


Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

## Automatic morphology phenotyping of tetra- and hexaploid wheat spike using computer vision methods


A.Yu. Pronozin<sup>1</sup>, A.A. Paulish<sup>2</sup>, E.A. Zavarzin<sup>2</sup>, A.Yu. Prikhodko<sup>2</sup>, N.M. Prokhoshin<sup>2</sup>, Yu.V. Kruchinina<sup>1,3</sup>, N.P. Goncharov<sup>1,4</sup>, E.G. Komyshev<sup>1,2,3</sup>, M.A. Genaev<sup>1,2,3</sup> 

<sup>1</sup> Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Kurchatov Genomics Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>4</sup> Novosibirsk State Agrarian University, Novosibirsk, Russia

 [mag@bionet.nsc.ru](mailto:mag@bionet.nsc.ru)

**Abstract.** Intraspecific classification of cultivated plants is necessary for the conservation of biological diversity, study of their origin and their phylogeny. The modern cultivated wheat species originated from three wild diploid ancestors as a result of several rounds of genome doubling and are represented by di-, tetra- and hexaploid species. The identification of wheat ploidy level is one of the main stages of their taxonomy. Such classification is possible based on visual analysis of the wheat spike traits. The aim of this study is to investigate the morphological characteristics of spikes for hexa- and tetraploid wheat species based on the method of high-performance phenotyping. Phenotyping of the quantitative characteristics of the spike of 17 wheat species (595 plants, 3348 images), including eight tetraploids (*Triticum aethiopicum*, *T. dicoccoides*, *T. dicoccum*, *T. durum*, *T. militinae*, *T. polonicum*, *T. timopheevii*, and *T. turgidum*) and nine hexaploids (*T. compactum*, *T. aestivum*, i:ANK-23 (near-isogenic line of *T. aestivum* cv. Novosibirskaya 67), *T. antiquorum*, *T. spelta* (including cv. Rother Sommer Kolben), *T. petropavlovskiyi*, *T. yunnanense*, *T. macha*, *T. sphaerococcum*, and *T. vavilovii*), was performed. Wheat spike morphology was described on the basis of nine quantitative traits including shape, size and awns area of the spike. The traits were obtained as a result of image analysis using the WERecognizer program. A cluster analysis of plants according to the characteristics of the spike shape and comparison of their distributions in tetraploid and hexaploid species showed a higher variability of traits in hexaploid species compared to tetraploid ones. At the same time, the species themselves form two clusters in the visual characteristics of the spike. One type is predominantly hexaploid species (with the exception of one tetraploid, *T. dicoccoides*). The other group includes tetraploid ones (with the exception of three hexaploid ones, *T. compactum*, *T. antiquorum*, *T. sphaerococcum*, and i:ANK-23). Thus, it has been shown that the morphological characteristics of spikes for hexaploid and tetraploid wheat species, obtained on the basis of computer analysis of images, include differences, which are further used to develop methods for plant classifications by ploidy level and their species in an automatic mode.

Key words: wheat spike morphology; wheat; phenomics; image processing; computer vision; machine learning; biotechnology.

**For citation:** Pronozin A.Yu., Paulish A.A., Zavarzin E.A., Prikhodko A.Yu., Prokhoshin N.M., Kruchinina Yu.V., Goncharov N.P., Komyshev E.G., Genaev M.A. Automatic morphology phenotyping of tetra- and hexaploid wheat spike using computer vision methods. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2021;25(1):71-81. DOI 10.18699/VJ21.009

## Автоматическое фенотипирование морфологии колоса тетра- и гексаплоидных видов пшеницы методами компьютерного зрения

А.Ю. Пронозин<sup>1</sup>, А.А. Паулиш<sup>2</sup>, Е.А. Заварзин<sup>2</sup>, А.Ю. Приходько<sup>2</sup>, Н.М. Прохошин<sup>2</sup>, Ю.В. Кручинина<sup>1,3</sup>, Н.П. Гончаров<sup>1,4</sup>, Е.Г. Комышев<sup>1,2,3</sup>, М.А. Генаев<sup>1,2,3</sup> 

<sup>1</sup> Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>3</sup> Курчатовский геномный центр Института цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>4</sup> Новосибирский государственный аграрный университет, Новосибирск, Россия

 [mag@bionet.nsc.ru](mailto:mag@bionet.nsc.ru)

**Аннотация.** Внутривидовая классификация культурных растений необходима для эффективного сохранения биологического разнообразия видов, изучения их происхождения, определения филогении и проведения межвидовой гибридизации при селекции. Современные возделываемые виды пшениц произошли

от трех диких диплоидных предков в результате гибридизации и нескольких раундов удвоения геномов и представлены ди-, тетра- и гексаплоидными видами. Поэтому идентификация плоидности пшеницы и определение их геномного состава являются одними из основных этапов их классификации на основе визуального анализа фенотипических признаков колоса. Цель работы – исследование морфологических характеристик колосов полиплоидных видов пшеницы методами высокопроизводительного фенотипирования. Выполнено фенотипирование количественных характеристик колоса 17 видов пшеницы (595 растений, 3348 изображений), включая восемь тетраплоидных: *Triticum aesthiopicum*, *T. dicoccoides*, *T. dicoccum*, *T. durum*, *T. militinae*, *T. polonicum*, *T. timopheevii*, *T. turgidum* и девять гексаплоидных: *T. compactum*, *T. aestivum* (в том числе изогенная линия сорта Новосибирская 67 АНК-23), *T. antiquorum*, *T. spelta* (включая стародавний сорт *T. spelta* Rother Sommer Kolben), *T. petropavlovskiyi*, *T. yunnanense*, *T. macha*, *T. sphaerococcum*, *T. vavilovii*. Морфология колоса описана на основе девяти количественных признаков, включающих форму, размер и остистость. Признаки были получены в результате анализа цифровых изображений с помощью программы WERecognizer. Кластерный анализ растений по характеристикам формы колоса и сравнение их распределений у тетра- и гексаплоидных видов показали более высокую вариабельность признаков у гексаплоидных видов по сравнению с тетраплоидными. При этом сами виды в пространстве характеристик колоса формируют два кластера. К первому относятся преимущественно гексаплоидные виды, за исключением одного тетраплоидного, дикорастущего *T. dicoccoides*, ко второму – тетраплоидные, за исключением трех гексаплоидных, *T. compactum*, *T. antiquorum*, *T. sphaerococcum*, и i:АНК-23. Показано, что морфологические характеристики колосов для гекса- и тетраплоидных видов, полученные на основе компьютерного анализа изображений, демонстрируют различия, которые в дальнейшем могут быть использованы для разработки методики эффективной классификации растений по плоидности и их видовой принадлежности в автоматическом режиме. Ключевые слова: пшеница; морфология колоса; феномика; обработка изображений; компьютерное зрение; машинное обучение; биотехнологии.

## Introduction

A number of important issues, including aspects of the effective conservation of the biological diversity of cultivated plant species, the study of their origin, and their phylogeny, presupposes a detailed development of intraspecific classifications (Dorofeev et al., 1979; Goncharov, 2011). The producing of such classifications, reflecting the phylogenesis and genetic structure of species, should be considered the main goal of modern taxonomy (Hammer et al., 2011). When developing the classification of cultivated plants, the most complete description of all existing large and small forms (taxons) is assumed (Sinskaya, 1969). On the one hand, this is determined, by the convenience of using such a division in experimental work, on the other hand, it is also determined in the breeding and testing of cultivated plants.

