

SZENT ISTVÁN UNIVERSITY

Doctoral School of Environmental Sciences

**ANALYSIS OF THE DIVERSITY OF
EINKORN THROUGH MORPHOMETRIC
METHODS**

Doctoral theses

ANDREA EMÓDI

Gödöllő

2019

NAME OF THE DOCTORAL SCHOOL:

DOCTORAL SCHOOL OF ENVIRONMENTAL SCIENCES

BRANCH OF SCIENCE:

ENVIRONMENTAL SCIENCES

HEAD OF THE DOCTORAL SCHOOL:

DR. ERIKA CSÁKINÉ MICHÉLI

PROFESSOR, HEAD OF INSTITUTE

SZENT ISTVÁN UNIVERSITY, FACULTY OF AGRICULTURAL AND

ENVIRONMENTAL SCIENCES,

INSTITUTE OF ENVIRONMENTAL SCIENCE,

DEPARTMENT OF SOIL SCIENCE AND AGRICULTURAL CHEMISTRY

SUPERVISOR:

DR. FERENC GYULAI

PROFESSOR, DOCTOR OF THE HUNGARIAN ACADEMY OF SCIENCES

SZENT ISTVÁN UNIVERSITY, FACULTY OF AGRICULTURAL AND

ENVIRONMENTAL SCIENCES,

ENVIRONMENT AND LANDSCAPE MANAGEMENT INSTITUTE,

DEPARTMENT OF NATURE CONSERVATION AND LANDSCAPE ECOLOGY

CO-SUPERVISOR:

DR. JÓZSEF BERKE

PROFESSOR, CANDIDATE OF THE HUNGARIAN ACADEMY OF SCIENCES

DENNIS GABOR COLLEGE, INSTITUTE OF INFORMATICS

.....

Approval of the head of the doctoral school

.....

Approval of the supervisor

1. Introduction and objectives

Einkorn (*Triticum monococcum* subsp. *monococcum*) is one of the first cereals of humankind. It was first produced approximately 12 000 years ago, and was the dominant grain crop of several historical eras. In Hungary it was produced in smaller areas until the 19th century after which it slowly disappeared from production. The nutritional value, vitamin, mineral matter and trace element content of prehistoric and ancient wheat species were significantly higher than that of current intensively produced winter wheat species. Due to the spread of environmentally sound agricultural practices and the growing popularity of healthy food consumption ancient species and landraces are becoming more and more important today. However, for restarting their production and their spreading more detailed information is needed on their origin, development, diversity and accessible genetic resources. The primary sources of available einkorn items are archaeological findings and gene bank samples that is mostly charred and recent grains. The metric recording and the most accurate description of the crops is essential, since their shape identifies the species and presumably also the different varieties. By knowing them in detail the genetics, origin and spread of cultivated plants can be traced back in history, and we may also find the key to the so far unanswered questions on the plant production relations of historical nations (Berke et al. 2013)¹.

Although genetics is a suitable method for studying such issues, its analysis requires the destruction of einkorn grains of gene bank or archaeological origin, therefore we used one of the latest test methods, digital seed morphometry, for the analysis of the oldest produced cereal by which information on taxonomic classification, domestication development and the diversity of the species also can be gained from the grain.

This is a rapid and accurate method also suitable for performing population analyses, and enabling for intelligent classification beyond recording the digital visual – and consequently the processable numeric – data of the grains. Computers can “learn” how to identify and classify an unknown sample based on morphometric parameters, and the process can be semi-automated (Berke et al. 2013)¹.

¹ BERKE J., PENKSZA K., GYULAI F., FERENCZ Z. (2013): Intelligent seed identification system in education. p. 26-28. In: 19th Multimedia in education conference presentations: Kosice, 13-14th June 2013

Since this method is relatively new, and is currently used only by a few people in natural sciences, its standard technology has not been developed so far. Accordingly, my primary objective was to develop a testing system suitable for the measuring of the grains of *T. monococcum* by which the analysis of einkorn accessions stored in the Hungarian gene banks can be carried out.

My research aimed at the following objectives:

- Developing a visual and a morphometric database of seeds for *Triticum monococcum* accessions stored in the Hungarian gene banks,
- Differentiation of domesticated einkorn (*T. monococcum* subsp. *monococcum*) from its wild ancestor (*T. monococcum* subsp. *aegilopoides*) based on the morphometric database of seeds, and studying the process of domestication,
- Identification and taxonomic classification of an unknown seed at the *varietas* level.

Through my results I aimed at collecting more detailed information on the species of *T. monococcum* besides the already available ones by using a new method, and providing useful data for further archaeological, gene bank and plant production activities.

2. Material and method

2.1. Development of the morphometric database of *Triticum monococcum* seeds by using digital image processing

In the first phase of the analysis participating samples needed to be defined, which was followed by the elaboration of an appropriate methodology for the development of the morphometric database of seeds.

