploid and hexaploid wheat genotypes. Using a colorimetric assay, in line with the results of this study, he confirmed that most of the hexaploid wheat cultivars had higher alpha-amylase inhibitory activity than tetraploids such as *T. durum* samples. Moreover, he also highlighted that ancient hexaploid wheat samples

such as spelt had similar alpha-amylase inhibitory activity as common wheat and that in general alpha-amylase inhibitory activity did not change with respect to ancient or recently developed wheat varieties, in agreement with the results of this study. Interestingly, within the common wheat genotypes, most of the hard red spring wheat tested by Gelinias et al. had higher alpha-amylase inhibitory activity than the hard red winter wheat. However, Gelinias et al. didn't have any information about the growing conditions and the agronomic performance of the tested wheat genotypes. In this study, the hard red winter common wheat genotypes (Turkey Red and Judee) showed higher activities than the hard red spring common wheat genotypes (Marquis and Vida) in USA where a winter and a spring cycle respectively were used. Differently, in Italy where a winter cycle was used for all the wheat genotypes, Marquis showed a similar inhibitory activity than Turkey Red and Judee, but Vida seemed not to be influenced by the different growing cycle adopted in Italy and showed similar inhibitory activities than in USA. So, beside the genotype, also the cropping cycle can influence the expression of these inhibitors. However, further studies on more samples are needed to confirm what observed in this study.

The lowest activity was observed in Einkorn sample in both growing areas with values which are much lower than the others or even close to zero. Capocchi et al. (2013) reported previously that complete inhibition of the amylase activity cannot be achieved due to substrate competition with the inhibitor for binding the α -amylase.

The very low level (or even close to zero) of alpha-amylase inhibitory activity of Einkorn is in agreement with previous studies. Reig-Otero et al. (2017) highlighted that immune transference testing and mass spectrometry studies couldn't detect inhibition of human α -amylases by einkorn samples; using a colorimetric assay any inhibitory activity against human and pest amylases was obtained in einkorn samples (Bedetti et al., 1974; Sanchez-Monge et al., 1996; Vittozzi & Silano, 1976). Zoccatelli et al. (2012) with an immunoblotting after SDS-PAGE and Urea-PAGE using polyclonal antibodies (PABs) raised against 0.19 and 0.28 alpha-amylase inhibitors couldn't obtain any positive signal with monococcum samples. Similarly, with a reverse zymographic technique using human salivary amylase and a pest amylase, no monococcum accession was able to show any inhibition activity. Rogniaux et al. (2015) used a targeted MS/MS approach and confirmed the low level of expression of the alpha-amylase inhibitors in the diploid species, in accordance with the results of previous 2D-electrophoresis that showed the absence of spots

corresponding to alpha-amylase inhibitors and almost no associated IgE binding (Larré et al., 2011).

It was previously suggested that the apparent absence of alpha-amylase inhibitory activity of monococcum (as observed also in this study) could be explained by the fact that the corresponding coding genes might be expressed at very low level or even silenced, perhaps as a result of gene mutations that prevent the translation into the mature protein (Garcia-Maroto et al., 1990). This hypothesis is in agreement with the gene sequencing results of this study. In fact, Einkorn didn't show any WTAI-CM3 gene sequence and so it should not be able to produce tetrameric alpha-amylase inhibitor proteins. Moreover, in agreement with Wang et al. (2005), all the WDAI gene sequences showed an insertion of C at position 160 which resulted in a premature stop codon and the impossibility to synthesize the correct WDAI mature protein. Lastly, Einkorn showed WMAI sequences, but it is known that WMAI proteins are highly active against α -amylase of *T. molitor* and only weakly inhibits the α -amylases from human saliva and pancreas (Payan et al., 2004).

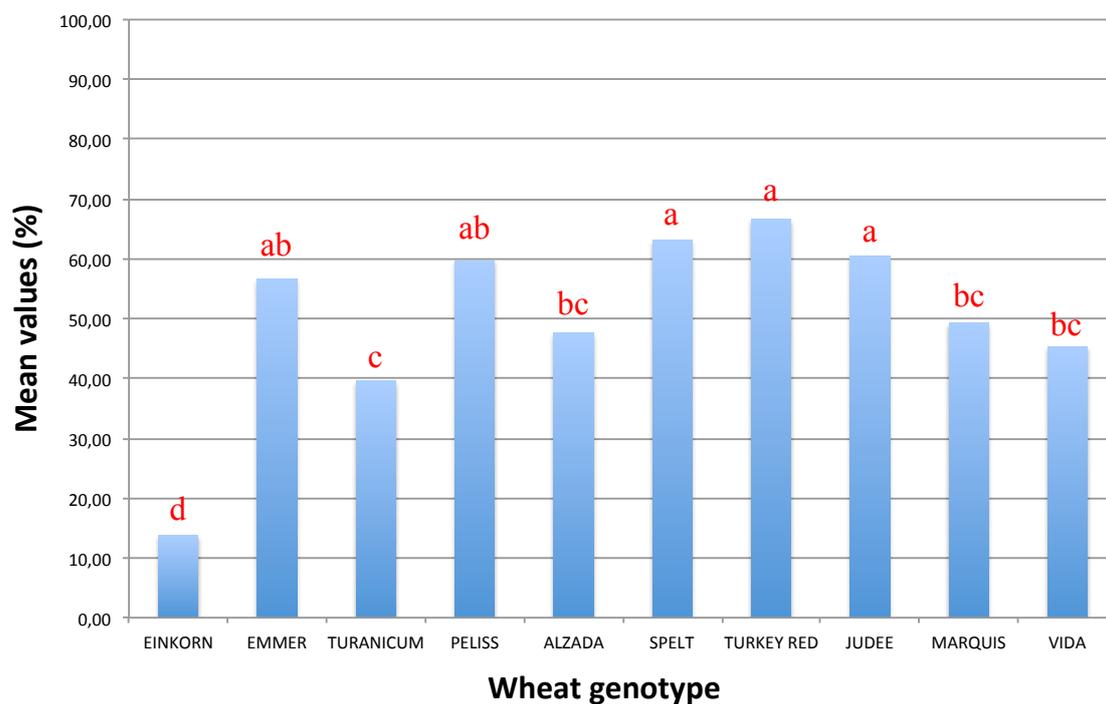
The apparent low levels of proteins belonging to the alpha-amylase inhibitors in monococcum wheat could result in a low pest defence performance in comparison to other wheats. However, this is not the case since monococcum is well known for its high resistance to pest and disease and it is successfully cultivated (Rajaram et al., 2001; Shi, Leath, & Murphy, 1998). It should be also pointed that in this study only the inhibitory activity against human amylase was assayed and that the only α -amylase inhibitor gene which seem to be functional in Einkorn is WMAI protein which is mainly active against insect enzymes. Moreover, Einkorn showed high levels of trypsin inhibitory activities in this study (see paragraph 4.8.2).

Tab. 26. Mean values (%) of the alpha-amylase inhibitory activity for wheat genotypes, years and growing areas. Different letters indicate statistically significant different means for $P < 0,05$.

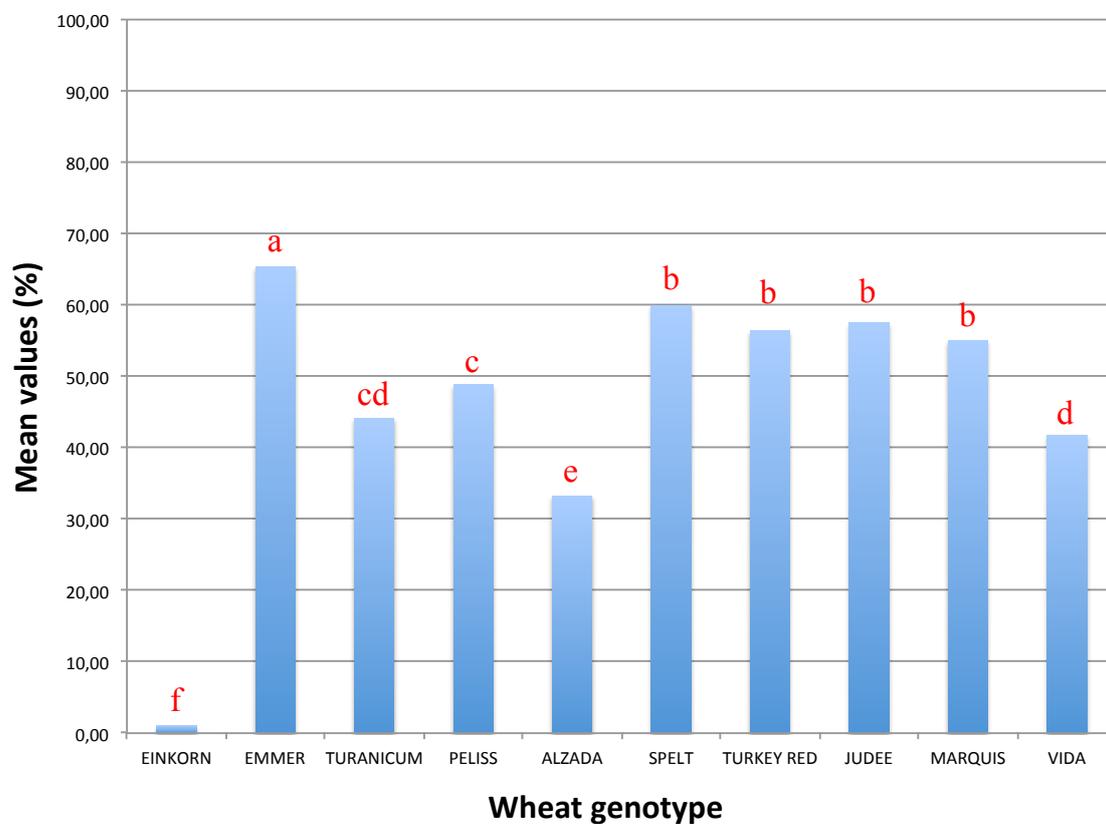
USA		ITALY	
Year	**	Year	**
2016	40,83 (b)	2017	53,43 (a)
2017	56,26 (a)	2018	47,06 (b)
2018	49,75 (a)	2019	38,52 (c)
Genotype	**	Genotype	**
EINKORN	13,76 (d)	EINKORN	1,12 (f)
EMMER	56,62 (ab)	EMMER	65,42 (a)
TURANICUM	39,6 (c)	TURANICUM	44,14 (cd)
PELISS	59,73 (ab)	PELISS	48,80 (c)
ALZADA	47,61 (bc)	ALZADA	33,28 (e)
SPELT	63,15 (a)	SPELT	59,95 (b)
TURKEY RED	66,63 (a)	TURKEY RED	56,42 (b)
JUDEE	60,50 (a)	JUDEE	57,47 (b)
MARQUIS	49,24 (bc)	MARQUIS	55,00 (b)
VIDA	45,35 (bc)	VIDA	41,77 (d)
Y*V	**	Y*V	**