The success and effectiveness of research work is often associated with the detailing and completeness of the experimental study, which depends on what the material is and how much it should be studied. In this regard, it is extremely important that the natural differentiation of one or another genus, the relationship between species, are reflected with high accuracy by a detailed taxonomy (Dorofeev, 1985). It should be noted, that for most of the plants important for agriculture, the volumes of the genus and species have not been unambiguously described yet (Rodionov et al., 2019).

A serious problem in the taxonomy of cultivated plants is the aspect of taxa aggregation vs. fragmentation, and in cases of cultivation it manifests itself especially in contrast (Golovnina et al., 2009; Goncharov, 2011). At the same time, the effective use of taxonomy of cultivated plants in the work of researchers causes certain difficulties. Both dichotomous tables (Dorofeev et al., 1979; Goncharov, 2009) and ideographic manual book (Zuev et al., 2019) require certain skills; therefore, the producing of a database and software that allows the identification of species by digital

images is a very promising direction. The development of these methods is mainly based on technologies for analyzing digital images of plant organs within the framework of computer phenomics (Afonnikov et al., 2016; Zhang et al., 2019; Demidchik et al., 2020; Yang et al., 2020).

Wheat is one of the world's most important food crops. The modern cultivated wheat species evolved from three wild diploid ancestors as a result of their hybridization and several rounds of genome doubling (polyploidization). Currently, cultivated wheat is represented by di- ( $2n = 2x = 14$ ,  $A^bA^b$  genome), tetra- ( $2n = 4x = 28$ ,  $BBA^uA^uDD$  genome) and hexaploid ( $2n = 6x = 42$ ,  $BBA^uA^uDD$  genome) species (Goncharov, Kondratenko, 2008). The main cultivated species, bread wheat (*Triticum aestivum* L.), is a hexaploid (genomic formula  $BBA^uA^uDD$ ). The ploidy level is one of the main taxonomic and classifying characteristics of wheat species (Dorofeev et al., 1984; van Slageren, Payne, 2013). It can be established by cytogenetic (Rodionov et al., 2020), molecular methods, as well as by comparing the morphological characteristics of plants (Dorofeev et al., 1984). In this work, we studied the morphological traits of the plant spikes of tetraploid and hexaploid wheat species based on the method of high-throughput phenotyping.

The aim of the research was to study the distribution of morphological traits of spikes of tetra- and hexaploid wheat species and compare their distributions.

## Material and methods

**Biological material.** We studied 17 polyploid wheat species, namely, nine hexaploids (*Triticum compactum* Host, *T. aestivum* L., *T. antiquorum* Heer ex Udacz., *T. spelta* L. (inclusion of *T. spelta* cv. Rother Sommer Kolben), *T. petropavlovskiyi* Udacz. et Migusch., *T. yunnanense* King ex S.L. Chen, *T. macha* Dekapr. et Menabde, *T. sphaerococcum* Perciv., *T. vavilovii* (Thum.) Jakubz.), the near-isogenic line ANK-23

**Table 1.** Characteristics of the studied wheat species

Species	Total					Ploidy	List of vegetation	
	photos	plants	accessions	table	pin			
<i>T. compactum</i> Host	472	101	10	177	295	Hexaploid	II18, IX16	
<i>T. aestivum</i> L.	456	80	8	166	290		II19, IX16, IX18, X14	
<i>T. antiquorum</i> Heer ex Udacz.	184	37	4	116	68		II18, X14	
<i>T. spelta</i> L.	164	49	5	40	124		II18	
<i>T. petropavlovskiyi</i> Udacz. et Migusch.	374	75	6	74	300		II17, IX17, IX18	
i:ANK-23	50	10	1	14	36		IX16	
<i>T. yunnanense</i> King ex S.L. Chen	191	43	3	43	148		IX17, IX18	
<i>T. spelta</i> cv. Rother Sommer Kolben	45	9	1	9	36		IX16, II18	
<i>T. macha</i> Dekapr. et Menabde	46	10	1	10	36		IX17, IX18	
<i>T. sphaerococcum</i> Perciv.	100	20	2	20	80		IX17	
<i>T. vavilovii</i> (Thum.) Jakubz.	15	3	1	3	12		II18	
<i>T. aethiopicum</i> Jakubz.	595	119	12	119	476		Tetraploid	X14
<i>T. dicoccoides</i> (Körn. ex Aschers. et Graebn.) Schweinf.	40	8	1	8	32			II16
<i>T. dicoccum</i> (Schrank) Schuebl.	41	9	1	9	32			II17
<i>T. durum</i> Desf.	275	56	5	55	220	II16, II17, II19, IX18		
<i>T. militinae</i> Zhuk. et Migusch.	40	8	1	8	32	IX17		
<i>T. polonicum</i> L.	95	19	2	19	76	II16, II19		
<i>T. timopheevii</i> (Zhuk.) Zhuk.	125	25	3	25	100	II16, IX18		
<i>T. turgidum</i> L.	40	8	1	8	32	II15		

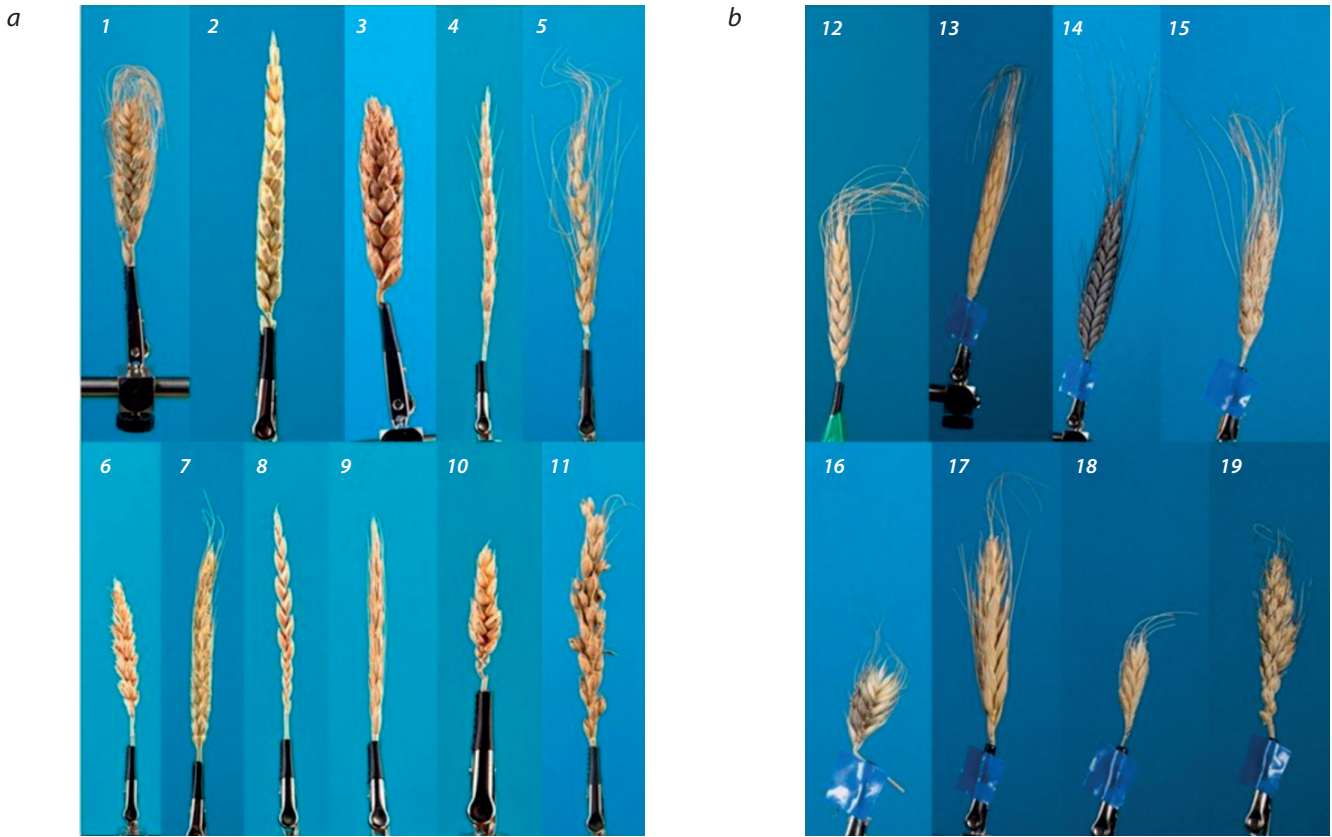
of bread wheat cv. Novosibirskaya 67 and eight tetraploids (*T. aethiopicum* Jakubz., *T. dicoccoides* (Körn. ex Aschers. et Graebn.) Schweinf., *T. dicoccum* (Schrank) Schuebl., *T. durum* Desf., *T. militinae* Zhuk. et Migusch., *T. polonicum* L., *T. timopheevii* (Zhuk.) Zhuk., *T. turgidum* L.); the sample consists of spikes of 595 individual plants, which was grown in nine vegetation seasons. The plants were grown in 2014–2019 in a greenhouse at the Shared Center ‘Laboratory of Artificial Plant Cultivation’ of the Institute of Cytology and Genetics SB RAS. A description of the material used is given in Table 1.

