

Variability and cluster analysis of arabinoxylan content and its molecular profile in crossed wheat lines

Marietta Szentmiklóssy^a, Kitti Török^a, Éva Pusztai^b, Sándor Kemény^b,
Karolina Tremmel-Bede^c, Marianna Rakszegi^c, Sándor Tömösközi^{a,*}

^a Research Group of Cereal Science and Food Quality, Department of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, Műegyetem rkp.3, 1111, Budapest, Hungary

^b Industrial Statistics Research Group, Department of Chemical and Environmental Process Engineering, Budapest University of Technology and Economics, Műegyetem rkp.3, 1111, Budapest, Hungary

^c Agricultural Institute, Centre for Agricultural Research, Brunszvik u. 2. Pf. 19, 2462, Martonvásár, Hungary

ARTICLE INFO

Keywords:

Crossed wheat lines
Variability
Arabinoxylan
Molecular weight distribution
Cluster analysis

ABSTRACT

Non-starch polysaccharides (NSP) in cereals, after starch and gluten proteins, determine the technological and nutritional properties of flour, dough and end product. From NSPs, the arabinoxylans are the most important components studied in wheat nowadays. The novelty of our study is the investigation of the variability and clustering ability of the wheat lines based on different quantitative and qualitative traits of arabinoxylans. The quantitative properties, the total and the water-extractable arabinoxylan content of wheat flours were measured separately by gas chromatography. The qualitative property referring to the structure of the molecule was the molecular weight distribution measured by size exclusion liquid chromatography. According to the results the variability of arabinoxylan properties in the breeding lines varied between wider ranges than the values measured for the parental varieties. The correlation between the quantitative parameters, and a trend between quantitative and qualitative parameters were described. During the cluster analysis, the parents were well separated into different groups. The parameters that played an important role in the clustering were the ratio of the probability of occurrence of molecular size ranges along with the quantitative traits of AX. The selected samples can be used for further targeted breeding while the methodology was used in this work can be suitable for selecting lines/varieties for special food or other industrial applications.

1. Introduction

The favourable chemical composition of wheat flour contributes to its unique technological properties and versatile usability. The crude protein content makes up 8–16% of the grain, and 80–90% of these are the storage proteins, which gluten proteins are responsible for the dough extension and viscoelastic properties. Relationships between the composition and quality of use are known as per previous researches (Békés et al., 2017; Lafandra et al., 2014; Thanhaeuser et al., 2014; Wieser and Kieffer, 2001). Furthermore, there is a wealth of knowledge

about the technological and nutritional properties of starch (Hou et al., 2020; Yan et al., 2016). More research investigates the technological properties of high amylose, waxy and regular amylopectin/amylose flours (Caramanico et al., 2017; McCann et al., 2018). However, less information is available on the variability and technological properties of non-starch polysaccharides (NSP). They occur mainly in the outer layers of the seed but can be also found in the endosperm (Lafandra et al., 2014). The NSP components, commonly with lignins, form the group of dietary fibres. The beneficial physiological effects of fibres are to initiate bowel movement, regulate nutrient absorption and residence

Abbreviations: H-MW, high molecular weight; L-MW, low molecular weight; R, ratio of high and low molecular ranges; SE-HPLC, size exclusion chromatography; TOTAX, total arabinoxylan content; TOT-A/X, arabinose to xylose ratio in total arabinoxylan content; WEAX, water-extractable arabinoxylan content; WE-A/X, arabinose to xylose ratio in water-extractable arabinoxylan content.

* Corresponding author.

E-mail addresses: szentmiklossy.marietta@mail.bme.hu (M. Szentmiklóssy), torokkitti@gmail.com (K. Török), pusztaie@mail.bme.hu (É. Pusztai), kemeny@mail.bme.hu (S. Kemény), bede.karolina@agr.ar.mta.hu (K. Tremmel-Bede), rakszegi.marianna@agr.ar.mta.hu (M. Rakszegi), tomoskozi@mail.bme.hu, tomoskozi@mail.bme.hu (S. Tömösközi).

<https://doi.org/10.1016/j.jcs.2020.103074>

Received 20 April 2020; Received in revised form 5 August 2020; Accepted 6 August 2020

Available online 16 August 2020

0733-5210/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

time. They can bind and thereby slow down the absorption of glucose. The most prominent NSP components in cereals are arabinoxylans, β -glucans, arabinogalactan peptides, fructans, glucomannans, galactomannans and cellulose (Andersson et al., 2013; Gebruers et al., 2008; Mendis and Simsek, 2014).

Arabinoxylans (AX) are heteropolymers of β -D-xylopyranose derivatives linked by β -1,4 glycosidic linkages, which can be substituted with α -L-arabinofuranose residues at C(O)-2 and/or C(O)-3 positions. This degree of substitution is characterized by the arabinose/xylose ratio (A/X). The mono-substituted α - (1,3) arabinofuranose residues can be ester linked with ferulic acid groups. Total arabinoxylan content (TOTAX) consists of water-extractable (WEAX) and water-unextractable arabinoxylan (WUAX) fractions, which differ in the A/X ratios and polymer chain lengths resulting differences in their biological function as well. While WUAX has a low degree of substitution and high molecular weight, WEAX characterized by higher degrees of substitution and lower MW. Most of the WUAXs occur in the outer layers of the seed and with cellulose, they are responsible for the physical protection of the grain cells. WEAXs, which are present in both the endosperm and the aleurone layers, provide physicochemical protection of the seed tissues/cells (Ordaz-Ortiz and Saulnier, 2005; Saulnier et al., 2007).

