Proteome analysis of diploid, tetraploid and hexaploid wheat: Towards understanding genome interaction in protein expression

Hexaploid wheat (Triticum aestivum L.) is derived from a complex hybridization procedure involving three diploid species carrying the A, B and D genomes. The proteome patterns of diploid, tetraploid and hexaploid wheat were analyzed to explore the genome interaction in protein expression. At least two species from each of the diploid and tetraploid were used to compare their proteome maps with a hexaploid wheat cv. Chinese Spring. The ancestral cultivars were selected based on their history of closeness with the cultivated wheat. Proteins were extracted from seed flour and separated by two-dimensional electrophoresis (2-DE) with isoelectric focusing of pH range from 4–10. 2-DE maps of cultivated and ancestral species were analyzed by computer assisted image analyzer. The region of high molecular weight glutenin subunits of hexaploid wheat showed similarity with those of the diploid donors, BB and DD genomes. The omega gliadin, which is controlled by B genome in common wheat, was assumed to have evolved as a result of interaction between AA and BB genomes. The low molecular weight glutenins and alpha and beta gliadin regions were contributed by the three genomes. This result suggests that the function of donor genomes particularly in the expression of proteins in hexaploid wheat is not totally independent; rather it is the product of interactions among the diploid genomes in the hexaploid nuclear constitutions. The expression of nonstorage proteins was affected substantially due to the removal of the D genome from hexaploid constitution. Location of the structural gene controlling one of the alpha amylase inhibitor proteins in the nonstorage protein region was identified in the short arm of chromosome 3D.

Keywords: Diploid / Genome / Hexaploid / Interaction / Seed proteome / Tetraploid / Wheat

1 Introduction

The seed proteins of common wheat (Triticum aestivum L.) are responsible for the ability of flour to form cohesive dough required to make good quality products such as breads, biscuits and cakes. Because of this uniqueness, breeding strategies for quality wheat protein in the last few years have focused on the correlation between storage proteins of wheat and its phylogenetics [1–4]. Bread wheat is a hexaploid species with three diploid genomes denoted by A, B and D and is composed of seven pairs of chromosomes in each genome. Kihara and his coworkers [5] first established the phylogenetics of common wheat, mainly on the basis of chromosome pairing affinity observed in F₁ hybrids. Later, the genomic relationship and phylogeny in wheat was established by restriction fragment length polymorphism analysis of nuclear DNA [6, 7]. Since then, the stability of new hybrid cultivars in breeding programs for protein quality is generally predicted based on phylogenetic information such as genome, gene, nuclear and cytoplasmic constitution. These have opened up a new avenue on genetic information such as nucleotide sequences, genetic maps, DNA markers and cytoplasmic interactions of the genome donors. But this information does not necessarily match quantitatively or qualitatively at the protein level, the frontier of the central dogma which regulates the complex biological function in an organism. Biological influences including the stability, half-life, post-transcriptional, cotranslational and degradative modifications of proteins along with the environmental factors affect the gene products [8]. This has led to a conclusion that there is no strict linear relationship between genes or genome information and protein complement of a cell [9].
Recent developments in proteomic techniques such as 2-DE, image analyses and mass spectrometry have provided a unique opportunity to look at the protein complement of the whole genome within a very short time. 2-DE, which analyzes complex mixtures of denatured proteins based on two independent criteria, charge (pI) and molecular weight (M_r), has become a powerful technique of choice for detecting and quantifying protein in plant proteome research including cereals. Using the 2-DE technique, Colas de Francs and Thiellement [2] investigated the chromosomal arm localization of structural genes controlling leaf-proteins of wheat cv. Chinese Spring (CS) by superimposing the dried gels of euploid and a series of ditelocentric lines (DT) which carry all the normal chromosome complement of wheat chromosomes except for one chromosome for which one arm is missing. Islam et al. [4, 10] reported the quantitative variations of wheat proteome due to the deletion of chromosomal arms. Although these studies have provided general information relating to chromosomal gene location and protein expression in proteome, the genome interaction in protein expression, a key factor considered to play a vital role in determining the protein function as a whole, has not been addressed yet. A clear understanding of the network of genome interactions, hence functionally interacting proteins, will provide valuable information to the cereal chemist/plant breeder, which will assist them in designing their research work precisely. To our knowledge, this is the first study to identify the genome interactions of the expression of proteins by proteome analysis of diploid, tetraploid and hexaploid wheat.