It should be noted that none of the large genebanks of the world have typical sets of wheat accessions (collections), so they usually reflect either the researchers view on the methods of selection of such sets (Palmova, 1935) or are determined by the representativeness of the researchers available material (Goncharov, Shumny, 2008). Standard taxonomic descriptions of specimens are given in publicly available databases on genebank websites (<http://db.vir.nw.ru/virdb/maindb>).

**Digital images obtaining.** In this work we used two protocols for receiving mature spikes photo. The first is that the spike is placed on the glass of a light table, which is located on a table with a blue top (background). The ca-

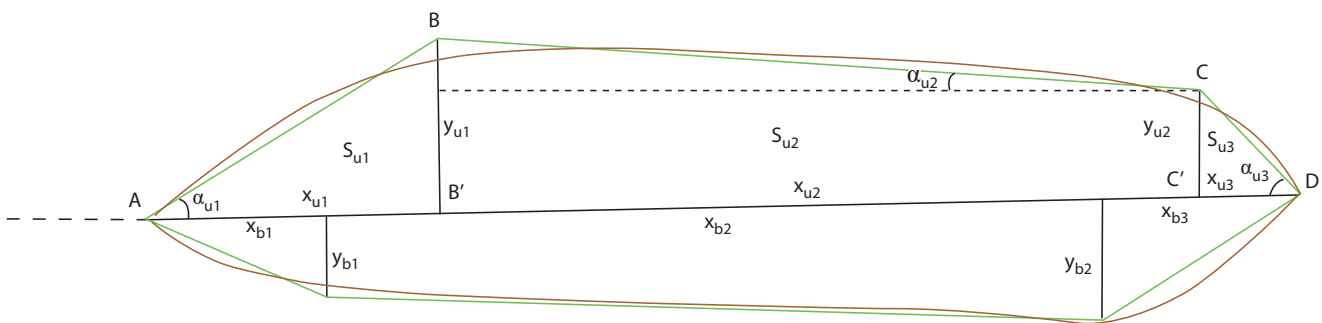
mera is fixed on a stand above the glass. With this method, the front projection of the spike can be captured. Second, the spike is held vertically in front of the blue background. The spike is supported by clip that are placed on a tripod. With this method, by rotating the spike about its axis, four or more projections of the spike can be captured (Genaev et al., 2018). According to the protocols, a ColorChecker must be present in the photographs. It is needed for colour normalisation and scaling. One plant in our dataset can correspond to up to five pictures of its spike taken with different protocols and in different projections. Examples of spikes images (one for each species) are shown in Fig. 1. In total 3348 spike images in different projections were captured by the two protocols, 2097 of them were of hexaploid species and 1251 were of tetraploid species. Of these, 915 images were obtained using the “on the table” protocol and 2433 “on the clip”.

**Evaluation of spikes quantitative characteristics.** WERrecognizer (Genaev et al., 2019) was used to estimate spikes quantitative characteristics based on image analysis. This program describes a wheat spike by geometric model of two quadrangles based on image analysis (Fig. 2). The geometry of this model is described by nine independent parameters. The parameters  $x_{u1}$ ,  $x_{u2}$ ,  $y_{u1}$ ,  $y_{u2}$  are for the



**Fig. 1.** Spike images of hexaploid (a) and tetraploid (b) wheat species.

1 – *T. compactum*; 2 – *T. aestivum*; 3 – *T. antiquorum*; 4 – *T. spelta*; 5 – *T. petropavlovskyi*; 6 – i:ANK-23; 7 – *T. yunnanense*; 8 – *T. spelta* cv. Rother Sommer Kolben; 9 – *T. macha*; 10 – *T. sphaerococcum*; 11 – *T. vavilovii*; 12 – *T. aethiopicum*; 13 – *T. dicoccoides*; 14 – *T. dicoccum*; 15 – *T. durum*; 16 – *T. millitinae*; 17 – *T. polonicum*; 18 – *T. timopheevii*; 19 – *T. turgidum*.



**Fig. 2.** Wheat spike shape represents in the form of two quadrangles (Genaev et al., 2019).

The black horizontal line shows the spike centerline. Brown line – spike contour. Green lines – the quadrilaterals that approximate the spike contour. The spike base – left dotted line. The figure for the upper quadrangle shows the main parameters that characterize spike geometry. Similar parameters are defined for the lower quadrangle.

upper quadrangle; the parameters  $x_{b1}$ ,  $x_{b2}$ ,  $y_{b1}$ ,  $y_{b2}$  are for the lower quadrangle; the common parameter for the two quadrangles is the ear length. The program additionally calculates a number of general features of the shape and size of the spike, as well as the characteristic of its awning. Details of the feature extraction algorithm are given in (Genaev et al., 2019).

In the present study, we used the model traits that we selected as the most informative for predicting spike density

index in our previous study (Genaev et al., 2019), as well as the general shape and spike trait characteristics. These traits characterise a complex view of the morphology (phenotype) of the spike by describing its shape (Circularity, Roundness), the physical dimensions of the ear body (Perimeter, Rachis length) and the area of the awns (Awns area), the traits obtained by approximating the ear by two quadrangles are related to the width ( $x_{u2}$ ,  $y_{bm}$ ) and length ( $x_{b2}$ ,  $y_{u2}$ ) of individual segments of the ear (Table 2).