In order to carry out morphometric analyses on seeds it was a prerequisite for *Triticum monococcum*. L. accessions to include at least 50 seeds, and also to have a known taxonomic classification. The Centre for Agricultural Research (Martonvásár) provided 17 samples for the analysis with the following information (sett Table 1):

Table 1: Traits of einkorn samples provided by the Centre for Agricultural Research (Martonvásár)

Species	Subspecies	Varietas	Convarietas	Sample size	Year of sample production
<i>Triticum monococcum</i>	<i>aegilopoides</i>			200	2013
<i>Triticum monococcum</i>	<i>aegilopoides</i>	<i>Nigrireuteri</i>		200	2012
<i>Triticum monococcum</i>	<i>aegilopoides</i>	<i>Symbolense</i>		200	2013
<i>Triticum monococcum</i>	<i>aegilopoides</i>	<i>Mayssuriani</i>		200	2013
<i>Triticum monococcum</i>	<i>aegilopoides</i>	<i>Thaoudar</i>		200	2013
<i>Triticum monococcum</i>	<i>aegilopoides</i>	<i>Pseudoboeoticum</i>		200	2013
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>Vulgare</i>		200	2013
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>Albohornemannii</i>		200	2010
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>Macedonicum</i>		200	2013
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>Hornemannii</i>		200	2013
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>Flavescens</i>		200	2013
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>Atriaristatum</i>		200	2013
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>Vulgare</i>	<i>Balkan</i>	50	2010
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>vulgare</i>	<i>Schwedisches Einkorn</i>	50	2012
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>vulgare</i>	GÖRÖG-LANDRACE	50	2012
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>vulgare</i>	Csanaki	50	2012
<i>Triticum monococcum</i>	<i>monococcum</i>	<i>vulgare</i>	EPEAUTRE	50	2012

These 17 samples fulfilled all the requirements. Their taxonomic classification was known up to the *varietas* level, they were all harvested from the same field, and habitat conditions did not affect the size of the grains. Since the majority of the samples were produced in 2013 enough samples were available also for those analyses where only seeds produced in the same year could be used in order to eliminate production year effects.

I used the software Fovea Pro 4.0 for the development of the morphometric database and for the measuring of *Triticum monococcum* accessions, consequently the methodology of measuring and necessary tools were primarily defined by the input demands of this program.

Accordingly, images to be taken must meet the following basic criteria:

- camera-sample distance shall stay constant during taking the images,
- the optical axle of the camera shall be fixed at right angles to the plane of the samples during taking the images,
- even lighting is required,
- proper settings for taking images shall be made.

Besides, there are also some expectations which can be met at different levels. From the point of measuring, their better accomplishment means achieving more accurate, simpler, faster and more reliable results and collecting data that can be used for even more detailed analyses. Many times these expectations are inversely proportional (e.g. increasing resolution decreases the number of images that can be taken within a certain period of time), so the most optimal trade-off needs to be found. Such expectations on the images include:

- taking images of as many objects as possible within a certain period of time,
- taking as high resolution images as possible,
- minimizing the occurrence of shading and other unwanted objects on the images,
- minimizing noises,
- avoiding the contact of objects with each other and the edge of the image.

By taking all these aspects into account, finding the proper equipment for making the images was the first task requiring great care during the development of the methodology.

2.2 Digital analysis methods used during the research

The Fovea Pro program and its toolkit were basically designed to create comprehensive algorithms in the form of plug-ins that can be used for various programs. This is especially useful when using the Adobe Photoshop program, since 8 and 16 bit images per channel can be measured this way. The major measurement

elements of the parameters which can be recorded by the Fovea Pro 4.0 software are the following:

- **Area:** Series of pixels falling within the area of the calibrated pixel size and the binary object to be measured, which is modified by specific measuring. Unit of measurement: mm^2
- **Convex Area:** In the case of objects like cereal seeds lacking dents or nooks it is equal with the measured data of the area. Unit of measurement: mm^2
- **Length:** The longest string or straight line between any two points located on the perimeter of the object, meaning the maximal length dimension. Unit of measurement: mm
- **Breadth:** The maximal thickness dimension of the measured object, usually measured at right angles to the length. Unit of measurement: mm
- **X-Feret and Y-Feret diameters:** Horizontal and vertical dimension of the object measured with a slide calliper, meaning the horizontal and the vertical size of the object. Measured values are given in pixels.
- **Equivalent Diameter:** The diameter of a circle covering the same area as the object. Unit of measurement: mm
- **Inscribed Radius:** Radius of the circle which can be drawn inside the object. Unit of measurement: mm
- **Circumed Radius:** Radius of the circle which can be drawn around the object. Unit of measurement: mm
- **Perimeter:** Pixels located along the borderline. Unit of measurement: mm
- **Convex Perimeter:** Perimeter of the convex polygon bordering the object. Unit of measurement: mm
- **Aspect Ratio:** The ratio of length and breadth, meaning one aspect of the extension of the object.
- **Roundness:** It measures how the two-dimensional extension of an object differs from the shape of a circle. Its formula is the following: $4 \times \text{area} / \pi \times \text{maximal dimension}$
- **Form factor:** It emphasizes the extension of the perimeter related to the objective area. Its formula is the following: $4 \pi \times \text{area} / \text{perimeter}^2$.
- **Convexity and Solidity:** Solidity means the ratio of the perimeter of the object and the perimeter of a polygon closely fitted around it, while convexity means the ratio of the area of the object and the area of a polygon closely

fitted around it, measuring the differences between the object and the polygon (Rovner and Gyulai 2007).