Fig. 22. Mean values of alpha-amylase inhibitory activity (%) in USA (a) and in Italy (b).

a. USA



b. Italy



The mean values for each genotype in each cropping year are shown in Fig. 23 a (USA) and b (Italy).

In USA, 2017 and 2018 cropping years showed higher alpha-amylase inhibitory activities than 2016. All the wheat genotypes showed a similar trend among the three cropping years with the exceptions of Einkorn and Turanicum which significantly increased and decreased respectively their activities in 2018. During the three-year growing period the weather conditions, such as temperature and rain precipitation could influence the biological activities. Therefore, the temperatures and the rainfall during the three-year growing period were recorded. As it is showed in Table 22, in USA there was more rainfall in spring and summer in 2016 compared to the same period in 2017 and 2018. This could be the reason why only in 2016 in USA there were positive values of DON content in some samples (even if below the law limits) (see Table 25). Considering the rainfall amount during the period May-August, there is a negative correlation with the alpha-amylase inhibitory activity.

In Italy, the highest alpha inhibitory activities are in 2017 and the lowest in 2019. Overall, there weren't samples showing a different trend compared to the average. The precipitation was most abundant in 2019 and also in this case this is correlated with the presence of positive levels of DON in 8 out of 10 wheat genotypes, even if below the law limits. Considering the rainfall amount during the period April-July, also in this case there is a negative correlation with the alpha-amylase inhibitory activity.

In both cultivation areas, higher precipitation correlates with lower alpha-amylase inhibitory activities, differently from what could be expected for a class of protein involved in plant defence. However, it should be taken into consideration that this study measured the inhibitory activity against human α -amylase and not against insect α -amylases and that it is known in literature that each type of ATIs shows different specificities against different enzymes (Priya et al., 2013).

Regarding the mean high temperatures, there is a positive correlation with the alpha-amylase inhibitory activity both in USA and in Italy considering the same period analysed for precipitation.

In this study the abiotic factors seem to influence the inhibitory activity of wheat genotypes. Yang et al. (2011) also reported changes in the levels of WMAI, WDAI and the WTAI subunits CM1, CM3 and CM17 in grain from the bread wheat cv. Vinjett subjected to

different combinations of temperature and water stress. Another study evidenced that the temperature stress conditions can influence the ATIs levels in seed cultivated in growth chambers. Low-temperature conditions during seed development increased the content of protein families like ATIs and this indicated a possible stress-related function of ATIs during seed development or in seed germination (Juhasz, et al. 2018).

Regarding the correlation between different cultivation areas, cropping years, climatic conditions and the alpha-amylase inhibitory activities of cereals, only a few studies are available. Piasecka et al. (2007) also highlighted a negative correlation between the monthly precipitation (particularly in January and in March) and the alpha-inhibitory activities in wheat, more specifically, against the hog pancreas α -amylase, but not against human α -amylase and insects α -amylase. In another study Piasecka et al. (2012) observed different alpha-amylase and trypsin inhibitory activities against mammalian enzymes in dependency of harvest year and wheat varieties. Using a peptidomic approach, Prandi et al. (2013) quantified CM3 protein levels in different durum wheat varieties and showed strong differences on the basis of the cultivation area. They concluded that this is in agreement with the defensive function of this protein, which thus can be influenced by environmental factors, such as weather conditions and biotic stress.

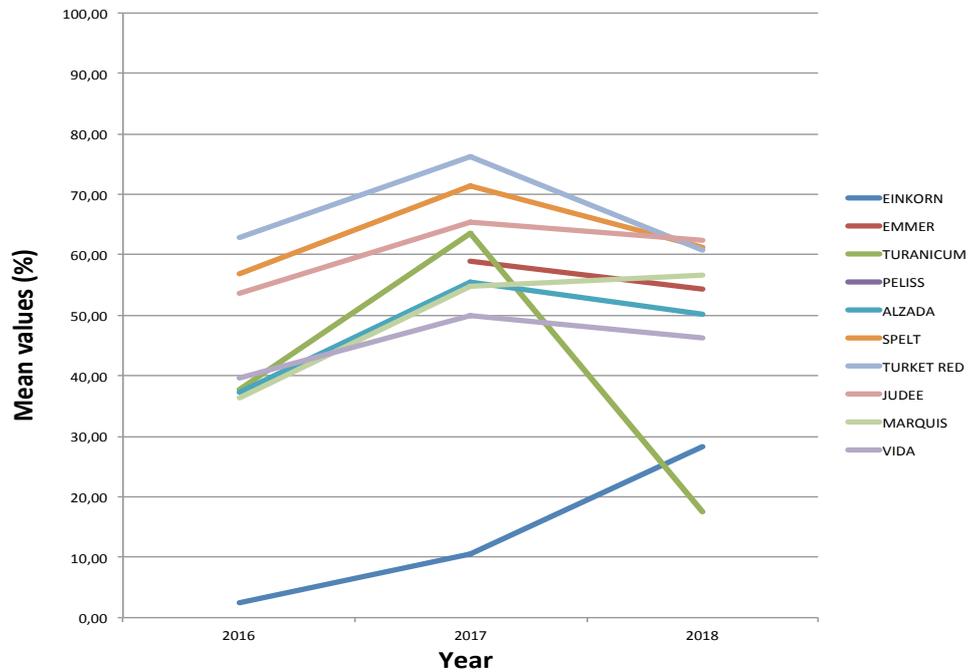
Considering the agronomic data of the Italian agronomic trial, it can be noted that the 2018-2019 cropping year, which showed the lowest alpha-amylase inhibitory activity, was characterized by the highest incidence of both weed and disease. More specifically, considering the mean values for the three cropping years for the hexaploid wheat genotypes, Vida showed the lowest activity and had the highest disease incidence. A similar behaviour was found for Aldada in the tetraploid wheat group. So, from this data, it doesn't seem that the alpha-amylase inhibitory activity is activated by the biotic stress monitored in this study like pathogens (especially fungal pathogens), however it was possible to evidence a negative correlation. Whether these biotic stress factors have a direct or indirect effect on the observed activities, it is not possible to say from this data. Moreover, the functional significance of the activation of this enzymatic activity remains to be tested and to be correlated to specific metabolic pathways in response to biotic and/or abiotic factors

The alpha-amylase inhibitory activity in this study was measured against a human enzyme and each type of inhibitor has a different specificity for difference sources of α -amylase. Different types of alpha-amylase inhibitors can be differentially activated by different types