**Table 2.** Description of the spike trait characters

Features	Description	Dimension
Awns area	Awns area	mm <sup>2</sup>
Circularity	The roundness index is equal to the ratio of the perimeter of a circle with an area equal to the contour area to the perimeter of the contour. The index indicates how close the shape of the contour is to that of a circle. The value varies from 0 to 1	Dimensionless
Roundness	The roundness index is equal to the ratio of the contour area to the area of a circle with a diameter equal to the centreline of the spike	
Perimeter	Perimeter of ear contour without awns	mm
Rachis length	Length of the broken line along the axis of the compound spike (spike axis line)	
X <sub>u2</sub>	Quadrangle model parameter related to the length of the left (the top one in Fig. 2) centre spike	
X <sub>b2</sub>	Quadrangle model parameter related to the length of the right (the bottom one in Fig. 2) centre part of the spike	
Y <sub>u2</sub>	Distance of vertex C to its projection C' on base AD (see Fig. 2)	
Y <sub>bm</sub>	Parameter of the quadrilateral model. Average value of the height of the right (bottom) quadrilateral	

**Data analysis.** In order to estimate the distribution of spikes in the feature space under study, we used a non-linear t-SNE dimensionality reduction algorithm (t-distributed stochastic neighbor embedding; Maaten, Hinton, 2008). This method allows to visualize multidimensional data by mapping objects in multidimensional space to a smaller (two- or three-dimensional) space. The basic idea behind t-SNE is to reduce the dimensionality of space while maintaining the relative pairwise distances between objects. The advantage of the t-SNE method is its tendency to localize isolated, dense spatial structures of arbitrary geometry. The t-SNE method was applied to ordinate images of spikes; the images of each of the projections of a single ear were treated as separate objects.

In order to assess the similarity of the quantitative characteristics of spikes for different species, we used hierarchical clustering (Johnson, 1967) of 17 wheat species according to the traits obtained by averaging over all spikes of the same species. Each species was characterized by a feature vector of length 9. A value of  $1 - r$  was used as a metric for the distance between species, where  $r$  is the value of the Pearson correlation coefficient between the values of the traits (Müllner, 2011). The linkage (UGMA algorithm) and dendrogram functions from the SciPy library (Virtanen et al., 2020) were used for clustering and dendrogram construction.

To compare the variance of traits in plants belonging to different ploidy types, we used  $F$  statistics (Snedecor, Cochran, 1989), which evaluates the significance of differences in the variance of two distributions. The data were normalized by the StandardScaler function of the scikit-learn library (Pedregosa et al., 2011). The test was performed independently for each of the nine traits described in Table 2. In this test, one spike image per plant was used, obtained in the “on the table” projection protocol.

## Results and discussions

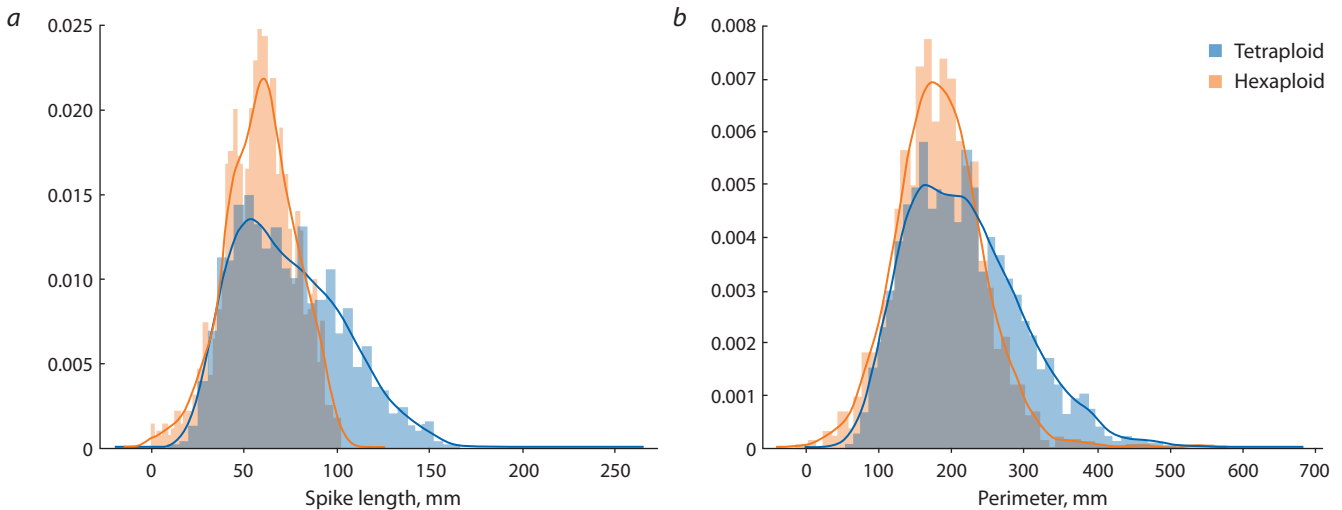
The mean, median, standard deviation and variance of the nine features calculated for the 17 wheat species are presented in the Supplementary 1<sup>1</sup>.

Let's review the distribution of spikes in our sample of plants according to the characteristic “area of the spikes”. The higher this parameter, the more awns were identified for the spike in the image. According to this characteristic, spikes of hexaploid wheat can be conditionally divided into three classes: awned (parameter value above 90), moderately awned (parameter value from 30 to 90), and awnless (parameter value below 30). The species *T. compactum*, *T. spelta*, *T. petropavlovskyi* and *T. vavilovii* are considered awned according to this criterion. *T. aestivum*, *T. yunnanense*, *T. macha* are moderately awned. The awnless ones are *T. antiquorum*, *i:ANK-23*, *T. spelta* cv. Rother Sommer Kolben, and *T. sphaerococcum* (see Supplementary 1). These data agree well with the appearance of the spikes (see Fig. 1, a). Thus, representatives of hexaploid wheat show considerable diversity in the presence/absence of awns.

If the classification above is applied to tetraploid wheat, only representatives of *T. militinae* (mean value of the parameter 24.09 mm<sup>2</sup>) can be assigned to the awnless category. Four species can be classified as moderately awned: *T. dicoccoides*, *T. polonicum*, *T. timopheevii* and *T. turgidum*. Three species are considered awned: *T. aethiopicum*, *T. dicoccum*, *T. durum*. In general, the representation of awned species (specimens) in tetraploid species is significantly higher than in hexaploid species.

Analysis of such characteristic as spike length shows that spikes can also be divided into three classes: length less than 60 mm (short), from 60 to 90 mm (medium) and more than 90 mm (long). According to this classification, the hexaploid

<sup>1</sup> Supplementary materials 1–4 are available in the online version of the paper: [http://vavilov.elpub.ru/jour/manager/files/SupplPronozinA\\_Engl.pdf](http://vavilov.elpub.ru/jour/manager/files/SupplPronozinA_Engl.pdf)



**Fig. 3.** Length (a) and perimeter (b) distribution of a wheat spike in tetraploid (blue) and hexaploid (orange) wheat species.

wheat species *T. spelta*, *T. petropavlovskiyi* and *T. vavilovii* can be classified as long spikes, *T. aestivum*, *T. yunnanense*, *T. spelta* cv. Rother Sommer Kolben and *T. macha* to medium spikes, and *T. compactum*, *T. antiquorum*, *T. sphaerococcum* and the near-isogenic lineage ANK-23 to short spikes. The boundary between species characterized by long and medium spikes is rather conditional. For tetraploid species we did not find any species which according to this parameter would fall into the category of long-boned. The medium-sized category could include *T. aethiopicum*, *T. dicoccoides*, *T. polonicum*, *T. turgidum*, the short spike category – *T. dicoccum*, *T. durum*, *T. timopheevii* and *T. militinae*.