- **Symmetry:** The symmetry of the object is encoded in the Fourier transformation of its borderline.
- **Radius Ratio:** Ratio of the radii of the circles that can be drawn within and around the object.
- **Elongation:** Ratio of the area of the object and the convex perimeter of the object (Emódi et al. 2015).
- **Intensity:** Mean intensity (mean brightness), minimal intensity, maximal intensity and the standard deviation of intensity are defined based on the pixel values measured on a 0 (black) – 255 (white) scale.
- **Mean red, green and blue colour values:** In the case of colourful images pixels have red, green and blue components, whose values range from 0 (black) to 255 maximal intensity levels. Data represent the values of all pixels located within the object.

2.3 Differentiation of domesticated einkorn from its wild ancestor

First I used the developed database to verify if the phenomenon published by Rovner and Gyulai (2007)² – that is the wild ancestor can be distinguished from domesticated einkorn by using a histogram – can be also witnessed in the first phase of the domestication process of einkorn being the ancestor of currently produced winter wheat.

Since my objective was to study domestication and to reveal genetic changes induced by this process, I used only samples produced in the same year, in order to eliminate environmental factors affecting the plants. As the accessions maintained by the Centre for Agricultural Research (Martonvásár) originated from the same field (being produced on neighbouring small plots), and production year effects could be also eliminated, the results purely reflect genetic information.

Data belonging to the different accessions have been copied from the database to a file. Five pieces of *T. monococcum* subsp. *aegilopoides* accessions and five pieces of *T. monococcum* subsp. *monococcum* accessions have been analysed, each containing 192 pieces of seeds. Since it is assumed that differences can be traced back to the

² Rovner I., Gyulai F. (2007): Computer-Assisted Morphometry: A New Method for Assessing and Distinguishing Morphological Variation in Wild and Domestic Seed Populations. In: *Economic Botany*, 154-172. p.

various levels of diversity and that human selection was based on the size of the grains, the area, perimeter, length and breadth of the seeds have been used to investigate the differences. I created histograms from the data of the different parameters one-by-one by using the program Microsoft Excel 2007, and separating accessions from the *aegilopoides* and the *monococcum subspecies* categories. Within the command “Data analysis ->Histogram” I selected the different dimension related parameters as input interval without setting any subdivisions. Then I set the diagram output demonstrating if there is a difference between the wild and the domesticated subspecies in the case of the given parameter. By selecting the data source of the diagram I set 2 for the value of the period after choosing “Moving mean” within the command “Creating trend line”. As a result, the trend line follows the frequency values of the different parameters, i.e. the trend line forming a unimodal or a multimodal histogram. By putting the polynomial trend line no. 2 next to it, the number of their intersection points show if the graph can be regarded multimodal or unimodal compared to the bell curve. If there are more than two intersection points, the graph can be regarded multimodal, while in the case of only two intersection points, the unimodal shape of the graph can be demonstrated.

2.4 Using histograms for the analysis of domestication impacts

Besides their unimodal or multimodal features histograms also carry other kinds of information. The distribution of the grains and the shift of the modus also demonstrate that – besides decreased diversity – which value intervals decreased, what was the direction of shift, and to what extent did the values of the different parameters changed as a result of domestication. The most expressive way to illustrate this is putting the histograms of the wild and the domesticated accessions in the same coordinate system.

2.5 Classification system suitable for the identification of an individual seed with unknown taxonomic background

Besides separating the wild and the domesticated accessions based on the developed database, I also aimed at going from the *subspecies* to the *varietas* level regarding the identification of the samples, and reaching individual seed level taxonomic classification beyond the identification of accessions.

I used the nearest neighbour classification method in Euclid space as the basis for the classification system. Although all the 17 einkorn accessions were included in the

analyses, some parameters have been selected, and after this accessions were analysed in a 42-dimension sample space, since calculations were made in the space of the measured parameters.

Regarding taxonomic classification, my objective was to identify unknown individual einkorn grains as precisely as possible through the development of the most accurate classification system, which can be used for the elaboration of an artificial intelligence based seed identification system. For this end, I searched for those parts of classification where further specifications could be carried out for greater reliability, since minimizing the occurrence of errors is essential both from the point of gene bank related and archaeological utilization.

Principal Component Analysis (PCA) proved to be the most appropriate method for defining which parameter should be more or less weighted in order to help the differentiation of the accessions. I performed this analysis by using the IBM SPSS program. I run the analysis without rotation and with no filtering, weighting or file distribution. Then I calculated the weight factors based on the PCA component matrix (the maximal values were regarded as weight factors for each parameter, meaning 8 factors above the level of 95% in this case).

3. Results

3.1 Elaborating the method of digital image processing for the species of *T. monococcum*

The objectives of my research were the morphometric description of *Triticum monococcum* seeds stored in the Hungarian gene banks and the elaboration of a recording method which can be used for such purpose. The database developed this way is the basis for the collection of seed size and shape related information (after the morphometric and biological selection of its parameters). The results of the analysis can help to better know and more accurately identify the einkorn accessions of archaeological sites and gene banks.

The image processing method used during the research was elaborated step-by-step by considering the needs of the different work phases and the witnessed hindering factors. The method developed this way is suitable for the rapid and accurate morphometric recording of the grains of einkorn and other *Triticum* species even in the case of larger sample sizes.