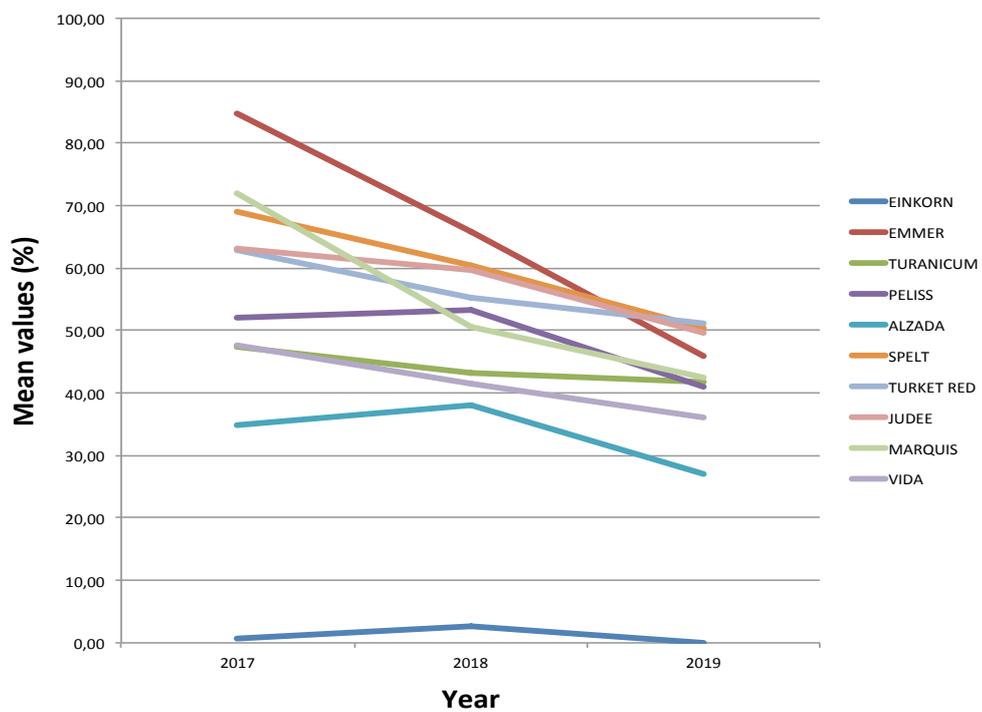
of stress and further studies are needed to understand possible mechanisms able to modulate the enzymatic activity in response to different types of biotic stress.

Fig. 23. Interaction of the α -amylase inhibitory activity with year for each genotype.

a. USA

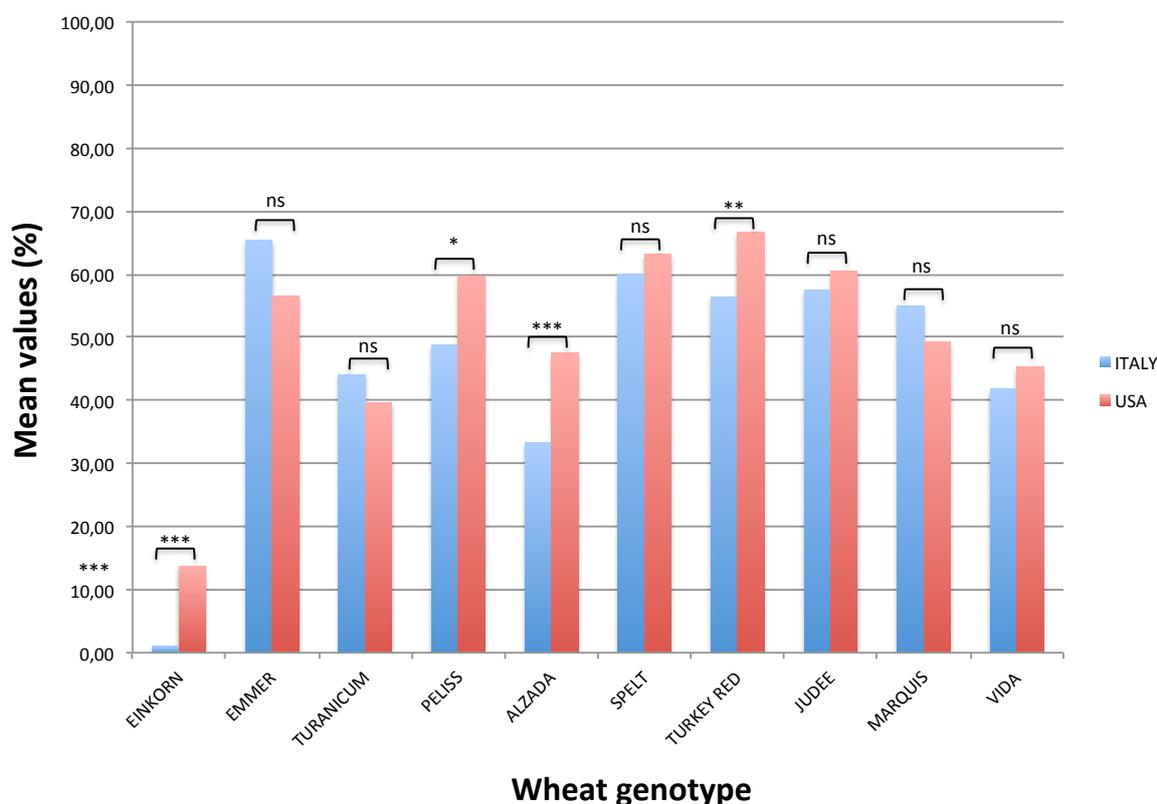


b. Italy



The two cultivation areas differently affected the wheat genotypes, in particular Einkorn, Peliss, Alzada and Turkey Red showed significantly higher α -amylase inhibitory activities in USA (Fig. 24).

Fig. 24. Differences between the mean values of alpha-amylase inhibitory activity in USA and in Italy for each wheat genotype. Significant level: *, $P < 0,05$; ** $P < 0,01$; *** $P < 0,001$. ns, not significant.



In addition, there were negative correlations between the protein content and the α -amylase inhibitory activities both in USA ($r = - 0,626$; $p < 0,01$) and in Italy ($r = - 0,526$; $p < 0,01$). Prandi et al. (2013) previously observed that locations that yielded more protein content consistently produced lower amounts of CM3 ATI.

4.8.2 Trypsin inhibitory activity assay

Trypsin inhibitors are anti-nutritional factors and defence proteins in plant against pest attacks. CM inhibitors were previously identified as the main contributors of trypsin inhibition in wheat (Call et al., 2019).

Literature survey on Science Direct and Web of Science database showed that data regarding trypsin inhibitory activity (TIA) of wheat are very scarce. A few studies compared TIA in wheat with other cereals and legumes. Boisen et al. (1989) reported significantly higher TIA values in rye and triticale than in wheat samples. Moreover, the average TIA of wheat was found to be approximately 1% of that of soy flour. Chang et al. (1979) also confirmed that TIA in cereals is much lower (1–3%) than in soy extracts, and Mossor and Skupin (1985), highlighted that trypsin inhibition activity in wheat is 1.5% of soybean. In agreement with the previous studies, Mikola & Mikkonen (1999) concluded that the amounts of trypsin inhibitors vary in different cereals with rye showing a 10-fold higher activity than that of wheat and oats.

Regarding more specifically the genus *Triticum*, a few studies can be cited. Priya et al. (2013) evaluated the specificity of the α -amylase and trypsin inhibitors in 54 genotypes of *T. aestivum* against the amylases and trypsins from both different insects and mammals and found large variation in trypsin inhibitory activity against insect and bovine trypsin. Also Warchalewski et al. (1989) showed that the trypsin inhibitory activity differed considerable among the six wheat varieties studied. Piasecka et al. (2012) studied two cultivars of *T. aestivum*, triticale and rye and found that the anti tryptic activity showed great differences between the two varieties of wheat and were lower than in rye.

A recent study (Kostekli et al., 2017) determined the anti tryptic activities (TIA) in wheat, rye mix, mixed cereals and whole wheat flours and breads made with these flours and studied the effects of fermentation, baking and in vitro digestion on TIA. He found that mixed cereal flour had the highest and wheat flour had the lowest TIA. Moreover, trypsin inhibitors found in mixed cereals lost their activities during fermentation and baking and trypsin inhibitors of wheat were pepsin stable.

To the best of our knowledge, this is the first study investigating the trypsin inhibitory activities in different wheat genotypes of genus *Triticum* with different ploidy and with different year of release.

Table 27 and Figure 25 show that there was high variability in trypsin inhibitory activity among the 10 wheat genotypes analysed. The general linear model (GLM) showed that the contribution of the fixed (genotype) and the random (year) factors to the variability of the trypsin inhibitory activity were significant, but not their interactions. The mean values of the trypsin inhibitory activity ranged between 18,61 % and 100 %, with the lowest activity observed in Peliss sample and the highest activity observed in Einkorn sample in both growing areas. After Einkorn, the hexaploid wheat genotypes Turkey Red, Judee and Marquis showed the highest inhibitory activities in both growing areas. Moreover, as for the alpha-amylase inhibitory activity, it is not possible to highlight differences among ancient and modern wheat genotypes.

Interestingly, as observed for the alpha-amylase inhibitory activity, the hard red winter common wheat genotypes (Turkey Red and Judee) showed higher activities than the hard red spring common wheat genotypes (Marquis and Vida) in USA where they were cultivated on a winter and on a spring cycle respectively. Differently, in Italy where all the genotypes were cultivated on a winter cycle, Marquis showed a similar inhibitory activity than Turkey Red and Judee, but Vida seemed not to be influenced by the different growing cycle adopted in Italy and showed similar inhibitory activities than in USA. The same behaviour of both Marquis and Vida were observed for the alpha-amylase inhibitory activity. So, beside the genotype, also the cropping cycle can influence the expression of these inhibitors. However, further studies on more samples are needed to confirm this effect.