The spike length distribution of the samples studied for hexaploid and tetraploid species is shown in Fig. 3, a. The Fig. 3, b shows the distribution of the parameter also characterizing the size of the spikes – the perimeter of the contour of the body of the spike in the image.

Fig. 3 shows that the distributions of both parameters in hexaploid wheat are more scattered, while the variability of these traits in hexaploid wheat is higher mainly due to the higher frequency of occurrence of ears with high values of these traits.

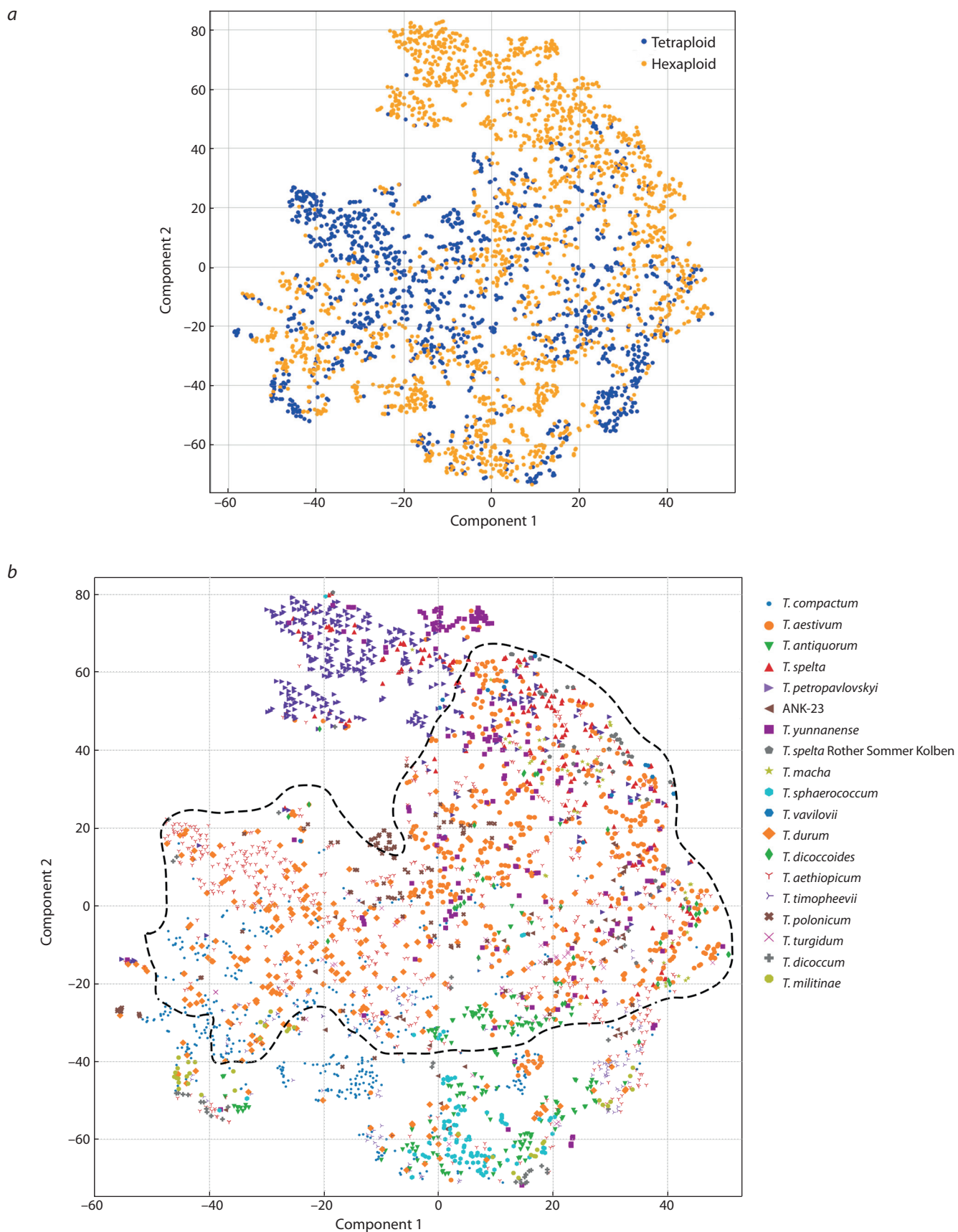
The distribution of the analyzed ears images in the space of nine features was visualized using the t-SNE method, resulting in a two-dimensional parameter space (components 1 and 2). The results of the transformation are shown in Fig. 4. In the resulting diagram, each point represents one of the analysed images of the spike. In Fig. 4, a the dots are coloured according to the type of ploidy of the plant (blue colour corresponds to tetraploid wheat species, orange to hexaploid ones). In Fig. 4, b the colour and shape of each dot corresponds to a particular wheat spike image.

The diagram in Fig. 4, a shows that the areas occupied by hexa- and tetraploid wheat species strongly overlap on the graph. This means that the spikes of these two groups are quite similar in their characteristics. This is consistent with the results presented in the Supplementaries 1 and 2

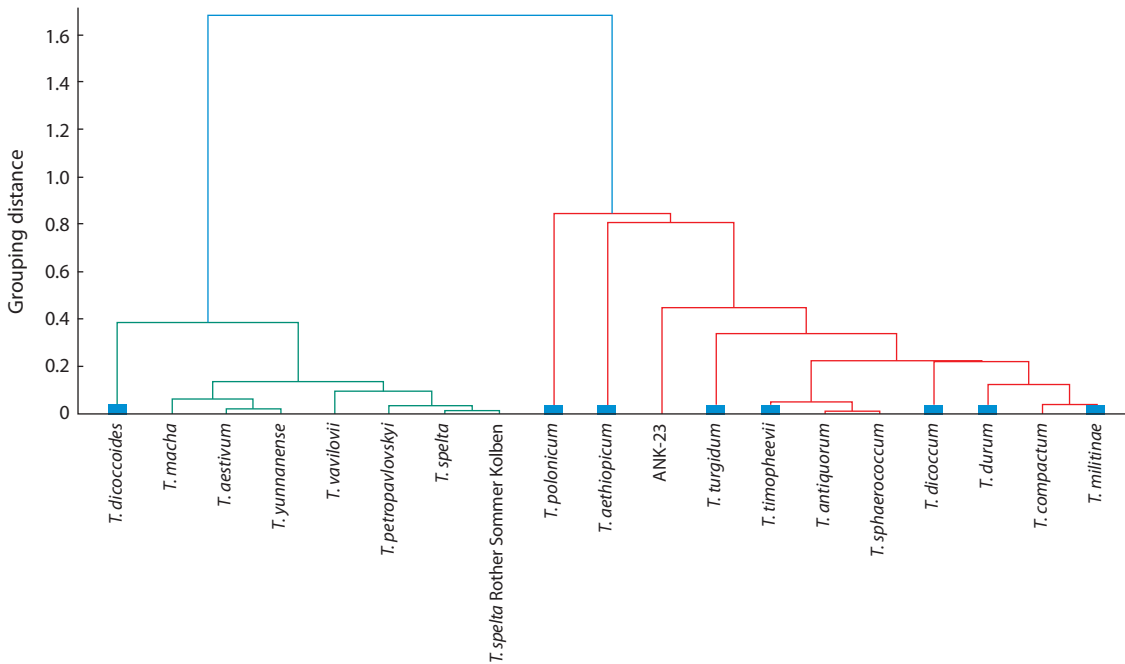
as well as in Fig. 3. However, it should be noted that in the diagram in Fig. 4, a samples of hexaploid species occupy a larger area, primarily due to the predominance of the corresponding points in the right part of the diagram. One can see that orange dots (hexaploid wheat) predominate in the area with values of component 1 more than –20, this predominance is even more pronounced in the upper right corner of the diagram (values of component 1 less than 0 and component 2 more than 20). This means that a number of spike trait characteristics have some values for hexaploid species specific only, but not for tetraploid ones. This agrees well with the result shown in Fig. 3. In particular, such areas may correspond to large values of the parameters “perimeter” and “ear length”.