Elements required by the method:

- grain(s) of the analysed *Triticum* accession without the chaff,
- frame for mirror reflex digital cameras,
- 40 mm Micro f/2.8 objective with a view angle of 38° 50',
- circular flash,
- tripod adjustable between 85-1220 mm and equipped with a 360° horizontal and vertical rotation head,
- air-level and 3D spirit level that can be mounted on cameras,
- remote trigger,
- green A4 sized decoration sponge sheet,
- plotting paper,
- Photoshop and Fovea software.

Steps of the method:

Taking images

1. Putting einkorn grains and a strip of plotting paper on the decoration sponge sheet functioning as the background.
2. Joining the tripod, the objective and the circular flash mounted on that.
3. Putting and fixing the frame on the tripod, and fixing it in a proper position above the sample.

4. Setting the optical axle of the camera at right angles to the plane of the samples with the help of the air-level and the spirit level.
5. Settings by applying f/3.2 shutter, 2.5 seconds of exposure time, 100 ISO value and 40 mm of focus distance.
6. Taking the image by eliminating external sources of light and the own flash of the camera with the help of the circular flash and the remote trigger.
7. Saving images in RAW and TIFF formats with 24 bit depth.

Image processing

1. Opening the image file with the program Photoshop
2. Selecting seeds with the “Magicwand” tool
3. Inverting selection and deleting the background
4. Saving the image in 16 bit TIFF format
5. Opening the 16 bit image file with the program Fovea Pro 4.0
6. Calibrating the image according to the plotting paper
7. Performing measuring tasks

3.2 Developing a morphometric database of the seeds of *Triticum monococcum* accessions stored in Hungarian gene banks

42 different dimensional, morphological and colour related parameters of 2545 seeds of 17 *Triticum monococcum* accessions have been measured based on the images taken by the developed method (see Figure 1).

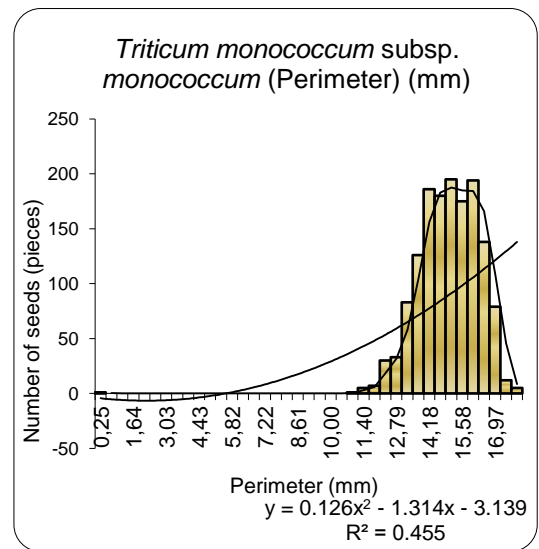
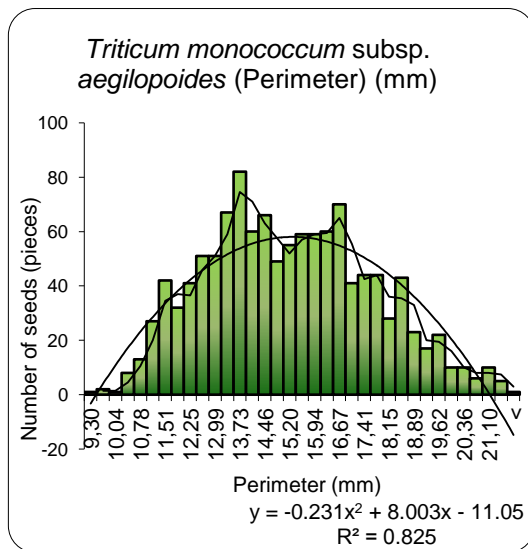
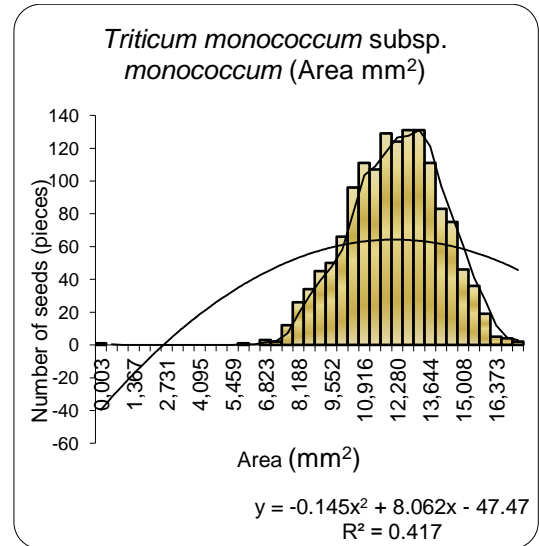
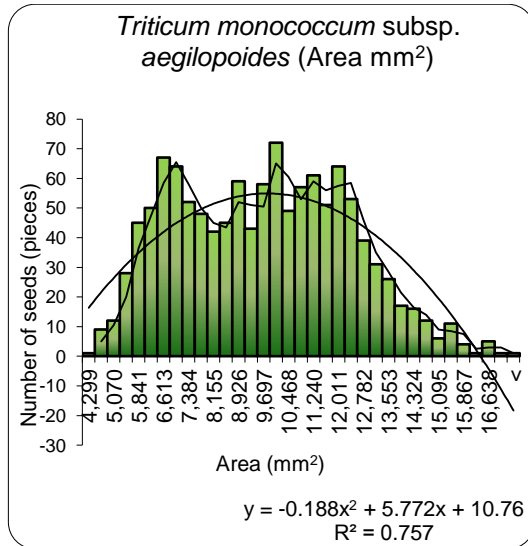