One of the unique technological property of the arabinoxylans is the high water binding capacity and thus the ability to form a highly hydrated dough matrix. The larger molecular weight WEAX are technologically advantageous, because they are able to stabilize the structure of gluten during dough formation, resulting in a larger, homogeneous and porous bread crumb structure. According to our present knowledge, water-unextractable arabinoxylans are technologically unfavourable, since they destabilize the gluten structure, reduce the rising of dough and consequently decrease bread volume (Bonnand-Ducasse et al., 2010). However, these effects depend highly on the source, composition, size and structure of the AX molecules. Another feature of arabinoxylans is that under oxidative conditions and at acidic pH, ferulic acid side chains can form cross-links. This process is called as oxidative gelling process (Németh et al., 2019; Saulnier et al., 2007).

The arabinoxylan content and properties of cereals are determined by both the genotype (G) and the environmental conditions (E). Several studies have already discussed the differences of AX characteristics of spring and winter wheat varieties and durum wheat (Finnie et al., 2006; Török et al., 2019; Vignola et al., 2016). These genetic experiments have shown that durum species had lower variability in all properties of AX than spring and winter bread wheat. In the latter ones, the WEAX content of the endosperm and the WE-A/X ratio of the bran showed the highest variation. The most intensively studied environmental effects are the temperature, the precipitation and the soil quality. Ciccioritti et al. (2011) observed a positive correlation between the precipitation and the TOTAX and WEAX content in wholemeal durum flour grown at two sites in two consecutive harvest years. The activity of endoxylanase enzymes may also play a role in the reduction of WUAX content, thereby increases the relative proportion of WEAX (Ciccioritti et al., 2011). Finnie et al. (2006) investigated 7 spring and 20 winter wheat varieties under 12 environmental conditions and concluded that the WEAX content primarily depends on the genotype, while the WUAX content depends more on the environment (Finnie et al., 2006). In general, it was established that the effects of environmental and genotype factors (GxE) are less significant, than their individual effects (Ciccioritti et al., 2011; Shewry et al., 2010; Vignola et al., 2016). Altogether, the literature does not agree on that certain AX traits are more influenced by the genotype or the environment.

The main goal of the whole project was to increase the dietary fibre content in wheat lines by traditional breeding, as nutritionally favourable, fibre-rich end products could be produced from the flour of these special wheat genotypes. In the breeding program run for several years, the effect of the genotype and the environment has already been investigated before in a three years field experiment including 27 breeding lines and 14 varieties (Török et al., 2019). In this study our aim

was the investigation of the variability of different AX properties, parameters (TOTAX, WEAX, substitution ratio, molecular size distribution) in crossed wheat lines and compare these parameters to the parental varieties in a given vintage. Furthermore, a partly new statistical evaluation method was developed for grouping the samples based on their arabinoxylan properties using statistical tools. This study and the developed method may help to understand the background of the variation of genetically determined arabinoxylan properties. Later it could help us to clarify the E effects and the GxE interactions, also.

2. Materials and methods

2.1. Sample collection

Forty-eight *Triticum aestivum* crossbreed lines were developed at the Agricultural Institute, Centre for Agricultural Research, (Martonvásár, Hungary) and harvested in 2017. Two parental lines were previously studied in the Healthgrain EU-FP6 project from which MV-Toborzó was a Hungarian, high-yielding, early maturing, winter wheat variety, with high fibre content while Tommi was a German wheat variety with late maturity and high phytochemical content (sterol, tocol, folate). The two varieties are very different for day/night sensitivity and dwarfing properties also, resulting high variation in the phenotypic traits of the plants in the progeny. White wheat flours (<250 μ m) were produced by a CD1 type laboratory mill (Chopin, Villeneuve-la-Garenne, France) according to the NF EN ISO 27971 standard.

The lines were grown at the Agricultural Institute, Centre for Agricultural Research, Martonvásár (latitude, 47_210N; longitude, 18_490E; altitude, 150 m) in 2017. The plots were 2.5 m long with six rows spaced at 20 cm. The soil was a chernozem with loam texture and pH 7.25. The previous crops were oilseed radish. The plots were treated with herbicide (4 L/ha U-46 D-fluid SL containing 500 g/L 2-methyl-4-chlorophenoxyacetic acid; 40 g/ha Granstar 50 SX containing 50% tribenuron methyl), insecticide (0.2 L/ha Karate Zeon 5CS containing 50 g/L k-cihalotrin), and fungicide (first: 1 L/ha Amistar Extra containing 200 g/L azoxistrobin and 80 g/L ciprokonazol, second: 1 L/ha Cherokee containing 50 g/L ciprokonazol, 62 g/L propiconazol and 375 g/L clor-etalonil). The growing conditions of the experiment are summarised in [Supplementary 1 Table 1](#), showing that 2016/2017 had a very cold winter with minimum temperature below minus 20° and the quantity of the precipitation was low during the last 100 days of the plant development.

2.2. Arabinoxylan quantification

The total and the water-extractable AX (TOTAX and WEAX) content of the wheat flours were measured separately by gas chromatography (GC). Sample preparation was carried out according to Gebruers et al. (2008) with some modifications. Briefly, for WEAX analysis samples were extracted in cold water and the supernatant was used for the next steps. All samples were hydrolysed using trifluoroacetic acid (TFA), then reduced and acetylated. The alditol acetates were then analysed on a Clarus 500 GC instrument (PerkinElmer, Waltham MA, USA) equipped with an autosampler and splitter injection port (split ratio 1:8). The alditols were separated on an Elite-17 (PerkinElmer, Waltham MA, USA) polar column (60 m, 0.25 mm, i.d.; 0.52 mm film thickness) and detected with a flame ionisation detector. Helium was the carrier gas. Separation was performed at 250 °C, and injection and detection were carried out at 300 °C. AX content (%) was calculated as 0.88 times the sum of xylose and arabinose and was expressed on a dry matter basis. A/X ratio was calculated as the ratio of arabinose to xylose without correction for arabinogalactan peptide, as such correction would lead to errors for wheat samples (Gebruers et al., 2008).