The diagram in Fig. 4, b shows that the areas occupied by samples of the different species overlap considerably. For example, *T. aestivum* and *T. durum* species overlap across the entire plot area (dotted line). At the same time, it should be noted that the images of spikes belonging to the same wheat species occupy mostly compact areas on the graph. At the same time, there are species for which the spike samples are divided into several clearly visible clusters according to their characteristics. Such species include *T. compactum* (small blue circle, component 1 from –60 to 0, component 2 from –60 to 0) and *T. petropavlovskiyi* (purple triangle, component 1 from –20 to 0, component 2 from 40 to 80).

Fig. 1 shows that hexaploids are represented by plants with two characteristic types of spikes: long and thin (*T. aestivum*, *T. spelta*, *T. petropavlovskiyi*, *T. yunnanense*, *T. spelta* cv. Rother Sommer Kolben); short and rounded (*T. compactum*, *T. antiquorum*, i:ANK-23, *T. sphaerococcum*, *T. macha*, *T. vavilovii*). In Fig. 4, b the group of plants with short and rounded spikes is located in the component 2 value range from –80 to 0 (lower part of the graph). Plants with long and thin spikes have component 2 values between 0 and 80 (upper part of graph). In Fig. 4, a, these two groups of plants correspond roughly to the two clouds of dots in hexaploid



**Fig. 4.** Clustering of spike digital images of individual genotypes by the t-SNE method, obtained on the basis of quantitative traits from Table 2. *a* – blue color corresponds to tetraploid wheat species, orange – hexaploid; *b* – the color and shape of each point corresponds to a specific type. The blue polygon marks the area occupied by *T. aestivum* and *T. durum* species. Clustering is called automatic partitioning into clusters. The automatic arrangement on the plane and in space is called ordination.



**Fig. 5.** Results of the hierarchical cluster analysis for nine signs of a wheat spike. Blue squares correspond to tetraploid.

wheat at the top and bottom of the graph, which overlap slightly in the central part of the graph. Thus, the diagrams in Fig. 4 provide a clear indication of the diversity of spikes in their characteristics within and between species.

To characterize in more detail the similarity of morphometric characteristics of spikes in different wheat species, we conducted a hierarchical cluster analysis for them based on a comparison of the mean values of the studied traits. (Fig. 5).

Fig. 5 shows that the wheat species were divided into two clusters (highlighted in red and green). The first cluster (red) predominantly includes tetraploid species (shown in blue rectangles near the terminal tree nodes). However, wild tetraploid wheat species *T. dicoccoides* is not included in this cluster, while among hexaploid species, *T. compactum*, *T. antiquorum* and *T. sphaerococcum* differing from all other species by compact spike shape, i.e. having the shortest spike of all studied hexaploid wheat species are included in it. It should be noted that in the work of A. Zatybekov et al. (2020), using economically important traits, samples of six tetraploid species were clustered arbitrarily, i.e. irrespective of their species identity. It is important to note, that remaining hexaploid species were clearly divided by spike length into two clusters of medium (*T. macha*, *T. aestivum*, and *T. yunnanense*) and long spikes (*T. vavilovii*, *T. petropavlovskiyi*, and *T. spelta*).

*T. spelta* and *T. spelta* cv. Rother Sommer Kolben (a German landrace) occur in the same cluster. This allows us to conclude that the “species” shape of spike during long-term wheat breeding did not change for a long time (in this case, more than fifty years) and may be successfully used for classification of the species.

It should be noted that the only wild tetraploid loose spike species in the genus, *T. dicoccoides*, has fallen to the hexaploids. While hexaploid wheat species with compact ear type – *T. compactum*, *T. antiquorum*, *T. sphaerococcum* and human-made near-isogenic line ANK-23 of spring bread wheat cv. Novosibirskaya 67 (Koval, 1997) – were included into tetraploid species. The latter leads to the conclusion that although near-isogenic lines are produced on a particular (specific) species, nevertheless, their species identity should be treated with caution.

Let’s take a look at *T. petropavlovskiyi*. The species was founded at the Chinese Pamir – route of the Great silk road. According to the results of the study of gliadins, all accessions of this species were very similar to such hybrid combination obtained from crossing bread wheat with *T. polonicum* (Watanabe et al., 2004). The authors of the “Cultural Flora of the USSR” also considered a possible hybrid origin of this species (Dorofeev et al., 1979). In addition, *T. petropavlovskiyi* also resembles bread wheat in a number of taxonomic traits (Goncharov, 2005). Previously, R.L. Boguslavsky (1982) described hybrids from crossing *T. aestivum* with *T. polonicum* produced by CIMMYT breeders as subspecies of *T. petropavlovskiyi* ssp. *mexicana* Bogusl. Based on the above, we considered it appropriate to conbided *T. petropavlovskiyi* as the subspecies of *T. aestivum*:

***Triticum aestivum* ssp. *petropavlovskiyi* comb. et stat. nov. (Udacz. et Migusch.) N.P. Gontsch.** – *T. turanicum* Jakubcz. convar. *montanostepposum* Jakubcz. f. *aristiforme* Jakubcz. 1959. *Bot. Zhur.* 10:1428, nom. illig. – *T. petropavlovskiyi* Udacz. et Migusch. 1970. *Vestn. Sel’skokhoz. Nauki.* 9:20.



**Table 3.** Results of using *F* statistics to confirm the hypothesis of a significant difference in the variance of two distributions

Features	<i>F</i> -statistics	<i>p</i> -value	Hexaploid dispersion	Tetraploids dispersion	Average hexaploids	Average tetraploids
Awns area	0.376	1.000	1.415	3.763	84.875	160.643
Circularity index	1.188	0.065	0.959	0.807	0.178	0.181
Roundness	1.828	1.110e-07	<b>1.312</b>	0.718	0.141	0.172
Perimeter	1.570	4.710e-05	<b>1.080</b>	0.688	218.124	185.015
Rachis length	3.500	< 1e-15	<b>1.320</b>	0.377	74.136	59.280
$\chi_{u2}$	3.928	< 1e-15	<b>1.336</b>	0.340	53.837	36.853
$\chi_{b2}$	4.437	< 1e-15	<b>1.331</b>	0.300	54.004	36.726
$\gamma_{u2}$	4.275	< 1e-15	<b>2.491</b>	0.583	3.844	4.171
$\gamma_{bm}$	1.081	0.248	0.695	0.643	0.225	0.246

Note. Significant dispersion differences are shown in bold.

**Typus:** described by an accession from China “Origin: China, Xinjiang Province, village Kurlia, K-48376, 1957. A.M. Gorsky exp[edition]. Reproduction of Central Asia, Tashkent, Central Asian Station of VIR. 08. VII. 1969, Collected/defined: R.A. Udachin & E.F. Migushova” in St. Petersburg (WIR!). (The herbarium specimens of the type and paratype of *Triticum aestivum* ssp. *petropavlovskiyi* are given in the Supplementaries 3 and 4).