Figure 1: Image of *T. monococcum* subsp. *monococcum* cv. *Epeautre* before (left) and after (right) pre-processing

3.3 Differentiation of domesticated einkorn from its wild ancestor by using the morphometric database of seeds

One of the main objectives of developing the database was to verify that domesticated and wild plants can be unequivocally distinguished by the morphometric measuring of their seeds and the histograms developed from the

collected data during the analysis of a domesticated cultivated plant species and its direct wild ancestor, even in case they taxonomically belong to the same species. Since it is assumed that differences can be traced back to the various levels of diversity and that human selection was based on the size of the grains, the area, perimeter, length and breadth of the seeds have been used to investigate the differences (see Figure 2).



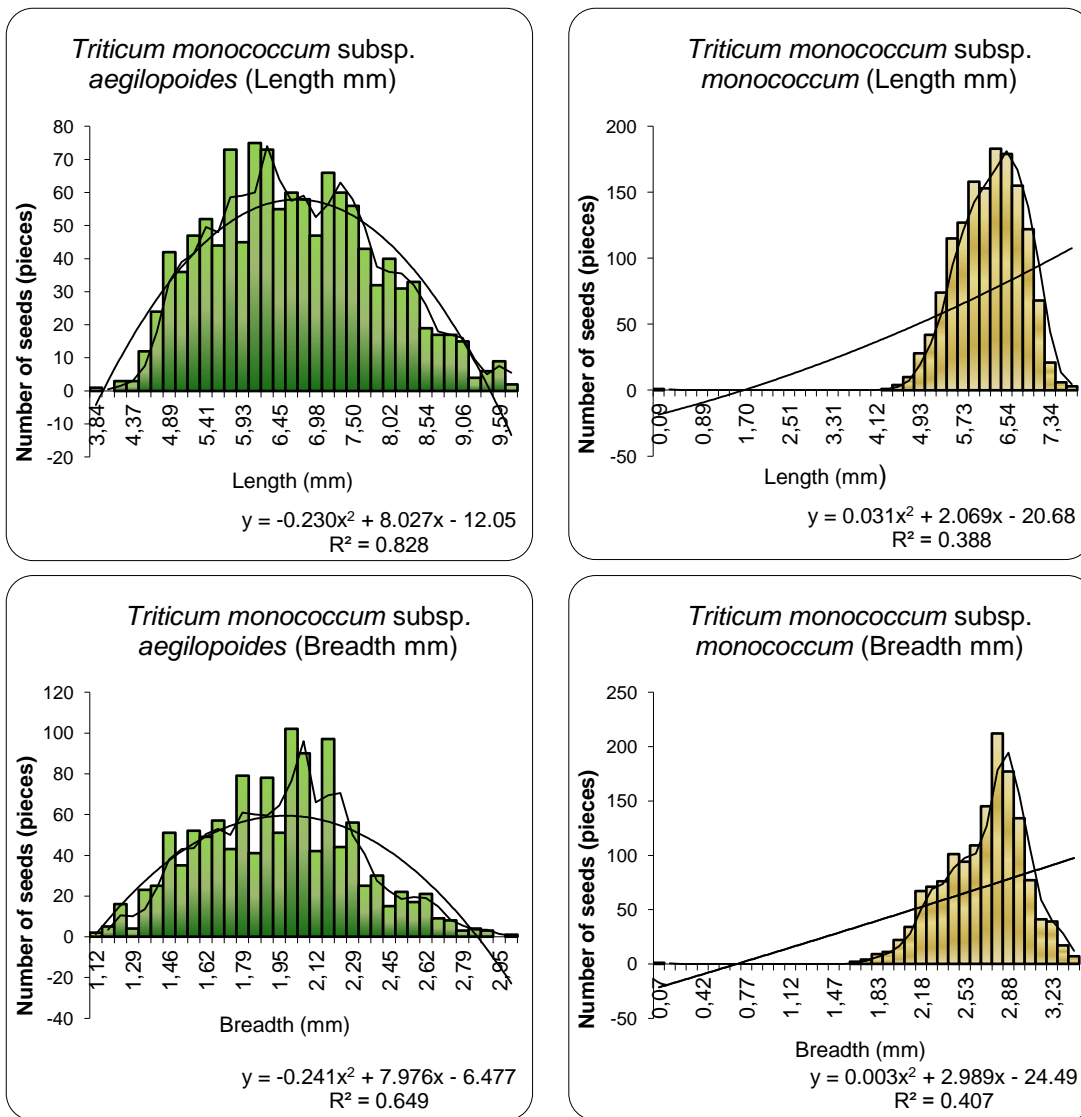


Figure 2: Multimodal histogram of wild einkorn (left) and unimodal histogram of domesticated einkorn (right)