2 Materials and methods

2.1 Plant materials

Seven diploid species, known to be potentially involved in the evolutionary process that gave rise to bread wheat, were selected for proteome analyses (Table 1).

Table 1. List of different diploid, tetraploid and hexaploid wheat species used for proteome analysis

<table>
<thead>
<tr>
<th>Species and variety</th>
<th>Genome formula</th>
<th>Comments and references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triticum urartu</td>
<td>AA</td>
<td>Wild-type and considered as one of the very close ancestors of common wheat [11–13]</td>
</tr>
<tr>
<td>var. nigrum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triticum boeoticum</td>
<td>AA</td>
<td>Wild-type [13]</td>
</tr>
<tr>
<td>Triticum monococcum</td>
<td>AA</td>
<td>Only cultivated species known to be A genome donor [13]</td>
</tr>
<tr>
<td>Aegilops squarrosa</td>
<td>DD</td>
<td>Recommended as a D donor by McFadden and Sears [14] and experimentally demonstrated by Kihara [15]</td>
</tr>
<tr>
<td>var. strangulata</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aegilops speltoides</td>
<td>SS (BB)</td>
<td>Considered as B genome progenitor of common wheat by Duad and Gustafson [16]; Kerby and Kuspira [17]; Talbert et al. [18]</td>
</tr>
<tr>
<td>var. typica</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aegilops longissimum</td>
<td>S'S'</td>
<td>Recommended as a donor of B genome by Vitozzi and Silano [19]; Tsunewaki and Oghihara [20]</td>
</tr>
<tr>
<td>Aegilops bicorne</td>
<td>S'S'</td>
<td>Suggested as a potential donor of B genome by Sears [21]</td>
</tr>
<tr>
<td>Triticum dicocoides</td>
<td>AABB</td>
<td>Wild-type wheat belongs to Emmer group [22]</td>
</tr>
<tr>
<td>var. kotoschyanum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>var. hokudai &amp; khalpi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>AABB</td>
<td>Developed by successive backcrosses (eg BC_{10}) of hybrid between hexa- and tetra-ploids and tetraploids were then selected from the self pollinated progenies of pentaploid [23–25]</td>
</tr>
<tr>
<td>var. tetra CS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cv. CS</td>
<td>AABBDD</td>
<td>Standard variety of common wheat widely used in studies for evolution and genetics</td>
</tr>
</tbody>
</table>
2.2 2-DE

Endosperms were dissected from seeds and ground to obtain flour. A portion of the flour (10 mg) was digested for 1 h with 300 μL of lysis buffer containing 8 M urea, 2% v/v NP-40, 0.8% v/v ampholine, pH 3.5–10, 5% v/v 2-mercaptoethanol and 5% w/v polyvinylpyrrolidone-40. Proteins from wheat endosperm flour (10 mg) were also extracted with 200 μL of Tris-HCl buffer (10 mM, pH 7.5) containing 0.1% SDS. After digestion, the samples were centrifuged at 15 000 × g for 10 min. To remove the non-protein contamination from the crude extract, proteins were subjected to acetone precipitation with 80% acetone cooled down to −30°C for more than 2 h. The pellet collected after centrifugation at 15 000 × g for 10 min was digested with 50 μL of lysis buffer. After centrifugation at 15 000 × g for 10 min, the supernatant was applied to IEF rod gels (pH 4–10). Sample solutions (40 μL) were loaded onto the acidic side of the IEF gels for the first dimension, and anodic and cathodic electrode solutions were filled in the upper and lower electrode chambers, respectively [26]. SDS-PAGE in the second dimension was performed with 17% separation and 5% stacking gels. Protein spots in 2-DE gels were visualized by Coomassie Brilliant Blue (CBB) R-250 staining [27]. Each sample was run at least two times.