Note that the results presented in Fig. 3 and 4, *a* show that hexaploid species have a greater variability in spike shape, size and awnness characteristics. Therefore, we hypothesized that the spike trait characteristics of hexaploid species may have a higher variation than those of tetraploid species. To test this assumption, we compared the variance of the estimated parameters using an *F*-distribution (Table 3).

The results presented in Table 3 show that the variance of most of the characters for hexaploids and tetraploids have significant differences ( $p < 0.05$ ). At the same time, the significant differences in variance were not found for such traits as  $\gamma_{bm}$  (quadrangle model parameter), Awns area and Circularity index. It is interesting to note that for all significant differences, we observe a higher variance in hexaploids than in tetraploids. Thus, the analysis showed that hexaploid species show higher diversity in spike morphometric trait characteristics compared to tetraploid species.

The data represent plants of 17 wheat species: 9 hexaploids (*T. compactum*, *T. aestivum*, *T. antiquorum*, *T. spelta* (including *T. spelta* cv. Rother Sommer Kolben), *T. petropavlovskiyi*, i:ANK-23 (near-isogenic line of bread wheat cv. Novosibirskaya 67), *T. yunnanense*, *T. macha*, *T. sphaerococcum*, *T. vavilovii*) and 8 tetraploids (*T. aethiopicum*, *T. dicoccoides*, *T. dicoccum*, *T. durum*, *T. militinae*, *T. polonicum*, *T. timopheevii*, *T. turgidum*). The results of their clustering are presented so that the colour and shape of each dot corresponds to a particular species (see Fig. 5).

It is well known that genome doubling as a result of duplications (autopolyploidy) or hybridization and subsequent polyploidization (allopolyploidy) leads to marked changes in plant phenotype (Finigan et al., 2012; Romanov, Pimonov, 2018; Rodionov et al., 2019). These changes in plants occur both at the cellular level (Liu et al., 2018) and at the organ level (Robinson et al., 2018). In many cases, in plants, an increase in ploidy leads to an increase in cell and organ size (Comai, 2005; Williams, Oliveira, 2020), increasing resistance to stress (Tan et al., 2015). Currently, researchers suggest that there are four types of molecular mechanisms of such variability: 1) increased gene/allele dosage, 2) increased genetic diversity, 3) altered genetic regulation, and 4) epigenetic rearrangements of the genome (Chen, 2007; Finigan et al., 2012).

The analysis of morphological characteristics of spikes of hexaploid ( $2n = 6x = 42$ ) and tetraploid ( $2n = 4x = 28$ ) wheat has shown, that most of spike characteristics have significantly higher variation in wheat with higher spike ploidy. Our results are in agreement with the ideas about the influence of ploidy on plant phenotype variability.

## Conclusion

A large-scale analysis of the spike digital images of 595 plants of 8 tetra- and 9 hexaploid wheat species was carried out. Nine quantitative traits describing the shape, size and awnedness of the spike were studied. The variability among the above genotypes was studied and it was shown that two clusters are formed in the spike characteristic space. The first cluster includes mainly hexaploid species (with the exception of wild tetraploid species *T. dicoccoides*). The second cluster includes tetraploid species (with the exception of three hexaploid species with compact spike shape – *T. antiquorum*, *T. sphaerococcum*, and near-isogenic line ANK-23). Analysis of variance of these characters in hexaploid and tetraploid wheats showed a significant in-

crease in variance for six of nine characters in the sample of hexaploids, i.e. greater ploidy level gives more variability in quantitative characters of spike morphology.

Thus, it is shown that morphological trait characteristics of spikes of hexa- and tetraploid species, obtained on the basis of computer image analysis, demonstrate the differences, which can be used in the future to develop a method of classification of plants by ploidy level and their species affiliation in automatic mode.