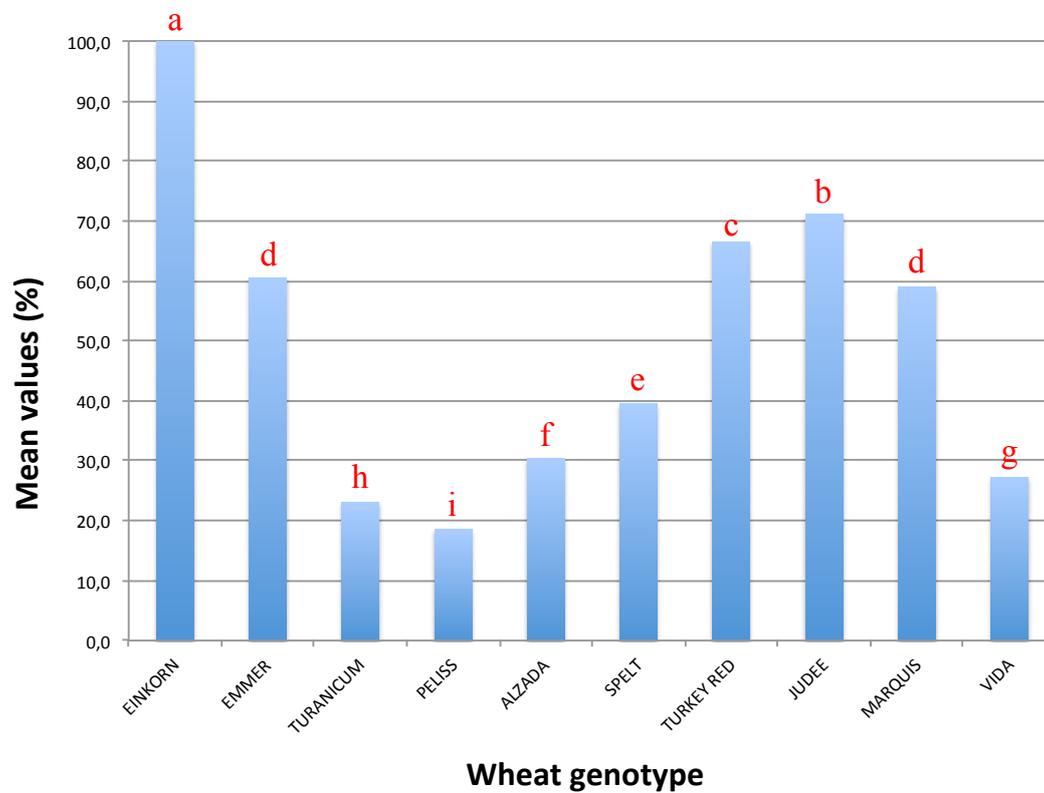
Regarding Einkorn sample, to the best of our knowledge there aren't other studies which evaluated the trypsin inhibitory activity in *T. monococcum* samples. In both growing areas and in all the three cropping years, Einkorn showed much higher trypsin inhibitory activity compared to all the other samples at the assay condition of this study. This result is in line with all the genetic data obtained in this study. In fact, taking into account the deduced amino acid sequences, Einkorn could be well differentiated from the other samples, as it can be also seen in Fig. 20. showing the PCA of all the four genes taken together. This very high anti tryptic activity could be related to its well-known high resistance to pest and disease.

Tab. 27. Mean values (%) of the trypsin inhibitory activity for wheat genotypes, years and growing areas. Different letters indicate statistically significant different means for $P < 0,05$.

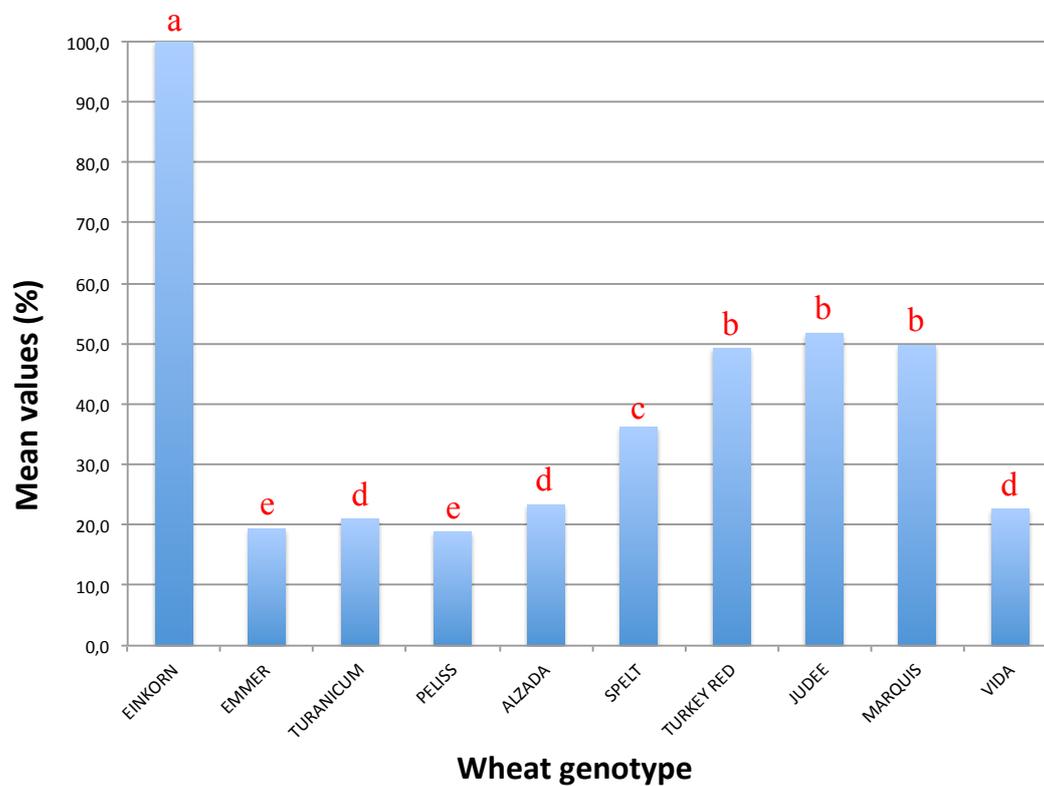
USA		ITALY	
Year	**	Year	**
2016	63,31 (a)	2017	42,53 (a)
2017	43,72 (c)	2018	39,57 (b)
2018	49,04 (b)	2019	35,47 (c)
Genotype	**	Genotype	**
EINKORN	100,00 (a)	EINKORN	100,00 (a)
EMMER	60,55 (d)	EMMER	19,34 (e)
TURANICUM	23,08 (h)	TURANICUM	20,93 (d)
PELISS	18,61 (i)	PELISS	18,77 (e)
ALZADA	30,45 (f)	ALZADA	23,30 (d)
SPELT	39,46 (e)	SPELT	36,18 (c)
TURKEY RED	66,54 (c)	TURKEY RED	49,27 (b)
JUDEE	71,06 (b)	JUDEE	51,70 (b)
MARQUIS	59,00 (d)	MARQUIS	49,76 (b)
VIDA	27,26 (g)	VIDA	22,64 (d)
Y*V	ns	Y*V	ns

Fig. 25. Mean values of trypsin inhibitory activity (%) in USA (a) and in Italy (b).

a. USA



b. Italy



Regarding the interaction of trypsin inhibitory activity with year for each genotype, in USA 2016 cropping year showed the highest and 2017 the lowest trypsin inhibitory activities. In Italy, the highest alpha inhibitory activities are in 2017 and the lowest in 2019. All the genotypes showed a similar trend among the three cropping years both in USA and in Italy.

The rainfall correlated with lower trypsin inhibitory activities in Italy but with higher activities in USA and the mean high temperatures showed a negative correlation with the trypsin inhibitory activity in USA and positive correlation in Italy considering the period May-August for USA and April-July for Italy. So, unlike what observed for alpha-amylase inhibitory activities, the trypsin inhibitory activity showed different correlations in the two growing areas.