2.3. Size distribution of arabinoxylans

Liquid chromatography method was carried out based on Loosveld et al. (1997), with some modification. 1000 mg of wheat flour samples were solubilised in 5 ml 0.3% (w/v) NaCl solution and kept at 4 °C for 60 min (vortexed in every 20 min). After centrifugation at 4000g for 5 min, 0.25 ml of the supernatant was pipetted into Eppendorf tube and 1 ml 0.3% (w/v) NaCl solution was added. To eliminate starch, 10 µl thermostable α -amylase were added to the samples, and they were thermostated at 98 °C at 300 rpm for 30 min. The samples were centrifuged at 4000 g for 5 min and the supernatants were filtered through a 0.45 µm RC (Regenerated Cellulose) filter. The samples were measured by a Flexar high-performance liquid chromatograph (PerkinElmer Inc), equipped with an autosampler, a column oven (38 °C), a RI detector (38 °C) and the modules were controlled by Chromera. A Shodex, OHpak 10 µm SB-804 HQ 300 mm × 8 mm, 200 Å gel filtration column (Phenomenex Inc, Torrance, California, USA) were used to analyse the molecular weight (MW) of water-extractable arabinoxylans. The samples were thermostated at 4 °C, the mobile phase was 0.3% (w/v) NaCl in ultrapure water (Merck Millipore, Merck KGaA, Darmstadt, Germany.), the flow rate was 0.3 ml/min and injection volume was 30 µl. Shodex Standard P-82 pullulans with an MW range of 5.900–708.000 was used as Molecular weight (MW) standard-line. The size distribution can be expressed as the ratio of the two areas of the water-extractable arabinoxylans. The high molecular weight ranges have always been presented in higher proportions than smaller ones, so the ratio of the area of the first peak to the second peak will be discussed below (R) (Fig. 2).

2.4. Statistical analysis

The analytical results were statistically evaluated by one-way ANOVA and correlation analysis (Supplement Table 1) using TIBCO Statistica 13.4.0 (TIBCO Software Inc., Palo Alto, USA) and RStudio 1.2.13.35 (RStudio Inc., Boston, USA) software. Significant differences were studied using t-tests. On every level of each factor, two replicates were measured. The homogeneity of variance and the normality of distribution were checked in all examined cases. The hierarchical (Ward's method and Euclidean distances) and K-means cluster analysis were performed for grouping the average values of all samples (Daultbakovbakytkan and Galym, 2019; Ibrahim et al., 2016; Ordaz-Ortiz

et al., 2005).

3. Result and discussion

3.1. Arabinoxylan content

The white flour of MV-Toborzó had 2.24% total arabinoxylan content, which is comparable with previously reported data (1.35–2.75%) (Shewry et al., 2013; Stone and Morell, 2009; Vignola et al., 2016). At the same time, Tommi had a 1.90% average TOTAX content. The water-extractable arabinoxylan content of MV-Toborzó was 0.95%, which corresponds to the upper third of the literature values (0.30–1.40%), while Tommi took a moderately low (0.63%) WEAX content. In the literature, TOT-A/X ranges from 0.5 to 0.7, and WE-A/X from 0.4 to 0.5 (Gebruers et al., 2008; Messia et al., 2017; Ordaz-Ortiz et al., 2005). MV-Toborzó had an average total arabinoxylan A/X value (0.62), while Tommi had higher TOT-A/X (0.73) value than as reported in the literature. The water-extractable A/X value of MV-Toborzó was the same (0.61), but in case of Tommi, it was almost twice as much (0.90) as the literature upper limit. Consequently, in terms of A/X values, Tommi was the parent that possessed higher values.

The TOTAX (1.51–2.69%) and WEAX (0.44–1.21%) results of the studied lines did not exceed the range reported in the literature, but there were smaller and larger values from both parents. More than one-third of the lines had slightly higher TOTAX values than MV-Toborzó (2.24%), although, the difference was not significant in any cases. One-sixth of the lines had smaller values than Tommi (1.90%). The variability of the TOT-A/X (0.60–1.17) and WE-A/X (0.59–0.97) values found in the lines were increased compared to the parental values. Some of the crossing lines had one and a half times or even higher than the literary maximum (Gebruers et al., 2008; Messia et al., 2017; Ordaz-Ortiz et al., 2005; Török et al., 2019).

The relationship between the TOTAX and WEAX content, and their A/X ratio changed similarly and more-or-less in the same direction. While there was no clear relationship between TOTAX and TOT-A/X, in contrast, the relationship of WEAX and WE-A/X values gave a significant negative correlation (−0.790), which means the higher the WEAX content of a line, the smaller its substitution. With decreasing WEAX-ranked results (Fig. 1), the two parents are located at the two border of the sample population. The MV-Toborzó was closer to the higher end values, while Tommi at situated at the lower end. It can be stated that in