The impacts of human selection, resulting in decreased diversity, are clearly visible on the histograms of all the four parameters. The histograms of *T. monococcum* subsp. *aegilopoides* cover wider intervals, the distribution of the grains is uneven, and similarly high seed numbers can be observed in the case of more measured values. This gives a multimodal appearance to the histogram. The reason behind this is that keeping a diverse gene pool is a tool of nature with the aim of preserving genes for future possibilities when they might be required for the survival of the given species (e.g. in the case of changing environmental conditions). Sensibility is increasing due to the growing intensity of natural selection pressures (e.g. stress impacts or climate change), and diversity can be used to counterbalance this phenomenon. During domestication humankind selected that segment from this gene pool which was regarded to be useful. From wild einkorn collected for food they kept

those seeds which were larger, making sowing and harvest easier. Consequently, the dimensional parameters of *T. monococcum* subsp. *monococcum* reflect a narrowed gene pool with a unimodal bell curve shaped histogram. This fact is equally confirmed by the data on area, length, breadth and perimeter. The unimodal histogram can be exactly distinguished from the multimodal one by drawing trend lines. If the polynomial trend line no. 2 intersects the 2 periodic trend line of the moving mean at more than twice, the histogram can be regarded multimodal, while in the case of only two intersection points the histogram is unimodal. According to this, it can be stated that the histograms developed by using the area, perimeter, length and breadth related data of *T. monococcum* subsp. *aegilopoides* seeds are multimodal, that is they can be regarded wild accessions; while the unimodal histograms of *T. monococcum* subsp. *monococcum* verify the process of domestication.

3.4 Analysis of the impacts of domestication by histograms

Dimensional parameters

Regarding the dimensional parameters histograms can be divided into three categories. The first category includes area, convex area and equivalent diameter (see Figure 3).

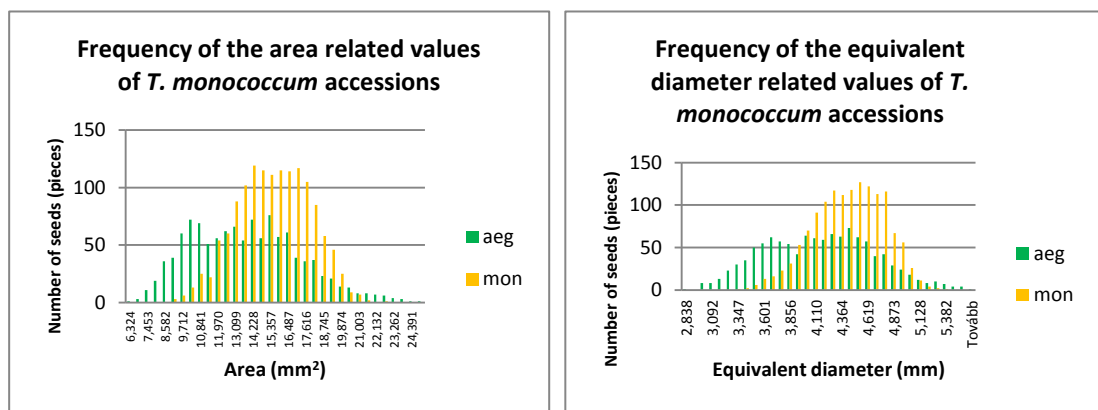


Figure 3: Histograms of the area and the equivalent diameter related data of wild and domesticated einkorn accessions

According to the histograms, it can be stated that the area related data of *T. monococcum* subsp. *aegilopoides* cover a wider range, and demonstrate a very diverse subspecies on the whole. On the contrary, *T. monococcum* subsp. *monococcum* covers a narrower range with higher seed numbers in the case of the typical values. This narrowed interval falls within the values of *aegilopoides*,

verifying that the area of *monococcum* seeds did not exceed the values of its wild ancestor during domestication, and it only became more uniform within the interval. Two modi can be observed on the multimodal histogram of *aegilopoides* from which the one belonging to the higher values overlaps with the modus of *monococcum* (14.22819 mm²), while the group belonging to the lower values have been already selected from the *monococcum* accession, and only a few seeds have such low area related values. The same is true for the value of the convex area closely correlating with the area and also for the data of the equivalent diameter.

The second group includes length, circumed radius, perimeter and convex perimeter (see Figure 4).

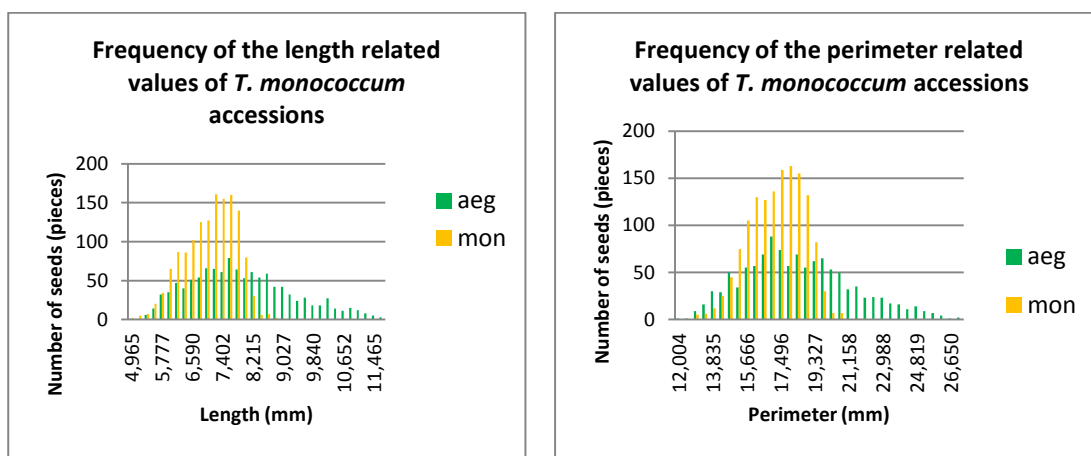


Figure 4: Histograms of the length and perimeter related data of wild and domesticated einkorn accessions