## References

- Afonnikov D.A., Genaev M.A., Doroshkov A.V., Komyshev E.G., Pshenichnikova T.A. Methods of high-throughput plant phenotyping for large-scale breeding and genetic experiments. *Russ. J. Genet.* 2016;52(7):688-701. DOI 10.1134/S1022795416070024.
- Boguslavsky R.L. A new botanical form of hexaploid wheat. *Nauchno-Tekhnicheskiiy Byulleten VIR = Scientific and Technological Bulletin of the Vavilov Institute of Plant Industry.* 1982;119:73-74. (in Russian)
- Chen Z.J. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* 2007;58:377-406. DOI 10.1146/annurev.arplant.58.032806.103835.
- Comai L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 2005;6(11):836-846. DOI 10.1038/nrg1711.
- Demidchik V.V., Shashko A.Y., Bandarenka V.Y., Smolikova G.N., Przhivalskaya D.A., Charnysh M.A., Pozhvanov G.A., Barkovskiy A.V., Smolich I.I., Sokolik A.I., Yu M., Medvedev S.S., Plant phenomics: fundamental bases, software and hardware platforms, and machine learning. *Russ. J. Plant Physiol.* 2020;67:397-412. DOI 10.1134/S1021443720030061.
- Dorofeev V.F. Intraspecific taxonomy of wheat. *Doklady VASKhNIL = Reports of the Academy of Agricultural Sciences.* 1985;9:1-4. (in Russian)
- Dorofeev V.F., Filatenko A.A., Migushova E.F., Udachin R.A., Yakubtsiner M.M. Flora of Cultivated Plants of USSR. Vol. 1. Wheat. Leningrad: Kolos Publ., 1979. (in Russian)
- Dorofeev V.F., Rudenko M.I., Filatenko A.A., Baras J., Segalova J., Lemann H. (Compilers). The International Comecon List of Descriptors for the Genus *Triticum* L. Leningrad: VIR Publ., 1984. (in Russian)
- Finigan P., Tanurdzic M., Martienssen R.A. Origins of novel phenotypic variation in polyploids. In: *Polyploidy and Genome Evolution.* Berlin; Heidelberg: Springer Press, 2012;57-76. DOI 10.1007/978-3-642-31442-1\_4.
- Genaev M.A., Komyshev E.G., Fu Hao, Koval V.S., Goncharov N.P., Afonnikov D.A. SpikeDroidDB: an information system for annotation of morphometric characteristics of wheat spike. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding.* 2018;22(1):132-140. DOI 10.18699/VJ18.340. (in Russian)
- Genaev M.A., Komyshev E.G., Smirnov N.V., Kruchinina Y.V., Goncharov N.P., Afonnikov D.A. Morphometry of the wheat spike by analyzing 2D images. *Agronomy.* 2019;9(7):390. DOI 10.3390/agronomy9070390.
- Golovnina K.A., Kondratenko E.Ya., Blinov A.G., Goncharov N.P. Phylogeny of the A genome of wild and cultivated wheat species. *Russ. J. Genet.* 2009;45(11):1360-1367. DOI 10.1134/S1022795409110106.
- Goncharov N.P. Comparative-genetic analysis – a base for wheat taxonomy revision. *Czech J. Genet. Plant Breed.* 2005;41:52-55.
- Goncharov N.P. Manual Book of Common and Hard Wheat Varieties. Novosibirsk: SO RAN Publ., 2009. (in Russian)
- Goncharov N.P. Genus *Triticum* L. taxonomy: the present and the future. *Plant Syst. Evol.* 2011;295(1-4):1-11. DOI 10.1007/s00606-011-0480-9.
- Goncharov N.P., Kondratenko E.Ya. Wheat origin, domestication and evolution. *Informatsionnyy Vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeders.* 2008;12(1/2):159-179. (in Russian)
- Goncharov N.P., Shumny V.K. From preservation of genetic collections to organization of National project of plant gene pools conservation in permafrost. *Informatsionnyy Vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeders.* 2008;12(4):509-523. (in Russian)
- Hammer K., Filatenko A.A., Pistrick K. Taxonomic remarks on *Triticum* L. and  $\times$ *Triticosecale* Wittm. *Genet. Resour. Crop Evol.* 2011;58(1):3-10. DOI 10.1007/s10722-010-9590-4.
- Johnson S.C. Hierarchical clustering schemes. *Psychometrika.* 1967;32(3):241-254.
- Koval S.F. The catalog of near-isogenic lines of Novosibirskaya-67 common wheat and principles of their use in experiments. *Russ. J. Genet.* 1997;33(8):995-1000.
- Liu W., Zheng Y., Song S., Huo B., Li D., Wang J. *In vitro* induction of allohexaploid and resulting phenotypic variation in *Populus*. *Plant Cell Tiss. Organ Cult.* 2018;134(2):183-192. DOI 10.1007/s11240-018-1411-z.
- Müllner D. Modern hierarchical, agglomerative clustering algorithms. *arXiv.* 2011;1109.2378.
- Palmova E.F. Introduction to Wheat Ecology. Moscow; Leningrad: Selkhozgiz Publ., 1935. (in Russian)
- Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Müller A., Nothman J., Louppe G., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 2011;12:2825-2830.
- Robinson D.O., Coate J.E., Singh A., Hong L., Bush M., Doyle J.J., Roeder A.H. Ploidy and size at multiple scales in the *Arabidopsis* sepal. *Plant Cell.* 2018;30(10):2308-2329. DOI 10.1105/tpc.18.00344.
- Rodionov A.V., Amosova A.V., Belyakov E.A., Zhurbenko P.M., Mikhailova Y.V., Punina E.O., Shneyer V.S., Loskutov I.G., Muravenko O.V. Genetic consequences of interspecific hybridization, its role in speciation and phenotypic diversity of plants. *Russ. J. Genet.* 2019;55(3):278-294. DOI 10.1134/S1022795419030141.
- Rodionov A.V., Shneyer V.S., Gnutikov A.A., Nosov N.N., Punina E.O., Zhurbenko P.M., Loskutov I.G., Muravenko O.V. Species dialectics: from initial uniformity, through the greatest possible diversity to ultimate uniformity. *Botanicheskiiy Zhurnal = Botanical Journal.* 2020;105(9):835-853. DOI 10.31857/S0006813620070091. (in Russian)
- Romanov B.V., Pimonov K.I. Phenomogenomics of Production Traits of Wheat Species. Persianovsky: Donskoy GAU Publ., 2018. (in Russian)
- Sinskaya E.N. Historical Geography of Cultural Flora (At the Dawn of Agriculture). Leningrad: Kolos Publ., 1969. (in Russian)
- Snedecor G.W., Cochran W.G. Statistical Methods. Ames, Iowa: Iowa State University Press, 1989.
- Tan F., Tu H., Liang W., Long J.M., Wu X.M., Zhang H.Y., Guo W.W. Comparative metabolic and transcriptional analysis of a doubled diploid and its diploid citrus rootstock (*C. junos* cv. Ziyang xiangcheng) suggests its potential value for stress resistance improvement. *BMC Plant Biol.* 2015;15:89. DOI 10.1186/s12870-015-0450-4.
- Udachin R.A., Migushova E.F. New in the knowledge of the genus *Triticum*. *Vestnik Selskokhozyaystvennoy Nauki = Herald of Agricultural Sciences.* 1970;9:20-24. (in Russian)
- van der Maaten L., Hinton G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 2008;9:2579-2605.
- van Slageren M., Payne T. Concepts and nomenclature of the Farro wheats, with special reference to Emmer, *Triticum turgidum* subsp.

- dicoccum* (Poaceae). *Kew Bull.* 2013;68:477-494. DOI 10.1007/S12225-013-9459-8.
- Virtanen P., Gommers R., Oliphant T.E., Haberland M., Reddy T., Cournapeau D., Burovski E., Peterson P., Weckesser W., Bright J., van der Walt S.J., Brett M., Jones E., Kern R., Larson E., Carey C.J., Polat I., Feng Yu, Moore E.W., VanderPlas J., Laxalde D., Perktold J., Cimrman R., Henriksen I., Quintero E.A., Harris C.R., Archibald A.M., Riberio A.H., Pedregosa F., van Mulbregt P. SciPy 1.0 Contributors. SciPy 1.0 – fundamental algorithms for scientific computing in Python. *Nat. Meth.* 2020;17(3):261-272. DOI 10.1038/s41592-019-0686-2.
- Watanabe N., Bannikova S.V., Goncharov N.P. Inheritance and chromosomal location of the genes for long glume phenotype found in Portuguese landraces of hexaploid wheat, 'Arrancada'. *J. Genet. Breed.* 2004;58:273-278.
- Williams J.H., Oliveira P.E. For things to stay the same, things must change: polyploidy and pollen tube growth rates. *Ann. Bot.* 2020; 125(6):925-935. DOI 10.1093/aob/mcaa007.
- Yakubtsiner M.M. More on Chinese wheats. *Botanicheskiy Zhurnal = Botanical Journal.* 1959;44(10):1425-1436. (in Russian)
- Yang W., Feng H., Zhang X., Zhang J., Doonan J.H., Batchelor W.D., Xiong L., Yan J. Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant.* 2020;13(2):187-214. DOI 10.1016/j.molp.2020.01.008.
- Zatybekov A., Anuarbek S., Abugalieva S., Turuspekov Y. Phenotypic and genetic variability of a tetraploid wheat collection grown in Kazakhstan. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding.* 2020;24(6):605-612. DOI 10.18699/VJ20.654.
- Zhang Y., Zhao C., Du J., Guo X., Wen W., Gu S., Wang J., Fan J. Crop phenomics: current status and perspectives. *Front. Plant Sci.* 2019; 10:714. DOI 10.3389/fpls.2019.00714.
- Zuev E.V., Amri A., Brykova A.N., Pyukkenen V.P., Mitrofanova O.P. Atlas of the Diversity of Soft Wheat (*Triticum aestivum* L.) by Ear and Grain Characteristics. St. Petersburg: Kopi-R Publ., 2019. (in Russian)

---

#### ORCID ID

A.Yu. Pronozin orcid.org/0000-0002-3011-6288  
Y.V. Kruchinina orcid.org/0000-0002-1084-9521

**Acknowledgements.** Preparation of spike samples, phenotyping, development of algorithms for shape analysis and classification were funded by the Kurchatov Genome Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662. Cultivation of experimental plants and *de visu* determination of their species belonging by the traits determining the architectonics of the spike were supported by grant of Russian Science Foundation 16-16-10021. The data were processed using the resources of the Bioinformatics Center supported by the budget project No. 0259-2021-0009. The authors A.A.P., E.A.Z., A.Yu.P., and N.M.P. were supported by the Mathematical Center in Akademgorodok, agreement with the Ministry of Science and Higher Education of the Russian Federation No. 075-15-2019-1675. The authors are grateful to D.A. Afonnikov for comments and recommendations during the work and I.G. Chukhina (VIR, St. Petersburg) for providing photos of herbarium specimens of *T. petropavlovskyi*.

**Conflict of interest.** The authors declare no conflict of interest.

Received October 27, 2020. Revised December 31, 2020. Accepted January 2, 2021.