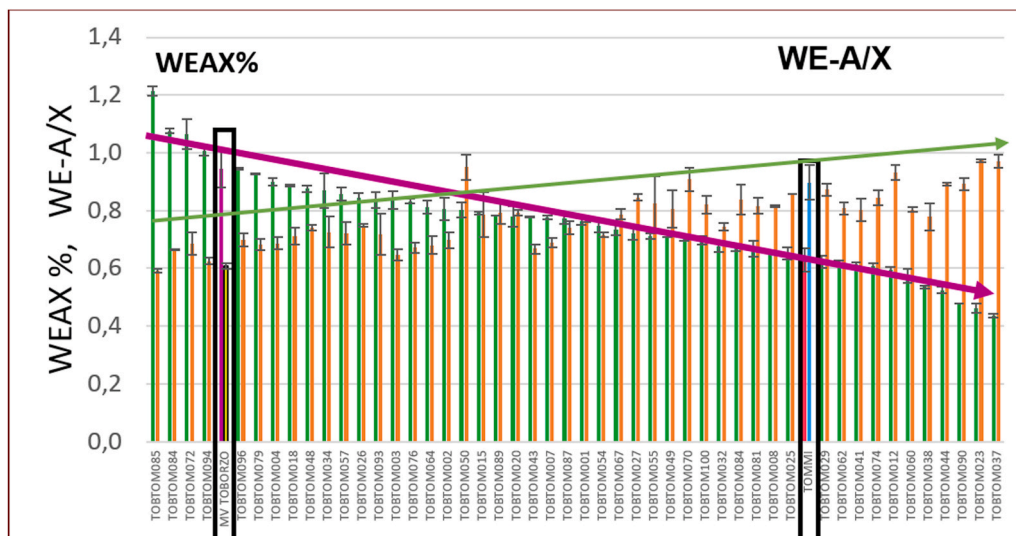


Fig. 1. Relationship between water-extractable arabinoxylans and their arabinose/xylose ratio; WEAX content (green) and WE-A/X ratio (red), of the parents are highlighted in different colours. Error bars represent \pm scatter of the mean (the lines had two technical repeats, the parental varieties had two independent replicate experiments with two technical repeats). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

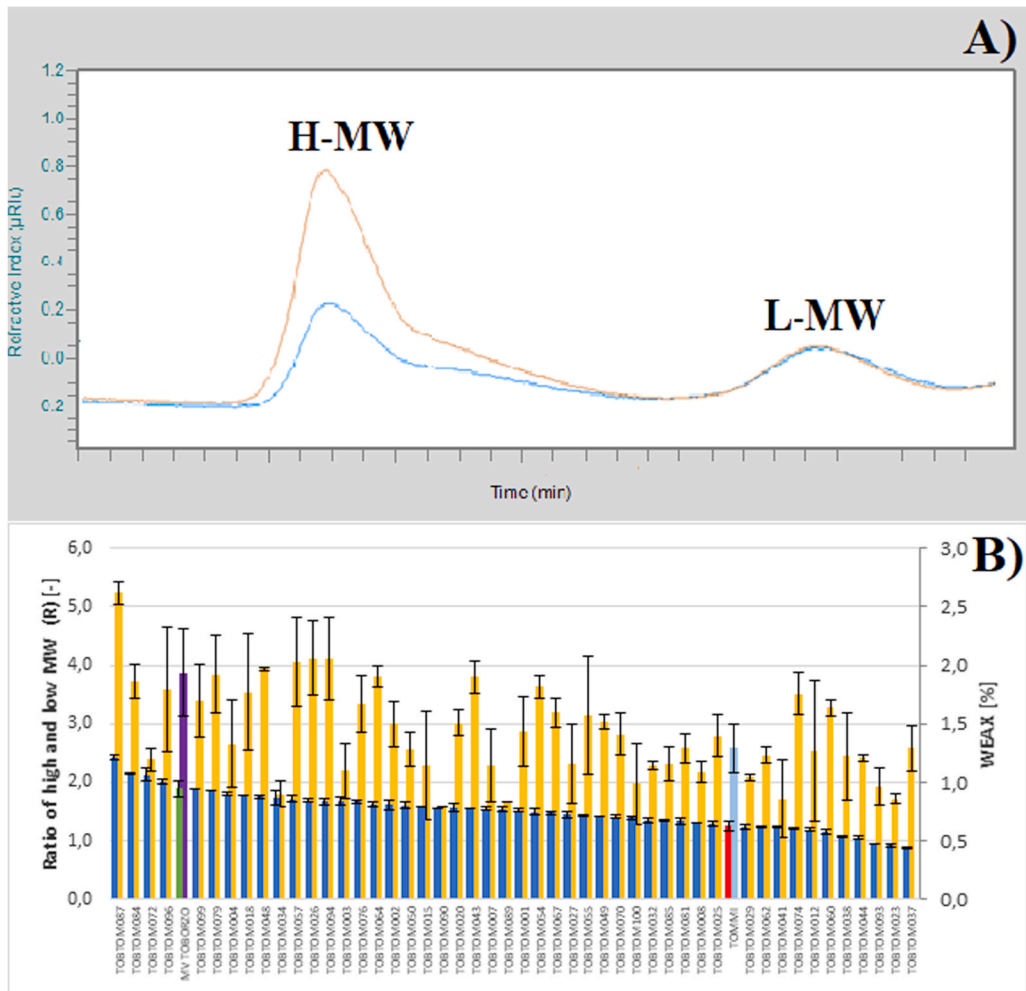


Fig. 2. A) Two breeding lines with different water-extractable arabinoxylan molecular weight profile; in the size exclusion chromatogram the first peak is the high molecular weight (H-MW), the second peak is the low molecular weight (L-MW) water-extractable arabinoxylan B) WEAX content (blue) and the ratio of high and low MW WEAX (R) (yellow), the parental species highlighted in different colours. Error bars represent \pm scatter of the mean (the lines had two technical repeats, the parental varieties had two independent replicate experiments with two technical repeats). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
K-means clustered by Statistica and RStudio.^a