2.3 Quantitative analysis of electrophoresis patterns

Following staining, the 2-DE gel maps were scanned using a flatbed scanner, and analyzed by PDQuest software V 6.2 (Bio-Rad Discovery Series, Bio-Rad, Hercules, CA, USA). After scanning, spots in gels were detected using the same parameters and quantified by 2-D Gaussian modeling. Taking full complement of genes in CS as standard, we made a matchset to compare protein spots.

2.4 Peptide mass fingerprinting

Proteins separated by 2-DE were digested in gels according to the method described by Hellman et al. [28]. The digests were desalted with ZipTips (Millipore, Boston, MA, USA) and subjected to nano ESI Q-TOF MS analysis for determination of peptide mass sequence.

3 Results and discussion

3.1 Protein extraction and proteome

Extraction of wheat seed proteins is known to be difficult due to two reasons: the presence of less ionizable side chains in the major class of wheat protein (glutenin) and the presence of wide range proteins, from very low abundant functional proteins to very high abundant proteins. A combination of urea-mercaptoethanol and NP-40 is a popular choice of solvent used for extraction of wheat seed proteins. In our study, we found that the urea-based solvent is suitable for the extraction of high molecular weight storage proteins but not suitable for low abundant functional proteins. The improved extraction of high molecular weight (HMW) wheat protein by urea-based solvent could be explained by the fact that the urea has similar molecules as the polypeptide backbone of the glutenin protein resulting in very small enthalpy [29]. Alternatively, in our study, good extraction of soluble and low abundant functional proteins was achieved by using 0.1% SDS (Fig. 1), a negatively charged detergent reported to be useful for the extraction of membrane proteins. In our study, we used lysis buffer (composition in Section 2.2) extracted proteins to compare 2-DE proteome maps of di-, tetra- and hexaploid wheat and 0.1% SDS in Tris-HCl extracted proteins to compare nonstorage functional proteins expression in hexa- and tetra-CS. To avoid any systematic error from the electrophoresis and staining procedures, we ran two rod gels in one large size SDS-PAGE using a recently introduced electrophoresis apparatus (Fig. 1).

The mature wheat grains contains several types of proteins including the storage proteins, gluten, and non-storage proteins which include enzymes like alpha amylase inhibitors and structural proteins such as those in membranes. The gluten can be classified into gliadin, a monomeric protein with intramolecular disulfide bonds, and glutenin, a polymeric protein with intra- and intermolecular disulfide bonds. Based on lactate-PAGE (Acid PAGE), gliadin proteins are classified into alpha, beta, gamma and omega gliadins [30]. The glutenin proteins are again classified into two groups based on their mobility in an electric field, HMW and low molecular weight (LMW) glutenins. Upon reduction, the glutenin dissociates into several subunits, denoted as subunit 1, subunit 2 and so on. Payne et al. [31] first attempted to divide the wheat endosperm proteome, the total proteins expressed by the genome, into several areas based on different types of protein expressed. Recently, Skylas et al. [32] have divided the wheat endosperm proteome into three regions; glutenin protein regions 1, glutenin protein region 2 and nonglutenin protein region 3. To compare the proteome maps of diploid, tetraploid and hexaploid wheat in detail, we divided the 2-DE map into four regions; HMW glutenin region, omega gliadin regions, LMW glutenin and alpha, beta and gamma region and nonstorage protein region (Fig. 2).
3.2 Proteome analysis of diploid, tetraploid and hexaploid wheat