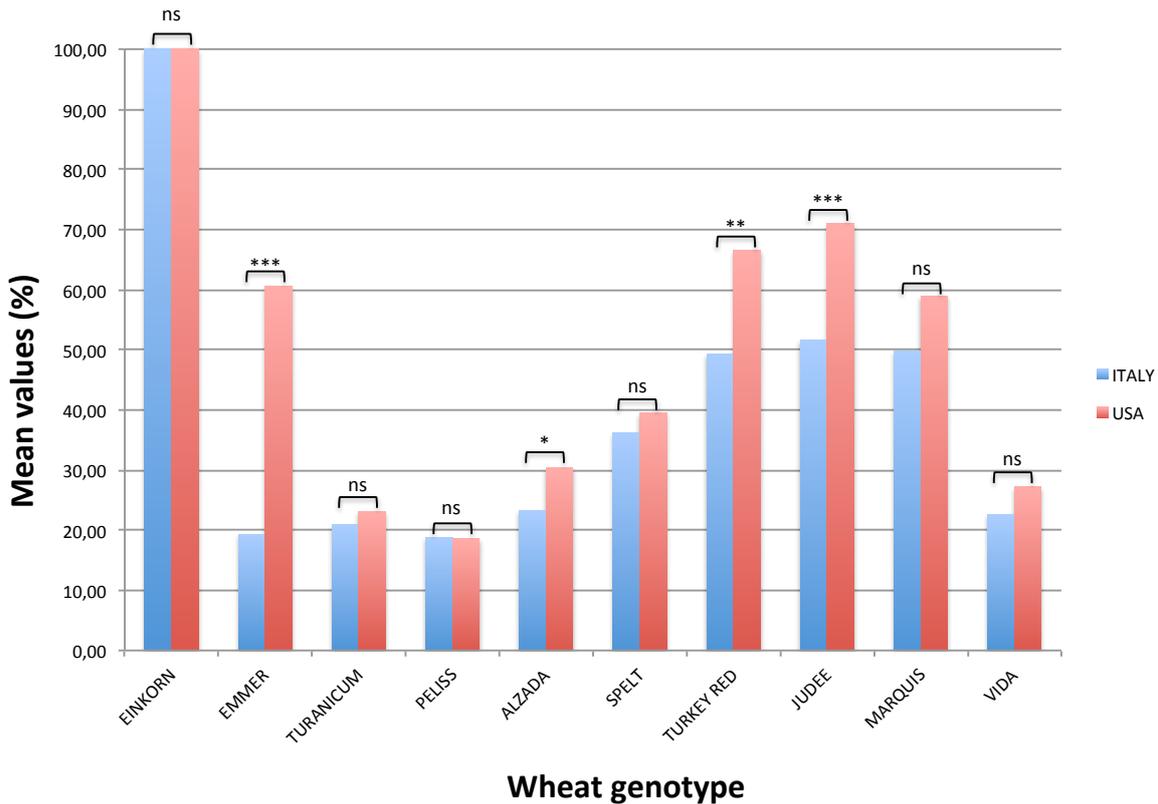
There are very few studies which investigated the correlation between different cultivation areas, cropping years and climatic conditions and the trypsin inhibitory activities of cereals. Piasecka et al. (2007) showed that inhibitory activities against bovine trypsin in two cultivars of *T. aestivum*, triticale and rye were statistically significant in all sources of changes (variety, genus, year of harvest as well as interaction between them) with the exception of rainfall.

Overall the 2018-2019 cropping year in Italy was characterized by higher disease incidence (also confirmed by higher DON levels this year), but it was not possible to evidence any correlation with the trypsin inhibitory activity of the samples.

The two cultivation areas differently affected the wheat genotypes, in particular Emmer, Judee, Turkey Red and Alzada showed significantly higher trypsin activities in USA, even three times more considering Emmer (Fig. 26).

As for the alpha-amylase inhibitory activities, there were negative correlations between the protein content and the trypsin inhibitory activities both in USA ($r = -0,273$; $p < 0,01$) and in Italy ($r = -0,489$; $p < 0,01$).

Fig. 26. Differences between the mean values of trypsin inhibitory activity in USA and in Italy for each wheat genotype. Significant level: *, $P < 0,05$; **, $P < 0,01$; ***, $P < 0,001$. ns, not significant.



Beside the important role of α -amylase/trypsin inhibitors for the plant protection against pest attacks, this class of enzymes has also an important role on human nutrition. It was also found that alpha-amylase and trypsin inhibitors in wheat bread were stable to digestive enzymes.

The alpha-amylase inhibitory activity of wheat-based food has significant relevance for its dietary and therapeutic support to potentially maintain postprandial glucose homeostasis, which is essential for prevention and overall management of early stages type 2 diabetes (Christopher et al., 2018). However, a side effect of this inhibition activity is that undigested and unabsorbed starchy constituents are fermented in the distal gut by the microbial flora and can cause mild to moderate symptoms, including flatulence and bloating, in sensitive individuals to carbohydrate malabsorption (Kawamori et al., 2009).

Regarding the trypsin inhibitory activity, high activities can be negative from a nutritional point of view because can limit the absorption of protein during food assumption. It was previously described that excessive amount of trypsin inhibitors in food can progressively lead to pancreas hyperplasia or hypertrophy and is correlated to a higher risk of pancreas cancer. Imbalanced and uncontrolled proteolysis can also lead to tissue damage such as inflammation, hypertension, gastric ulcer, tumor growth, and metastasis (Call et al., 2019).

For these reasons, protease inhibitors are considered as “antinutritional compounds”, due to their inhibitory activity on digestive enzymes in humans and animals.

According to the results of this study, the genotypes which showed in general the lowest inhibitory activities against both types of enzymes were Turanicum and Alzada and the highest activities the hexaploid Judee, Turkey Red and Marquis. However, considering the potential effect on glycaemic control, Emmer cultivated in Italy showed the best combination with high alpha-amylase inhibitory activity, but low trypsin inhibitory activity. The evidenced differences, even if to be confirmed in further studies also *in vivo*, can be useful information for the selection of the better genotypes for specific food needs.

However, food processing causes significant loss in trypsin inhibitory activities (Kostekli et al, 2017). Regarding alpha-amylase inhibitory activity, little residual inhibitory activity towards α -amylase from human saliva was detected in commercial and laboratory-made cereal foods (Gelinas et al., 2018).

So, both the alpha-amylase and trypsin inhibitory activities should be tested in processed food to assess the residual activity and drawing final conclusions about the impact of a specific wheat genotype on human health.

5. Conclusions

Wheat amylase-trypsin inhibitors (ATIs) are an important family of wheat proteins, which play an important role in plant defence against pest attacks. ATIs are also of great interest for their impact on human health. Besides being involved in wheat allergy, recently, ATIs have been identified as major stimulators of innate immune cells by activation of the TLR4 complex.

The present literature lacks information about the differences in genetic sequences and activities of ATIs among different types of wheat, especially considering the ancient wheat genotypes which have gained an increased interest over the last few years for their healthier nutritional profile. Moreover, very little information is available about the influence of growing conditions on the expression of these inhibitors.

To the best of our knowledge, this is the first study investigating both the alpha-amylase and trypsin inhibitory activities in ancient and modern wheat genotypes with a different ploidy level and cultivated in two growing areas for three consecutive years with the aim to find possible influences of the growing conditions. Moreover, the genetic sequences of four representative alpha-amylase/trypsin inhibitor genes have been determined and differences among the samples have been discussed.

In literature, information about some ATIs gene sequences (in particular CMx genes) are scarce and are not available for many types of wheat. In this study we obtained the sequences of ATIs genes of ten different wheat genotypes with a different ploidy level, and some of these, for example *T. monococcum* and *T. turgidum* spp. *turanicum*, were never been sequenced for some or all ATIs genes before. So, this study has implemented the available information about ATIs sequences.

Considering the deduced amino acid sequences of all the four ATIs genes studied, the PCA analysis evidenced that the ten wheat genotypes can be differentiated on the basis of their ploidy level. This is in agreement with the fact that ATIs genes are present in different chromosomal sets. Einkorn was the most divergent sample both considering all the four genes taken together and each specific gene. It is also important to point out that PCA analysis doesn't take into consideration the biological significance of a specific amino acid change.

We obtained a different genetic variability intra genotype. The hexaploid wheat accessions showed high diversity among the five genetic sequences obtained in each genotype. Differently, the tetraploid wheat accessions were characterized by a higher homology of the genetic replicates in each accession. The diploid sample showed high variability which is probably due to the fact that this sample is a population of different individuals.

The literature lacks information on which sequences of ATIs have immunogenic and inflammatory potential, so it was not possible to make any consideration on the basis only of the deduced amino acid sequences. However, the genetic sequence differences highlighted in this study among the ten genotypes can be the basis for further studies aimed at identifying those sequence in ATIs genes which are linked to higher inflammatory and allergenic activity.

Considering the *in vitro* inhibitory activities, there was high variability among the ten wheat genotypes analysed and the contribution of the genotype and the cropping year was significant for both the alpha-amylase and the trypsin inhibitory activity. Overall, the hexaploid wheat genotypes showed the highest inhibitory activities. Einkorn had a completely different behaviour for the two activities: showed a very low or even absent alpha-amylase inhibitory activity and the highest trypsin inhibitory activity. It is not possible to differentiate ancient and recently developed wheat genotypes on the basis of their *in vitro* ATIs activity against mammalian enzymes

The weather conditions differently affected the two inhibitory activities. In both cultivation areas, higher precipitation and lower high mean temperatures correlated with lower alpha-amylase inhibitory activities. However, it is not possible to establish if these factors have a direct effect on the observed activities. Moreover, the functional significance of the activation of this enzymatic activity remains to be tested and to be correlated to specific metabolic pathways in response to biotic and/or abiotic factors. Regarding the trypsin inhibitory activity, the weather conditions showed different correlations considering the two growing areas: the rainfall correlated with lower trypsin inhibitory activities in Italy but with higher activities in USA and the mean high temperatures a negative correlation with the trypsin inhibitory activity in USA and positive in Italy.