	#Clusters	Name	AX properties
All parameters, K-means to 6 clusters in Statistica	1.	Tommi, TOBTOM49, -50, -62, -70, -74, -81, -100	low TOTAX, high WE-A/X
	2.	TOBTOM1, -2, -3, -4, -7, -15, -27, -89, -	low TOT-A/X
	3.	TOBTOM90, -93	high TOTAX, high TOT-A/X
	4.	MV-Toborzó, TOBTOM18, -57, -64, -72, -76, -79, -85, -87, -94, -96, -99	high WEAX low WE-A/X
	5.	TOBTOM26, -43, -48, -54, -55, -67, -84	high TOTAX
	6.	TOBTOM8, -12, -20, -23, -25, -29, -32, -34, -37, -38, -41, -44, -60	low WEAX high WE-A/X
All parameters, K-means to 6 clusters in RStudio	1.	same as „All parameters, K-means to 6 clusters in Statistica” 2. cluster	
	2.	MV-Toborzó, TOBTOM57, -64, -72, -76, -79, -85, -87, -94, -96, -99	high WEAX, low WE-A/X
	3.	TOBTOM60, -62, -67 -70, -74, -84, -90	shorter L-MW
	4.	Tommi, TOBTOM8, -12, -20, -23, -25, -29, -32, -34, -37, -38, -41, -44	low WEAX, high WE-A/X
	5.	TOBTOM18, -26, -43, -48, -54, -55	high TOTAX
	6.	TOBTOM49, -50, -82, -93, -100	high WE-A/X
Five parameters, K-means to 6 clusters in RStudio	1.	Tommi, TOBTOM8, -12, -20, -23, -25, -29, -32, -34, -37, -38, -41, -44, -50, -62	low TOTAX, low WEAX, high WE-A/X
	2.	same as „All parameters, K-means to 6 clusters in Statistica” 2. cluster	
	3.	same as „All parameters, K-means to 6 clusters in RStudio” 5. cluster	
	4.	same as „All parameters, K-means to 6 clusters in RStudio” 2. cluster	
	5.	TOBTOM49, -60, -67 -70, -74, -81, -84	low WEAX, shorter L-MW
	6.	TOBTOM90, -93 (same as „All parameters, K-means to 6 clusters in Statistica” 3. cluster)	high TOTAX, high TOT-A/X
All parameters K-means to 3 clusters in RStudio	1.	TOBTOM1, -2, -3, -4, -7, -15, -72, 85, -89, -90, -99	similar to Tommi but low TOTAX, high TOT-A/X
	2.	MV-Toborzó, TOBTOM18, -26, -43, -48, -54, -57, -64, -67, 87, -76, -79, -94, -96	similar to MV-Toborzó
	3.	Tommi, TOBTOM8, -20, -23, -27, -29, 32, -34, -37, -38, -41, -44, -49, -50, -55, -62, -84, -93, -100	similar to Tommi
Five parameters, K-means to 3 clusters in RStudio	1.	Tommi, TOBTOM8, -12, -20, -23, -25, -27, -29, -32, -34, -37, -38, -41, -44, -49, -50, -55, -60, -62, -70, -74, -81, -84, -100	similar to Tommi
	2.	TOBTOM90, -93 (same as „All parameters, K-means to 6 clusters in Statistica” 3. cluster)	high TOTAX, high TOT-A/X
	3.	MV-Toborzó, TOBTOM1, -2, -3, -4, -7, -15, -18, -26, -43, -48, -54, -57, -62, -64, -67, -76, -79, -85, -87, -89, -94, -96, -99	similar to MV-Toborzó

^a “All parameters” are included (H-MW) high molecular weight; (L-MW) low molecular weight; (R) ratio of high and low molecular ranges; (TOTAX) total arabinoxylan content; (TOT-A/X) arabinose to xylose ratio in total arabinoxylan content; (WEAX) water-extractable arabinoxylan content, (WE-A/X) arabinose to xylose ratio in water-extractable arabinoxylan content, “Five parameters” are not included H-MW, and L-MW.

lines, but they did not differ significantly from the parents as the standard deviation was high. It is worth examining the evolution of the R together with the WEAX content to consider if there is a correlation between the evolution of quantitative and qualitative parameters (Fig. 2). The R values were primarily determined by the probability of occurrence of the H-MW WEAX, since the variability of the L-MW WEAX area was minimal. Fig. 2 shows that lines with higher WEAX content also had higher R, showing a trend between the two parameters. Connecting this phenomenon with the previously observed WEAX - WE-A/X reverse trend, it can be concluded that the lines with higher L-MW WEAX have lower WEAX contents and higher WE-A/X ratios. This can promote technological properties by improving dough stabilization through cross-linking. Technological studies are necessary to verify if this indirect information is valid in practice. The molecular size of WEAX is certainly technologically important as they influence the retrogradation of starch. L-MW retard the recrystallization of amylose and high-MW suppress the recrystallization of amylopectin (Hou et al., 2020; Li et al., 2013).

3.3. Cluster analysis

3.3.1. Hierarchical clustering

In the cluster analysis, the parents and the breeding lines grouped by taking into consideration of the quantitative values of arabinoxylan (i. e. TOTAX, TOT-A/X, WEAX and WE-A/X) and the measured quality parameters, like molecular size ranges and rates. Based on the smallest distances and plot of linkage distances across steps, six groups were identified in hierarchical clustering (Fig. 3). Examination of the properties of the groups revealed that R was one of the most important property for grouping, followed by quantitative properties (WEAX, TOTAX) and the molecular size ranges were the least meaningful (Table 1).

Clusters 1 to 3 separated well from clusters 4 to 6 (Table 1). The two

parents located in the two furthest clusters from each other (in cluster 3 and cluster 6), showing the great genetic distances and high variation in their properties for successful crossing. The cluster 1 was the smallest, with three members (TOBTOM67, -90, -93). They were characterized by their low ratio of arabinoxylans' size ranges. They had low WEAX content and moderate WE-A/X. On the other hand, they had high TOTAX, TOT-A/X and molecular size values. The cluster 2 and 3 had the nearest-neighbour linkages. The nine members of cluster 2 characterized by high WE-A/X and low molecular size. The cluster 3 consisted of eleven lines and Tommi, as parent. Its main difference from the cluster 2 was the value of R and the molecular sizes, which were characterised by medium values. As a summary, the left branches of the dendrogram (cluster 1–3) can be characterised by low or medium R values. It means that the large and small molecular weight WEAX components are almost equally likely to occur. They also showed high degree of AX substitution and low WEAX content of the Tommi variety.