Since the discovery of seven chromosomes in the haploid wheat cell by Sakamura [33] and subsequent demonstration of inheritance of D donor in common wheat by Kihara [15], the phylogenetic study of hexaploid wheat which involves complex processes of interactions among the diploid donors has become an area of interest for many cytogenetists, molecular biologists and conventional plant breeders. A number of techniques including genome analysis by chromosome pairing [34] and RFLP analysis of nuclear DNA [35] have been used to study wheat phylogenetic interaction of diploid species. These methods are
useful to understand wheat phylogeny in general. However, the complex processes of genome interactions in protein expression among the donor species, which could play a vital role in the study of wheat evolution, has not been explored yet. In order to understand this complex process of genome interaction, we analyzed the seed proteome of diploid, tetraploid and hexaploid species of wheat using proteomic techniques. Our results showed that the function of donor genomes, particularly in the expression of proteins in hexaploid is not totally independent, rather it is the product of interactions among the diploid genomes in the hexaploid nuclear constitutions.

We used two to three species from each of the diploid and tetraploid wheat for proteome analysis (Table 1 and Fig. 3). These species were selected based on their phylogenetic relationship as revealed by cytogenetical and taxonomical observations (see references in Table 1). The endosperm proteome of CS, a common variety of hexaploid wheat, was used as a template to compare the proteome maps of diploid and tetraploid species. This variety has been extensively used for phylogeny, cytogenetical and proteomic studies in the last few decades [4, 35, 36]. Protein from all the samples was extracted by lysis buffer and separated using 2-DE. After scanning, the tissue maps were compared taking the outline of the CS 2-DE map as standard (Fig. 3).

The HMW glutenin subunit-2, which is controlled by the D genome, was missing in all the A genome progenitors used in this study, but a close match was found in each of the three species of *Aegilops*. On the other hand, HMW glutenin subunit-12, which is also controlled by the D genome in hexaploid, was found only in *Ae. squarrosa* (DD). Based on this result, it can be assumed that HMW glutenin subunit-2 in hexaploid could be inherited from any of the species of *Ae. longissimum*, *Ae. bicorn*, and that of subunit-12 could be from *Ae. squarrosa*. Therefore, *Ae. squarrosa* can be assumed as one of the potential sources of HMW glutenin subunits controlled by the D genome in hexaploid wheat. Using morphological features and genome analysis of eight species of *Aegilops*, Kihara [15] found a perfect match between the *Ae. squarrosa* and hexaploid wheat. The HMW glutenin subunits-7 and -8, which are controlled by the B genome in hexaploid species, show a very close match to *longissimum*, *bicorn* and *speltoids* when compared by 1-D SDS-PAGE (data not shown), but substantial variations in pI of these subunits were observed when compared with the 2-DE maps of hexaploid wheat. It could be possible that these subunits originally came from the diploid species, but the pI has been changed in the polyploid nuclear constitution of hexaploid wheat. It is also possible that we did not include the diploid species of the B genome, which are closely related to inheritance of these subunits. However, the positional polymorphism (pI changes or horizontal shift) in the inheritance of polypeptides is reported in maize, barley and pea [37], in beans [38], in mice [39] and in humans [40].

The omega gliadin in hexaploid wheat consists of four to six spots and is controlled by genes located on the short arm of chromosome 1B [4, 10, 31]. These proteins are extensively used for classification of wheat varieties [41]. While comparing the region of omega gliadin in 2-DE maps, we found that neither the species AA nor the BB showed this protein in their proteome. Interestingly, this protein was present in the proteome maps of all the tetra-

---

**Figure 3.** Comparison of 2-DE maps of diploid, tetraploid and hexaploid wheat species. Diploid and tetraploid species are known to be potentially involved in the evolutionary process of hexaploid species.
poloids used in this study. Two different reasons can be proposed to account for this. (1) The true diploid donor of gliadin was not included among the species we had chosen in this study. Therefore, to decipher the paradigm of B genome donors of hexaploid wheat, it is better to include a much larger number of species. (2) This protein evolved in the context of the interaction between A and B donors in the processes of polyploidization which could have undergone some thousands years of genetic interaction to come to the present conditions. To compare the different diploid ancestors in hexaploid wheat, it has been a common interest of geneticists to estimate to what extent gene expression is affected when several genomes are associated in the polyploid nuclear constitution. While studying gene expression in the early stages of a newly synthesized wheat allotetraploid, Kaskush et al. [42] reported that 60 out of 3072 transcripts were reproducibly altered in the allotetraploid due to gene loss and gene silencing. Therefore, gene interaction in hexaploid wheat is not a new phenomenon, but for the first time, we found that this interaction is not only confined to the gene/mRNA level but also to the protein level. It is likely that the gliadin, which is currently controlled by the B genome in hexaploid wheat, was an initial product of interaction between A and B genomes.