There were negative correlations between the protein content and both the α -amylase and trypsin inhibitory activities considering both USA and in Italy.

6. Bibliography

Almagro Armenteros, J. J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., Heijne, G. and Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423 (2019).

Altenbach, S., Vensel, W. H. & Dupont, F. M. The spectrum of low molecular weight alpha-amylase/protease inhibitor genes expressed in the US bread wheat cultivar Butte 86. *BMC Res. Notes* 4, 242 (2011).

Alves, T. O., D’Almeida, C.T.S., Victorio, V.C.M., Souza, G.H.M.F., Cameron, L.C. Ferreira, M.S.L. Immunogenic and allergenic profile of wheat flours from different technological qualities revealed by ion mobility mass spectrometry. *J. Food Compos. Anal.* 73, 67–75 (2018).

Aune, D., Keum, N., Giovannucci, E., Fadnes, L.T., Boffetta, P., Greenwood, D.C., Tonstad, S., Vatten, L.J., Riboli, E., Norat, T. Whole grain consumption and risk of cardiovascular disease, cancer, and all cause and cause specific mortality: systematic review and dose-response meta-analysis of prospective studies. *BMJ* 353:i2716 (2016).

Baker, J. E., Woo, S. M., Throne, J. E. & Finney, P. L. Correlation of α -Amylase Inhibitor Content in Eastern Soft Wheats with Development Parameters of the Rice Weevil (Coleoptera: Curculionidae). *Environ. Entomol.* 20, 53–60 (1991).

Bardella, M. T., Elli, L., & Ferretti, F. Non celiac gluten sensitivity. *Current Gastroenterology*, 18, 63 (2016).

Bedetti, C., Bozzini, A., Silano, V., & Vittozzi, L. Amylase protein inhibitors and the role of *Aegilops* species in polyploid wheat speciation. *Biochim Biophys Acta.* 362, 299–307 (1974).

Bilodeau, M. & Duchesne, P. Principal component analysis from the multivariate familial correlation matrix. *J. Multivar. Anal.* 82, 457–470 (2002).

Boisen, S. Comparative studies on trypsin inhibitors in legumes and cereals. In J. Huisman, T. F. B. van der Poel, & I. E. Liener (Eds.). *Recent advances of research in antinutritional factors in legume seeds*, Proceedings of the First International Workshop on 'Antinutritional Factors (ANF) in Legume Seeds', November 23-25, 1988, Wageningen, The Netherlands

(pp. 118–120). Wageningen: Centre for Agricultural Publishing and Documentation (Pudoc) (1989).

Brik, Y. Plant Protease Inhibitors: Significance in nutrition. Plant protection, Cancer prevention and genetic engineering, p 52-96. Springer-Verlag, Berlin (2003).

Buonocore, V., Petrucci, T. & Silano, V. Wheat protein inhibitors of α -amylase. *Phytochemistry*, 16, 811–820 (1977).

Call, L. Reiter, E.V., Wenger-Oehn, G., Strnad, I., Grausgruber, H., Schoenlechner, R., D'Amico, S. Development of an enzymatic assay for the quantitative determination of trypsin inhibitory activity in wheat. *Food Chem.* 299, 125038 (2019).

Capocchi, A., Muccilli, V., Cunsolo, V., Saletti, R., Foti, S., Fontanini, D. A heterotetrameric alpha-amylase inhibitor from emmer (*Triticum dicoccon* Schrank) seeds. *Phytochemistry* 88, 6–14 (2013).

Carbonero, P. & García-Olmedo, F. A Multigene Family of Trypsin/ α -Amylase Inhibitors from Cereals. *Seed Proteins* 617–633 (1999).

Ceroni, A., Passerini, A., Vullo, A. & Frasconi, P. Disulfind: A disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res.* 34, 177–181 (2006).

Chang, C. R., & Tsen, C. C. Note on trypsin inhibitor activity in the acetate extract of cereal samples. *Cereal Chemistry*, 56, 493–494 (1979).

Choudhury, A., Maeda, K., Murayama, R., Di Magno, E.P. Character of a wheat amylase inhibitor preparation and effects on fasting human pancreatico-biliary secretions and hormones. *Gastroenterology*. 111(5):1313-20 (1996).

Christopher, A., Sarkar, D., Zwinger, S. & Shetty, K. Ethnic food perspective of North Dakota Common Emmer Wheat and relevance for health benefits targeting type 2 diabetes. *J. Ethn. Foods* 5, 66–74 (2018).

Cuccioloni, M., Mozzicafreddo, M., Ali, I., Bonfili, L., Cekarini, V., Eleuteri, A.M., Angeletti, M. Interaction between wheat alpha-amylase/trypsin bi-functional inhibitor and mammalian digestive enzymes: Kinetic, equilibrium and structural characterization of binding. *Food Chem.* 213, 571–578 (2016).

Cuccioloni, M., Mozzicafreddo, M., Bonfili, L., Cecarini, V., Giangrossi, M., Falconi, M., Saitoh, S., Eleuteri, A.M. & Angeletti, M. Interfering with the high-affinity interaction between wheat amylase trypsin inhibitor CM3 and toll-like receptor 4: In silico and biosensor-based studies. *Sci. Rep.* 7, 4–6 (2017).

Dinu, M., Whittaker, A., Pagliai, G., Benedettelli, S. & Sofi, F. Ancient wheat species and human health: Biochemical and clinical implications. *J. Nutr. Biochem.* 52, 1–9 (2018).

Dupont, F. M., Vensel, W. H., Tanaka, C. K., Hurkman, W. J. & Altenbach, S. B. Deciphering the complexities of the wheat flour proteome using quantitative two-dimensional electrophoresis, three proteases and tandem mass spectrometry. *Proteome Sci.* 9, 10 (2011).

Fasano, A., Sapone, A., Zevallos, V. & Schuppan, D. Nonceliac Gluten Sensitivity and Wheat Sensitivity. *Gastroenterology* 148, 1195–1204 (2015).

Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791 (1985).

Finnie, C., Melchior, S., Roepstorff, P., Svensson, B. Proteome analysis of grain filling and seed maturation in barley. *Plant Physiol*;129: 1308–1319 (2002).

Fontanini, D., Capocchi, A., Muccilli, V., Saviozzi, F., Cunsolo, V., Saletti, R., Foti, S., Gallechi, L. Dimeric inhibitors of human salivary α -amylase from emmer (*Triticum dicoccon* Schrank) seeds. *J. Agric. Food Chem.* 55, 10452–10460 (2007).

Fontanini, D., Capocchi, A., Saviozzi, F. & Gallechi, L. Simplified electrophoretic assay for human salivary α -amylase inhibitor detection in cereal seed flours. *J. Agric. Food Chem.* 55, 4334–4339 (2007).

Franco, O. L., Rigden, D.J., Melo, F.R., Bloch Jr, C., Silva C.P. and Grossi de Sa, M.F. Activity of wheat α -amylase inhibitors towards bruchid α -amylases and structural explanation of observed specificities. *Eur. J. Biochem.* 267, 2166–2173 (2000).

García-Maroto, F., Carbonero, P. & García-Olmedo, F. Site-directed mutagenesis and expression in *Escherichia coli* of WMAI-1, a wheat monomeric inhibitor of insect α -amylase. *Plant Mol. Biol.* 17, 1005–1011 (1991).

Gasteiger, E. The proteomics protocols handbook. *Biochem.* 71, 696–696 (2006).

Geisslitz, S., Ludwig, C., Scherf, K. A. & Koehler, P. Targeted LC-MS/MS Reveals Similar Contents of α -Amylase/Trypsin-Inhibitors as Putative Triggers of Nonceliac Gluten Sensitivity in All Wheat Species except Einkorn. *J. Agric. Food Chem.* 66, 12395–12403 (2018).

Gélinas, P. & Gagnon, F. Inhibitory activity towards human α -amylase in wheat flour and gluten. *Int. J. Food Sci. Technol.* 53, 467–474 (2018).

Gélinas, P., McKinnon, C. & Gagnon, F. Inhibitory activity towards human α -amylase in cereal foods. *LWT – Food Sci Technol* 93, 268–273 (2018).

Gómez, L., Martin, E., Hernandez, D. et al. Members of the alpha-amylase inhibitors family from wheat endosperm are major allergens associated with baker's asthma. *FEBS Lett* 261, 85–88 (1990).

Gómez, L., Sanchez-Monge, R., García-Olmedo, F. and Salcedo, G. Wheat tetrameric inhibitors of insect alpha-amylase: allopoloid heterosis at the molecular level. *Proceedings of the National Academy of Sciences USA.* 86, 3242-3246 (1989).

Gregorini, A., Colomba, M., Julia Ellis, H. & Ciclitira, P. J. Immunogenicity characterization of two ancient wheat α -gliadin peptides related to coeliac disease. *Nutrients* 1, 276–290 (2009).

Guruprasad, K., Reddy, B.V., Pandit, M.W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng. Dec;* 4(2):155-61 (1990).