The first branch on the right part of the dendrogram counts nine members for the cluster 4. The TOT-A/X were always low, but clustering for the other parameters were not observed. The five members of the cluster 5 was characterized by the high R value, high TOTAX and medium-high WEAX values. The degrees of substitution were variable in this group and the large molecular size ranges widened. The last group, cluster 6, reflects the characteristics of MV-Toborzó, the parental variety and eleven breeding lines, having consistently high WEAX and low WE-A/X values. However, the right side of the dendrogram (cluster 4–6) had a low degree of substitution and a high value of R, which means that larger molecular weight WEAX components are more likely to occur than smaller ones.

3.3.2. K-means clustering

Another statistical approach was to perform clustering using a K-means method, so the results by different algorithms can show, whether the lines are comparable or not. Based on the hierarchical clustering

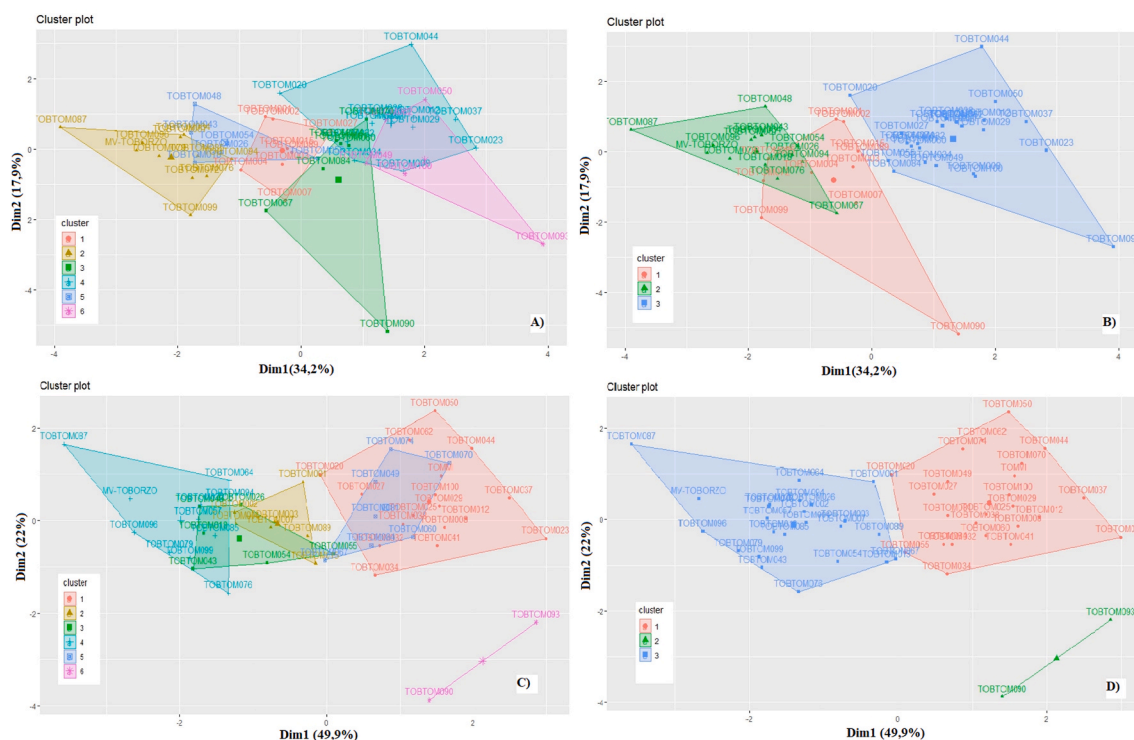


Fig. 4. K-means clustering in R-Studio, A) six groups with all analysed parameters, B) three groups with all parameters, C) six groups with five parameters, D) three groups with five parameters.

number, six clusters were formed (Table 2). Similarly, to the previous results, the two parental varieties were separated into two different clusters. Except for seven lines, they formed groups identical to the hierarchical clustering. Of these, five lines moved from the hierarchical group two to the group three of K-means. The similarity between the groups formed by the two different methods demonstrated the correctness of the hypothesis that it is possible to distinguish and classify the breeding lines by the investigated AX-related properties.

The K-means cluster analysis also proceeded in RStudio, which allows us to visualize the results (Fig. 4, Table 2). Based on the hierarchical clustering of the Statistica program, the starting cluster number was chosen to be six in the RStudio program as well. As shown in Fig. 4A, the groups did not separate, they strongly overlapped and some of the patterns were outliers. In order to avoid overlapping, the number of the clustering groups was then decreased from six to three (Fig. 4B). In this case cluster 1 and 2 still overlaps, and TOBTOM90 and -93 are outliers, as they stayed very far from the group centres.

Due to the high variance of the molecular sizes and its low effect on the variability and the uncertainty of the measurement, these quality parameters were ignored and only the quantitative results with the ratio of the molecular size ranges were considered for further evaluation. Based on the five parameters, using the K-means method formed six groups (Fig. 4C). Five of the clusters overlapped with each other, but the previous outlier TOBTOM90 and -93 lines formed a separate group (cluster 6). By reducing the number of clusters from six to three (Fig. 4D), the sample distribution was a bit unbalanced among the groups as one of them contained 2, while the other two clusters included 24-24 lines. One of the groups included the MV-Toborzó (cluster 3), while another one contained Tommi (cluster 1). This figure (Fig. 4D) shows that the first and third clusters are close to each other, and most of the lines are between the two parents. This is in line with our expectations since it was a crossing between two varieties with known characteristics. The novelty of these results is that the TOBTOM90 and -93 separated completely from parental properties.

In light of these visualized RStudio results, the K-means clustering

performed with three groups in Statistica. The groups obtained through the Data Mining function were the same as the reduced parametric three-cluster members of the R-Studio. So, the assumption that molecular sizes have no meaningful effect on grouping seems to be reinforced. It can be found in some literature that the samples are grouped according to the studied properties, but the statistical evaluation hasn't been performed based on arabinoxylan properties (Dauletbakovbakytkan and Galym, 2019; Ibrahim et al., 2016; Ordaz-Ortiz et al., 2005). Therefore, the novelty of our paper is the statistical methodology we applied for grouping and selecting sample population on the basis of AX properties, it could contain a new approach to evaluating results.