The inheritance of LMW glutenin and alpha, beta and gamma gliadins is complicated and is contributed by A, B and D genomes. The mixed contribution of the protein expression in this region is also apparent from our previous studies. The inheritance of the nonstorage protein region is also not very clear except quadruplets, which showed close similarity with all the Aegilops species.

3.3 Genome interaction and protein expression in tetra CS

To understand the genome interaction in protein expression, we also analyzed proteome maps of tetra CS (AABB) by comparing them with that of hexaploid species. The tetra CS is developed by successive backcrosses of hybrid between hexa- and tetraploid, and the extracted tetraploid were then selected from the self pollinated progenies of pentaploid [23–25]. The tetra CS used in our study is different from those reported by Kaltsikes et al. [24, 25]. In our materials, the tetraploid was extracted by crossing hexaploid with the tetra Thatcher (AABB) component derived from common wheat cultivars (Thatcher), instead of natural durum wheat, and then ten successive backcrosses (BC10) were performed, compared with three to seven backcrosses reported by Kaltsikes et al. [24, 25]. Therefore, the genome compliments of tetra CS in the present experiment are almost similar to the AABB background of hexaploid CS.

As shown in Fig. 4, the HMW subunits 2 and 12, which are controlled by the D genome in hexaploid wheat, did not appear in proteome maps of tetra CS (indicated by arrow in 2-DE maps). These results are in agreement with the findings reported by Payne et al. [31] and Islam et al. [4, 10]. Surprisingly, two new spots, 3 and 6, in addition to the four spots of omega gliadin appeared in this region. These spots have a pI similar to the gliadin, but their molecular weight is slightly less than that of gliadin. To determine whether they are similar to gliadin or not, peptide sequence of spot 6 were determined by nano ESI Q-TOF MS after digestion of the protein with trypsin. Selected peptides were subjected for MS/MS analyses and the sequence information were interrogated by MassLynx (V3.5). Results revealed that the peptide fragments 866.46 showed homology with gliadin proteins (Fig. 5), indicating that these spots are very similar in their structural configuration with omega gliadin. But, the presence of these spots only in tetra CS has raised a question of its origin: where does it come from? Theoretically, it is not possible that these spots were initially inherited from Thatcher, one of the parents used to produce tetra CS, and retained in the tetra CS even after 10th generation of backcrossing. However, to confirm this assumption practically, we analyzed the proteome map of Thatcher, but could not ob-

![Figure 4. 2-DE maps of hexa- and tetraploid CS. Endosperm proteins were extracted by urea based lysis buffer and applied in IEF rod gels (pH 4–10) and run on large size SDS-PAGE.](image)
serve any such spots in the proteome map (Fig. 6), suggesting that the new spots did not come from the var. Thatcher. Therefore the appearance of the new spots may have three other explanations. (1) As the spots are structurally similar to omega gliadin, the genes controlling this protein are also located in the B genomes, but their expression is regulated by a suppressor located in the D genome. (2) It could be a product of genome interaction between AA and BB genomes, which remain suppressed in the presence of the D genome. (3) These could also be the co- or post-translational modified products of gliadin, and the presence of D genome prevents such modification of the protein. The actual mechanism behind the appearance of the new spots is unclear. But what is clear here is that interactions among the genomes in the polyploid wheat dictate the expression of these proteins in wheat seed proteome. Although these proteins, spots 3 and 6 appeared in tetra CS, the natural tetraploid did not show these spots, raising an issue of difference between these two types of tetraploid. It is possible that, if our assumption of genome interaction is true, these proteins initially appeared in the natural tetraploid and then subsequently disappeared over the course of time for the same reason of genomic interactions.