Horton, H. R., Moran, L. A., Scrimgeour, K. G., Perry, M. D., & Rawn, J. D. *Principles of Biochemistry* (4th ed.). upper Saddle River, NJ, USA: Pearson Prentice Hall (2006).

Huebener, S., Tanaka, C.K., Uhde, M., Zone, J.Z., Vensel, W.H., Kasarda, D.D., Beams, L., Briani, C., Green, P.H.R., Altenbach, S.B. and Alaedini A. Specific non-gluten proteins of wheat are novel target antigens in celiac disease humoral response. *J. Proteome Res.* 14, 503–511 (2015).

Jambu, M. *Exploratory and multivariate data analysis. Statistical Modeling and Decision Science.* San Diego, CA,: Academic Press. 432 p. (1991).

Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292: 195-202 (1999).

Juhász, A., Belova, T., Florides, C.G., Maulis, C., Fischer, I., Gell, G., Birinyi, Z., Ong, J., Keeble-Gagnère, G., Maharajan, A., Ma, W., Gibson, P., Jia, J., Lang, D., Mayer, K.F.X., Spannagl, M., International Wheat Genome Sequencing Consortium, Tye-Din, J.A., Appels, R., Olsen, O. Genome mapping of seed-borne allergens and immune-responsive proteins in wheat. *Sci. Adv.* 4, eaar8602 (2018).

Junker, Y., Zeissig, S., Kim, S.J., Barisani, D., Wieser, H., Leffer, D.A., Zevallos, V., Libermann, T.A., Dillon, S., Freitag, T.L., Kelly, C.P. and Schuppan, D. Wheat amylase trypsin inhibitors drive intestinal inflammation via activation of toll-like receptor 4. *J. Exp. Med.* 209, 2395–2408 (2012).

Kawamori, R., Tajima, N., Iwamoto, Y., Kashiwagi, A., Shimamoto, K. & Kaku, K. Voglibose for prevention of type 2 diabetes mellitus: a randomised, double-blind trial in Japanese individuals with impaired glucose tolerance. *The Lancet*, 373, 1607–1614 (2009).

Kodama, T., Miyazaki, T., Kitamura, I., Suzuki, Y., Namba, Y., Sakurai, J., Torikai, Y. and Inoue, S. Effects of single and long-term administration of wheat albumin on blood glucose control: randomized controlled clinical trials. *Eur J Clin Nutri*, 59, 384–392 (2005).

Kostekli, M. & Karakaya, S. Protease inhibitors in various flours and breads: Effect of fermentation, baking and in vitro digestion on trypsin and chymotrypsin inhibitory activities. *Food Chem.* 224, 62–68 (2017).

Kucek, L. K., Veenstra, L. D., Amnuaycheewa, P. & Sorrells, M. E. A grounded guide to gluten how modern genotypes and processing impact wheat sensitivity. *Compr. Rev. Food Sci. Food Saf.* 14, 285–302 (2015).

Kumar, S., Stecher, G., and Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870-1874. (2016).

Larré, C., Lupi, R., Gombaud, G., Brossard, C., Branlard, G., Moneret-Vautrin, D.A., Rogniaux, H., Denery-Papini, S. Assessment of allergenicity of diploid and hexaploid wheat genotypes: Identification of allergens in the albumin/globulin fraction. *J. Proteomics* 74, 1279–1289 (2011).

Liu, Y. & Wang, J. Molecular Evolution of Exogenous Alpha-amylase Inhibitors in Triticeae - An Update. (2012).

Madeira, F., Park, Y., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn R.D. and Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641 (2019).

Mikola, M., & Mikkonen, A. Occurrence and stabilities of oat trypsin and chymotrypsin inhibitors. *J Cereal Sci*, 30, 227–235 (1999).

Molberg, O., Uhlen, A.K., Jensen, T., Flaete, N.S., Fleckenstein, B., Arentz-Hansen, H., Raki, M., Lundin, K.E.A. & Sollid, L.M. Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. *Gastroenterology*, 128, 393-401 (2005).

Mossor, G., & Skupin, J. Some biochemical properties of trypsin inhibitor type antinutrients derived from extracts of wheat grain, Beta variety. *Die Nahrung*, 29, 491–500 (1985).

Nagy-Gasztonyi, M., Nagy, A., Németh-Szerdahelyi, E., Pauk, J. & Gelencsér, É. The activities of amylases and α -amylase inhibitor in wide-range herbicide resistant wheat lines. *Czech J. Food Sci.* 28, 217–224 (2010).

Ninomiya, K., Ina, S., Hamada, A., Yamaguchi, Y., Akao, M., Shinmachi, F., Kumagai, H. and Kumagai, H. Suppressive effect of the α -amylase inhibitor albumin from buckwheat (*Fagopyrum esculentum* moench) on postprandial hyperglycaemia. *Nutrients* 10, (2018).

Oda, Y., Matsunaga, T., Fukuyama, K., Miyazaki, T. & Morimoto, T. Tertiary and quaternary structures of 0.19 α -amylase inhibitor from wheat kernel determined by X-ray analysis at 2.06 Å resolution. *Biochemistry* 36, 13503–13511 (1997).

Oneda, H., Lee, S. & Inouye, K. Inhibitory Effect of 0.19 α -Amylase Inhibitor from Wheat Kernel on the Activity of Porcine Pancreas α -Amylase and Its Thermal Stability. *J. Biochem.* 135, 421–427 (2004).

Pandey, B., Saini, M. & Sharma, P. Molecular phylogenetic and sequence variation analysis of dimeric α -amylase inhibitor genes in wheat and its wild relative species. *Plant Gene* 6, 48–58 (2016).

Pastaria International 6/2015 pag. 24-36

Payan, F. Structural basis for the inhibition of mammalian and insect α -amylases by plant protein inhibitors. *Biochim. Biophys. Acta - Proteins Proteomics* 1696, 171–180 (2004).

Piasecka-Kwiatkowska, D., Madaj, D., Warchalewski, J.R. The biological activity of wheat, rye and triticale varieties harvested in four consecutive years. *Acta Sci. Pol., Technol. Aliment.* 6(4), 55-65 (2007).

Piasecka-Kwiatkowska, D., Warchalewski, J. R., Zielińska-Dawidziak, M. & Michalak, M. Digestive enzyme inhibitors from grains as potential components of nutraceuticals. *J. Nutr. Sci. Vitaminol. (Tokyo)*. 58, 217–220 (2012).

Poerio, C., Caporale, L., Carrano, P., Pucci, V., Buonocore, Assignment of the five disulfide bridges in an alpha-amylase inhibitor from wheat kernel by fast-atom-bombardment mass spectrometry and Edman degradation, *Eur. J. Biochem.* 3 595–600 (1991).

Poutanen, K., Shepherd, R., Shewry, P.R., Delcour, J.A., Björck, I., Van Der Kamp, J.W. Beyond whole grain: The European HEALTHGRAIN project aims at healthier cereal foods. *Cereal Food World* 53:32–35 (2008).

Prandi, B., Faccini, A., Tedeschi, T., Galaverna, G. & Sforza, S. LC/MS analysis of proteolytic peptides in wheat extracts for determining the content of the allergen amylase/trypsin inhibitor CM3: Influence of growing area and variety. *Food Chem.* 140, 141–146 (2013).

Priya, S., Kumar, S., Kaur, N. & Gupta, A. K. Specificity of α -amylase and trypsin inhibitor proteins in wheat against insect pests. *New Zeal. J. Crop Hortic. Sci.* 41, 49–56 (2013).

Purahong, W., Alkadri, D., Nipoti, P., Pisi, A., Lemmens, M., Prodi, A. Validation of a modified Petri-dish test to quantify aggressiveness of *Fusarium graminearum* in durum wheat. *Eur J Plant Pathol*;132:381–91 (2012).

Rajaram, S., Pena, R. J., Villareal, R. L., Mujeeb-Kazi, A., Singh, R., & Gilchrist, L. Utilization of wild and cultivated emmer and of diploid wheat relatives in breeding. *Isr J Plant Sci.* 49, S93–S104 (2001).

Reig-Otero, Y., Mañes, J. & Manyes, L. Amylase-Trypsin Inhibitors in Wheat and Other Cereals as Potential Activators of the Effects of Nonceliac Gluten Sensitivity. *J. Med. Food* 21, 207–214 (2018).

Rinat Islamov, A. Two independent reactive sites in the bifunctional inhibitor of alpha-amylase and trypsin? *Struct. Biosynth. Funct.* 141–146 (2012).

Rogniaux, H., Pavlovic, M., Lupi, R., Lollier, V., Joint, M., Mameri, H., Denery S. and Larré, C. Allergen relative abundance in several wheat varieties as revealed via a targeted quantitative approach using MS. *Proteomics* 15, 1736–1745 (2015).

Rost, B. & Sander, C. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*. May;19(1):55-72 (1994).

Saitou, N. and Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425 (1987).