3.3.3. Comparison of the results of different cluster analysis

Comparing the different clustering methods, many similarities can be observed. Some lines always moved together, so they were closer to each other in their properties. One such "group" is TOBTOM1, -2, -3, -4, -7 and -15 with high TOT-A/X and low WE-A/X. The TOBTOM57, -64, -76, -79, -85, -87, -94, -96 and -99 have high WEAX and R values and low WE-A/X, they also similar to MV-Toborzó. Most of the lines, which always moved together, were the TOBTOM8, -12, -20, -23, -27, -29, -32, -34, -37, -38 and -41. They inherited the properties of Tommi, such as low WEAX with high WE-A/X and low R.

4. Conclusion

In this work, the variability and clustering traits of AXs were examined in crossed wheat lines, which were members of a wheat mapping population. The correlation between WEAX and WE-A/X and trends between quantitative and qualitative parameters were examined.

Parental attributes separated well during clustering into two furthest apart groups. The properties of the four other groups examined and deciphered the decisive changes at the branching. It is possible to decode the importance of parameters by examining the properties of the groups of classical hierarchical clustering. At the junctions, the properties of the branches were also examined. In the case of the first branching, the

method distinguished two groups based on the ratio of molecular size ranges, WEAX and TOT-A/X properties. At level two branching, changes in TOTAX and molecular size ranges were affected.

As a summary, it can be concluded that the identified reasons (parameters) of correlations and clustering can be used for selection of lines with different technological and nutritional character. The described methodology can be a useful tool in the selection of lines with targeted quality in breeding programs, also. It can help to select nutritionally and techno-functionally favourable lines during breeding work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research and breeding activities were supported by the projects of OTKA K112169 and OTKA K112179. The work was also connected to the goals of the BME-Biotechnology FIKP grant of EMMI (BME FIKP-BIO).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jcs.2020.103074>.