Figure 6. 2-DE maps of Thatcher and Thatcher + CS. Endosperm proteins were extracted by lysis buffer applied in IEF rod gels (pH 4–10), and then run on large size SDS-PAGE (17%). In Thatcher + CS, flour from Thatcher and CS was mixed at 1:1 ratio, and followed the same procedure as mentioned above.
The nonstorage protein region of tetra- and hexaploid CS also showed a substantial difference in protein expression (Fig. 4). This region contains mainly functional proteins including the heat shock proteins (smHSP) and alpha amylase inhibitors. One spot of quadruplets, which are known as alpha amylase inhibitor, disappeared in tetra CS, suggesting that the gene controlling this protein spot is located in the D genome. Using ditelocentric lines of the D genome, we confirmed the location of this gene on the short arm of chromosome 3D (data not shown). To our knowledge, this is the first report of identifying the gene location of alpha amylase inhibitor. Importantly, the deletion of the D genome has seriously affected protein expression, particularly in the nonstorage protein regions, which was clearly revealed when proteins were extracted with 0.1% SDS (Figs. 1 and 7). To estimate quantitative variations of protein expression, particularly in the nonstorage protein region, we analyzed the spot intensities of 0.1% SDS extracted protein by PDQuest (V 6.2, Bio-Rad). As shown in Fig. 7, out of the 19 spots detected, 12 spots were up-regulated and only two down-regulated. The substantial variation of protein intensities due to the deletion of the D genome, hence genes, indicate that the magnitude of protein expression in a proteome depends on the context of its interactions with other proteins under a set of conditions. Although this phenomenon contradicts the traditional concept of protein function where a protein function is defined as its action in the presence of a substrate and a catalyst, it is quite supportive to the post-genomic view of protein expression, where a protein function is defined by its interaction with all other functionally interacting proteins [43]. The imbalance of the whole network of protein expression as revealed from our study by up- and down-regulation due to genome deletion, could also result from some of the functional linkages in the metabolic or signaling pathways responsible for the protein expression. This result is also consistent with the current theory of action resonance, which provides an interactive wave-guide for the mutual evolution of all organisms occurring within each physicochemical environment [44].

4 Concluding remarks

The proteome analysis of diploid, tetraploid and hexaploid wheat opens up new possibilities to look at genome interactions in protein expression, compared to the conventional genome analysis and RFLP methods. Taking an example of omega gliadin, we have demonstrated that the function of donor genomes, particularly in the expression of proteins, in hexaploid wheat is not totally independent; rather it is the product of interactions among the diploid genomes in the hexaploid nuclear constitutions. Gene location of four nonstorage proteins identified in the D genome can be used as a marker for the analyses of functional proteins such as heat shock protein. Substantial up-regulation of nonstorage proteins caused by the removal of the D genome might open up a new avenue to study protein-protein interactions in hexaploid wheat.

We thank the Japanese Society for promotion of Science, for their financial assistance in this research and a postdoctoral fellowship for Dr. Nazrul Islam, a nominee of the Australian Research Council and The Australian Academy of Science. We also thank Prof. Yoshihiko Furuta of Gifu University, Japan, for providing seeds of tetra Chinese Spring and Mr. C. Shang and Y. Ogura for helping in peptide analysis by ESI Q-TOF MS. Part of the research was also supported by the Ministry of Agricul-

---

**Figure 7.** Comparison of protein expression in the nonstorage protein region between hexa and tetra CS using matchset (A). A synthetic image created in PDQuest used as reference map to compare protein spots between CS and tetra CS (C). The intensity of protein expression is shown in bar graph (B), left bar represents protein intensity in hexaploid CS and right bar represents the same spot in tetraploid CS.
5 References