Salcedo, G., Sánchez-Monge, R., García-Casado, G., Armentia, A., Gomez L. And Barber, D. The Cereal α -Amylase/Trypsin Inhibitor Family Associated with Bakers' asthma and Food Allergy. *Plant Food Allergens* 70–86 (2004).

Salcedo, G., Quirce, S. & Diaz-Perales, A. Wheat allergens associated with Baker's asthma. *J. Investig. Allergol. Clin. Immunol.* 21, 81–92 (2011).

Sanchez de la Hoz, P. S., Castagnaro, A. & Carbonero, P. Sharp divergence between wheat and barley at loci encoding novel members of the trypsin/ α -amylase inhibitors family. *Plant Mol. Biol.* 26, 1231–1236 (1994).

Sánchez-Monge, R., Gómez, L., García-Olmedo, F. and Salcedo, G. A tetrameric inhibitor of insect α -amylase from barley. *FEBS Letters* 207, 105-109 (1986).

Sanchez-Monge, R., Garcia-Casado, G., Malpica, J. M., & Salcedo, G. Inhibitory activities against heterologous alpha-amylases and in vitro allergenic reactivity of Einkorn wheats. *Theor Appl Genet.* 93, 745–750 (1996).

Sharma, P., Shankar, P.R., Subramaniam, G., Kumar, A., Tandon, A., Suresh, C.G., Rele M.V. and Kumar L.S. Cloning and Sequence Analysis of the Amylase Gene from the Rice Pest Scirpophaga incertulas Walker and its Inhibitor from Wheat (Variety MP Sehore). *Int. J. Insect Sci.* 1, IJIS.S3384 (2009).

Shi, A. N., Leath, S., & Murphy, J. P. A major gene for powdery mildew resistance transferred to common wheat from wild einkorn wheat. *Phytopathology*. 88, 144–147 (1998).

Silano, V., Furia, M., Gianfreda, L., Macri, A., Palessandolo, R., Rab, A., Scardi, V., Stella, E., Valfre, F. Inhibition of amylases from different origins by albumins from the wheat kernel, *Biochim. Biophys. Acta* 391 170 – 178 (1975).

Singh, A., Gaur, A. K. & Mishra, D. P. Molecular cloning and characterization of CM3 gene, from *T. Aestivum* and *T. Durum* genome. *Int. J. Sci. Res. Publ.* 2, 1–6 (2012).

Šotkovský, P., Sklenar, J., Halada, P., Cinova, J., Setinova, I., Kainarova, A., Golias, J., Pavlaskova, K., Honzova, S. and Tuckova, L. A new approach to the isolation and characterization of wheat flour allergens. *Clin. Exp. Allergy* 41, 1031–1043 (2011).

Tabachnick, B. and Fidell, L. *Using Multivariate Statistics*. Harper & Row Publishers New York. (2013).

Taddei, F., Gazza, L., Conti, S., Muccilli, V., Foti, S., Pogna, N.E. Starch-bound 2S proteins and kernel texture in einkorn, *Triticum monococcum ssp monococcum*. *Theor. Appl. Genet.* 119, 1205–1212 (2009).

Tatham, A. S. & Shewry, P. R. Allergens to wheat and related cereals. *Clin. Exp. Allergy* 38, 1712–1726 (2008).

Van den Broeck, H.C., de Jong, H.C., Salentijn, Em.M.J., Dekking, L., Bosch, D., Hamer, R.J., Gilissen, J.W.J., van der Meer, I.M. & Smulders, M.J.M. Presence of celiac disease epitopes in modern and old hexaploid wheat varieties: wheat breeding may have contributed to increased prevalence of celiac disease. *Theor. Appl. Genet.*, 121, 1527-1539 (2010).

Vittozzi, L., & Silano, V. The phylogenesis of protein alpha-amylase inhibitors from wheat seed and the speciation of polyploid wheats. *Theor Appl Genet.* 48, 279–284 (1976).

Walsh, B.J. & Howden, M.E.H. A method for the detection of IgE binding sequences of allergens based on a modification of epitope mapping. *J Immunol Methods* 121, 275–280 (1989).

Wang, J. R., Deng, M., Liu, Y.X., Qiao, X., Chen, Z.H., Jiang, Q.T., Pu, Z.E., Wei, Y.M., Nevo, E., Zheng, Y.L. Adaptive polymorphism of tetrameric alpha-amylase inhibitors in wild emmer wheat. *Genes and Genomics* 33, 357–364 (2011).

Wang, J. R., Yan, Z.H., Wei, Y.M., Nevo, E., Baum, B.R., Zheng, Y.L. Molecular characterization of dimeric alpha-amylase inhibitor genes in wheat and development of genome allele-specific primers for the genes located on chromosome 3BS and 3DS. *J. Cereal Sci.* 43, 360–368 (2006).

Wang, J. R., Wei, Y.M., Long, Y.X., Yan, Z.H., Nevo, E., Baum, B.R., Zheng, Y.L. Molecular evolution of dimeric α -amylase inhibitor genes in wild emmer wheat and its ecological association. *BMC Evol. Biol.* 8, (2008).

Wang, J. R., Pu, Z.E., Lan, Y.J., Baum, B.R., Yan, Z.H., Zheng, Y.L., Wei, Y.M. Phylogenetic analysis of the dimeric alpha-amylase inhibitor sequences from an orthologous region in 21 different genomes of the tribe Triticeae (Poaceae). *Biochem. Syst. Ecol.* 38, 708–714 (2010).

Wang, J. R., Zhang, L., Wei, Y.M., Yan, Z.H., Baum, B.R., Nevo, E., Zheng, Y.L. Sequence polymorphisms and relationships of dimeric α -amylase inhibitor genes in the B genomes of *Triticum* and S genomes of *Aegilops*. *Plant Sci.* 173, 1–11 (2007).

Wang, J. R., Wei, Y.M., Deng, M., Nevo, E., Yan, Z.H., Zheng, Y.L. The impact of single nucleotide polymorphism in monomeric alpha-amylase inhibitor genes from wild emmer wheat, primarily from Israel and Golan. *BMC Evol. Biol.* 10, (2010).

Wang, J. R., Wei, Y. M., Yan, Z. H. & Zheng, Y. L. Detection of single nucleotide polymorphisms in 24 kDa dimeric α -amylase inhibitors from cultivated wheat and its diploid putative progenitors. *Biochim. Biophys. Acta - Gen. Subj.* 1723, 309–320 (2005).

Wang, J. R., Wei, Y. M., Yan, Z. H. & Zheng, Y. L. Genetic mapping of the wheat dimeric α -amylase inhibitor multi-gene family using allele-specific primers based on intergenomic SNPs. *Plant Mol. Biol. Report.* 24, 287–294 (2006).

Wang, J. R., Wei, Y. M., Yan, Z. H. & Zheng, Y. L. Sequence variations and haplotype identification of wheat dimeric α -amylase inhibitor genes in einkorn wheats. *Biochem. Genet.* 45, 803–814 (2007).

Wang, J. R., Wei, Y. M., Yan, Z. H. & Zheng, Y. L. SNP and haplotype identification of the wheat monomeric α -amylase inhibitor genes. *Genetica* 134, 277–285 (2008).

Warchalewski, J.R., Madaj, D., Skupin, J. The varietal differences in some biological activities of proteins extracted from flours of wheat seeds harvested in 1986. *Nahrung/Food* 33, 9, 805-821 (1989).

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191 (2009).

Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Honigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N. and Rost, B. PredictProtein - An open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* 42, 337–343 (2014).

Yang, F., Jørgensen, A.D., Li, H., Søndergaard, I., Finnie, C., Svensson, B., Jiang, D., Wollenweber, B., and Jacobsen, S. Implications of high-temperature events and water deficits on protein profiles in wheat (*Triticum aestivum* L. cv. Vinjett) grain. *Proteomics* 11, 1684–1695 (2011).

Zapatero, L., Martínez, M.I., Alonso, E., Salcedo, G., Sánchez-Monge, R., Barber, D., Lombardero, M. Oral wheat flour anaphylaxis related to wheat alpha-amylase inhibitor subunits CM3 and CM16. *Allergy*, 58, 956 (2003).

Zevallos, V. F., Raker, V., Tenzer, S., Jimenez-Calvente, C., Ashfaq-Khan, M., Rüssel, N., Pickert, G., Schild, H., Steinbrink, K. and Schuppan, D. Nutritional Wheat Amylase-Trypsin Inhibitors Promote Intestinal Inflammation via Activation of Myeloid Cells. *Gastroenterology* 152, 1100–1113.e12 (2017).

Zoccatelli, G., Segal, M., Bolla, M., Cecconi, D., Vaccino, P., Rizzi, C., Chignola, R., Brandolini, A. Expression of α -amylase inhibitors in diploid *Triticum* species. *Food Chem.* 135, 2643–2649 (2012).

Zou, W., Schulze, B.L., Tan, X., Sissons, M., Warren, F.J., Gidley, M.J., Gilbert, R.G.
The role of thermostable proteinaceous α -amylase inhibitors in slowing starch digestion in pasta. *Food Hydrocoll.* 90, 241–247 (2019).

Zuckerkindl, E. and Pauling, L. Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York (1965).