References

- Andersson, A.A., Andersson, R., Piironen, V., Lampi, A.M., 2013. Contents of dietary fibre components and their relation to associated bioactive components in whole grain wheat samples from the HEALTHGRAIN diversity screen. *Food Chem.* 136, 1243–1248. <https://doi.org/10.1016/j.foodchem.2012.09.074>.
- Békés, F., Schoenlechner, R., Tömösközi, S., 2017. Ancient wheats and pseudocereals for possible use in cereal-grain dietary intolerances. In: *Cereal Grains: Assessing and Managing Quality*, second ed. Elsevier Ltd, pp. 353–389. <https://doi.org/10.1016/B978-0-08-100719-8.00014-0>.
- Bonnand-Ducasse, M., Della Valle, G., Lefebvre, J., Saulnier, L., 2010. Effect of wheat dietary fibres on bread dough development and rheological properties. *J. Cereal. Sci.* 52, 200–206. <https://doi.org/10.1016/j.jcs.2010.05.006>.
- Caramanico, R., Barbiroli, A., Marengo, M., Fessas, D., Bonomi, F., Lucisano, M., Pagani, M.A., Iametti, S., Marti, A., 2017. Interplay between starch and proteins in waxy wheat. *J. Cereal. Sci.* 75, 198–204. <https://doi.org/10.1016/j.jcs.2017.04.008>.
- Ciccoritti, R., Scalfati, G., Cammerata, A., Sgrulletta, D., 2011. Variations in content and extractability of durum wheat (*Triticum turgidum* L. var durum) arabinoxylans associated with genetic and environmental factors. *Int. J. Mol. Sci.* 12, 4536–4549. <https://doi.org/10.3390/ijms12074536>.
- Dauletbaekovbakytkan, D., Galym, D., 2019. Modeling the quality of new wheat varieties in Kazakhstan. *J. Food Ind.* 3, 30. <https://doi.org/10.5296/jfi.v3i1.15912>.
- Finnie, S.M., Bettge, A.D., Morris, C.F., 2006. Influence of cultivar and environment on water-soluble and water-insoluble arabinoxylans in soft wheat. *Cereal Chem.* 83, 617–623. <https://doi.org/10.1094/CC-83-0617>.
- Gebruers, K., Dornez, E., Boros, D., Fras, A., Bedó, Z., Rakszegi, M., 2008. Variation in the content of dietary fiber and components thereof in wheats in the healthgrain diversity screen. *J. Agric. Food Chem.* 56, 9740–9749. <https://doi.org/10.1021/jf800975w>.
- Hou, C., Zhao, X., Tian, M., Zhou, Y., Yang, R., Gu, Z., Wang, P., 2020. Impact of water extractable arabinoxylan with different molecular weight on the gelatinization and retrogradation behavior of wheat starch. *Food Chem.* 318, 126477. <https://doi.org/10.1016/j.foodchem.2020.126477>.
- Ibrahim, O.M., Tawfik Elh, M.M., Badr, A., Wali, A.M., 2016. Evaluating the performance of 16 Egyptian wheat varieties using self-organizing map (SOM) and cluster analysis. *J. Appl. Sci.* 16, 47–53. <https://doi.org/10.3923/jas.2016.47.53>.
- Izydorczyk, M.S., 2009. Arabinoxylans. In: Phillips, G., Williams, P. (Eds.), *Handbook of Hydrocolloids*, second ed. Woodhead Publishing, Woodhead, pp. 653–692. <https://doi.org/10.1533/9781845695873.653>.
- Lafiandra, D., Riccardi, G., Shewry, P.R., 2014. Improving cereal grain carbohydrates for diet and health. *J. Cereal. Sci.* 59, 312–326. <https://doi.org/10.1016/j.jcs.2014.01.001>.
- Li, W., Hu, H., Wang, Q., Brennan, C.S., 2013. Molecular features of wheat endosperm arabinoxylan inclusion in functional bread. *Foods* 2 (2), 225–237. <https://doi.org/10.3390/foods2020225>.
- Loosveld, A.M.A., Grobet, P.J., Delcour, J.A., 1997. Contents and structural features of water-extractable Arabinogalactan in wheat flour fractions. *J. Agric. Food Chem.* 45, 1998–2002. <https://doi.org/10.1021/jf960901k>.
- McCann, T.H., Homer, S.H., Øiseth, S.K., Day, L., Newberry, M., Regina, A., Lundin, L., 2018. High amylose wheat starch increases the resistance to deformation of wheat flour dough. *J. Cereal. Sci.* 79, 440–448. <https://doi.org/10.1016/j.jcs.2017.12.001>.
- Mendis, M., Simsek, S., 2014. Arabinoxylans and human health. *Food Hydrocoll.* 42, 239–243. <https://doi.org/10.1016/j.foodhyd.2013.07.022>.
- Messia, M.C., Candighiotta, T., De Arcangelis, E., Marconi, E., 2017. Arabinoxylans and β -glucans assessment in cereals. *Ital. J. Food Sci.* 29, 112–122.
- Morales-Ortega, A., Carvajal-Millan, E., López-Franco, Y., Rascón-Chu, A., Lizardi-Mendoza, J., Torres-Chavez, P., Campa-Mada, A., 2013. Characterization of water extractable arabinoxylans from a spring wheat flour: rheological properties and microstructure. *Molecules* 18, 8417–8428. <https://doi.org/10.3390/molecules18078417>.
- Németh, R., Bender, D., Jaksics, E., Calicchio, M., Langó, B., D'Amico, S., Török, K., Schoenlechner, R., Tömösközi, S., 2019. Investigation of the effect of pentosan addition and enzyme treatment on the rheological properties of millet flour based model dough systems. *Food Hydrocoll.* 94, 381–390. <https://doi.org/10.1016/j.foodhyd.2019.03.036>.
- Ordaz-Ortiz, J.J., Devaux, M.F., Saulnier, L., 2005. Classification of wheat varieties based on structural features of arabinoxylans as revealed by endoxylanase treatment of flour and grain. *J. Agric. Food Chem.* 53, 8349–8356. <https://doi.org/10.1021/jf050755v>.
- Ordaz-Ortiz, J.J., Saulnier, L., 2005. Structural variability of arabinoxylans from wheat flour. Comparison of water-extractable and xylanase-extractable arabinoxylans. *J. Cereal. Sci.* 42, 119–125. <https://doi.org/10.1016/j.jcs.2005.02.004>.
- Saulnier, L., Sado, P.E., Branlard, G., Charmet, G., Guillon, F., 2007. Wheat arabinoxylans: exploiting variation in amount and composition to develop enhanced varieties. *J. Cereal. Sci.* 46, 261–281. <https://doi.org/10.1016/j.jcs.2007.06.014>.
- Shewry, P.R., Hawkesford, J.M., Piironen, V., Lampi, A.-M., Gebruers, K., Boros, D., 2013. Natural variation in grain composition of wheat and related cereals. *J. Agric. Food Chem.* 61, 8295–8303. <https://doi.org/10.1021/jf3054092>.
- Shewry, P.R., Piironen, V., Lampi, A.M., Edelmann, M., Kariluoto, S., Nurmi, T., Fernandez-Orozco, R., Ravel, C., Charmet, G., Andersson, A.A.M., Åman, P., Boros, D., Gebruers, K., Dornez, E., Courtin, C.M., Delcour, J.A., Rakszegi, M., Bedo, Z., Ward, J.L., 2010. The HEALTHGRAIN wheat diversity screen: effects of genotype and environment on phytochemicals and dietary fiber components. *J. Agric. Food Chem.* 58, 9291–9298. <https://doi.org/10.1021/jf100039b>.
- Stone, B., Morell, M.K., 2009. Carbohydrates. In: Khan, K., Shewry, P.R. (Eds.), *Wheat: Chemistry and Technology*. Elsevier Inc., Minnesota, pp. 299–362. <https://doi.org/10.1016/B978-1-891127-55-7.50016-1>.
- Thanhaeuser, S.M., Wieser, H., Koehler, P., 2014. Correlation of quality parameters with the baking performance of wheat flours. *Cereal Chem.* 91, 333–341. <https://doi.org/10.1094/CCHEM-09-13-0194-CESI>.
- Török, K., Szentmiklóssy, M., Tremmel-Bede, K., Rakszegi, M., Tömösközi, S., 2019. Possibilities and barriers in fibre-targeted breeding: characterisation of arabinoxylans in wheat varieties and their breeding lines. *J. Cereal. Sci.* 86, 117–123. <https://doi.org/10.1016/j.jcs.2019.01.012>.
- Vignola, M.B., Moiraghi, M., Salvucci, E., Baronia, V.T., Pérez, G., 2016. Whole meal and white flour from Argentine wheat genotypes: mineral and arabinoxylan differences. *J. Cereal. Sci.* 71, 217–223. <https://doi.org/10.1016/j.jcs.2016.09.002>.
- Wieser, H., Kieffer, R., 2001. Correlations of the amount of gluten protein types to the technological properties of wheat flours determined on a micro-scale. *J. Cereal. Sci.* 34, 19–27. <https://doi.org/10.1006/jcs.2000.0385>.
- Yan, Z., Qi, G.U.O., Nan, F., Jin-rong, W., Shu-jun, W., Zhong-hu, H.E., 2016. Characterization of A- and B-type starch granules in Chinese wheat cultivars. *J. Integr. Agric.* 15, 2203–2214. [https://doi.org/10.1016/S2095-3119\(15\)61305-3](https://doi.org/10.1016/S2095-3119(15)61305-3).