In the case of these parameters a more intense diversity loss can be observed. The values of *aegilopoides* cover a wider range, while *monococcum* grains are concentrated at the lower values of the interval, and after reaching and exceeding the modus the frequency of seeds suddenly decreases and ceases. In the case of *monococcum* no seeds can be found above certain length (8.9776 mm) and perimeter (21.2394 mm) values, where the grains of the wild subspecies still can be found in larger numbers. Since this trait presumably can be traced back to such behaviour of the length related value – which probably has not been affected by controlled human selection –, it is more likely that this is a natural evolutionary reaction originating from the size and breadth driven selection of the seeds. As the grains became wider and rounder, their potential length decreased. This was also confirmed by the values of seed breadth and inscribed radius belonging to the third group of dimensional parameters (see Figure 5).

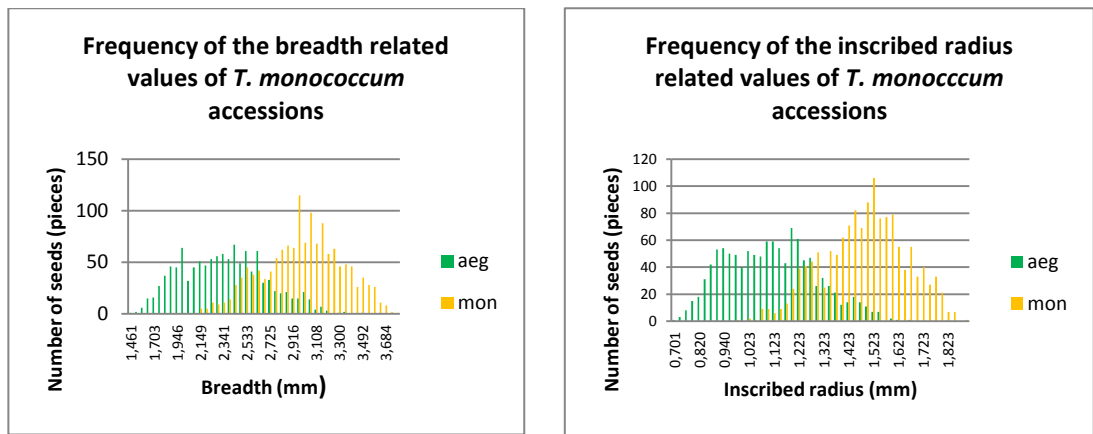


Figure 5: Histograms of the breadth and inscribed radius related data of wild and domesticated einkorn accessions

Here an evolutionary change resulting from the selection impact also can be unequivocally observed. The lowest values of *aegilopoides* cannot be found in the case of *monococcum* since they have been selected. Besides, the grains of *monococcum* reach such high breadth values that cannot be observed in the case of the wild *aegilopoides*, demonstrating that the domesticated subspecies exceeds its wild ancestor regarding these parameters. However, the two subspecies have not separated totally so far. The same statements are also true for inscribed radius correlating with the breadth of the grain. This also verifies that although *T. monococcum* subsp. *monococcum* is a domesticated subspecies, it is still in the first phase of domestication.

Morphological parameters

The phenomenon described above can be also witnessed in the case of the morphological parameters. The histograms of form factor, X-feret diameter, roundness and radius ratio also show that the values of *aegilopoides* are lower, that is their grains do not really resemble the form of a perfect circle. There is a value interval in the case of both subspecies which is not typical of the other one at all, and the modi of the *aegilopoides* subspecies also fall into this range. Although some *aegilopoides* grains can be found at the modus values of *monococcum* in the case of all the parameters, their ratio is negligible (see Figure 6).

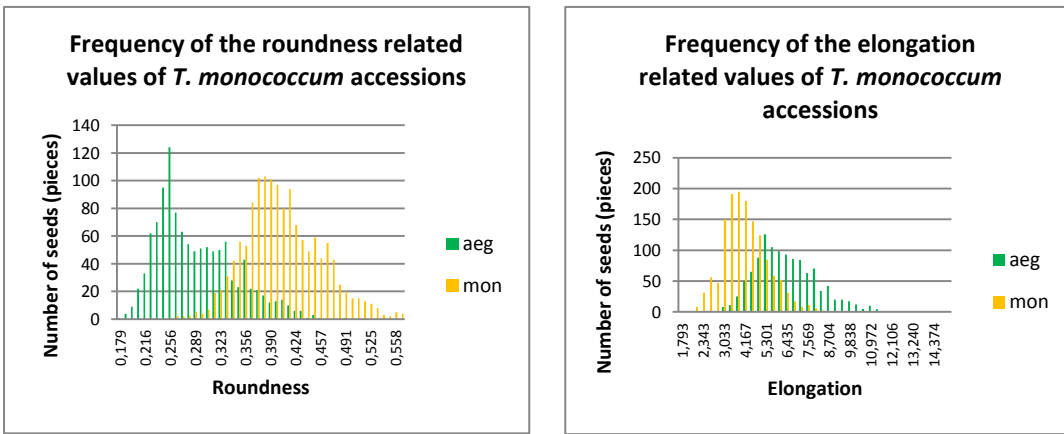


Figure 6: Histograms of the roundness and elongation related data of wild and domesticated einkorn accessions

From the point of morphology it is an interesting result that the above mentioned facts are true for roundness, while in the case of elongation this sharp separation cannot be totally observed. The histogram of elongation shows that the highest values of *monococcum* are quite close to the modus of *aegilopoides*, while when speaking about roundness the two modi are perfectly separated, and are located far from each other. Accordingly, a more significant selection impact with smaller evolutionary changes can be witnessed in the case of elongation, while evolutionary changes are more typical of roundness.

From the morphological parameters aspect ratio (the ratio of length and breadth) also needs to be mentioned (see Figure 7).

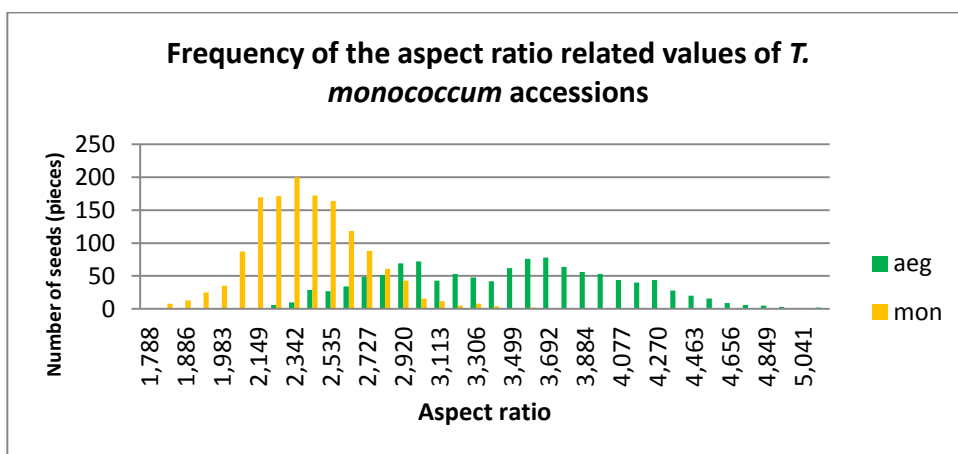


Figure 7: Histogram of the aspect ratio related data of wild and domesticated einkorn accessions

Aspect ratio is primarily determined by genetic information, and environmental factors affect its values only in extreme situations, therefore it provides a good basis

for collecting genetic information, reflects changes related to them, and is also important during the identification of the samples. The histogram shows that the ratio of length and breadth significantly narrowed during domestication, and is demonstrated through the lower values compared to the wild ancestor, which are becoming even lower due to the process of domestication. This latter statement is verified by the fact that on the histogram of domesticated einkorn values never witnessed in the case of the wild ancestor also can be found due to the seeds included in the samples.

Intensity and colour parameters

Significant difference cannot be detected between the two subspecies according to the intensity and colour parameters. Typically, the intervals of the histograms are the same, the distribution of *aegilopoides* is more even – similarly to the other types of parameters –, and *monococcum* covers narrower intervals with higher values (see Figure 8).

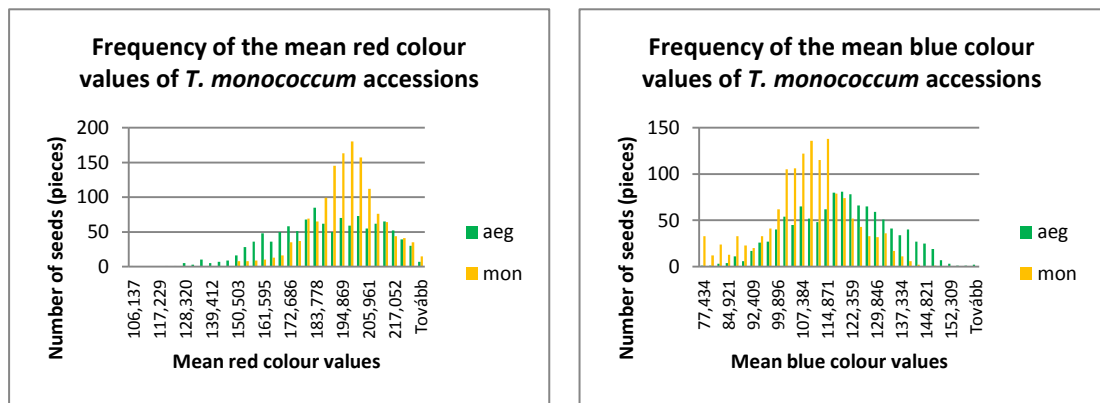


Figure 8: Histograms of the mean read and blue colour values of wild and domesticated einkorn accessions

All in all, it can be stated that by comparing the morphometric histograms of a domesticated and a wild einkorn accession, the impacts of domestication on the morphology of the grains can be traced. The method verified that it also can be a tool for the differentiation of wild and domesticated plants.

3.5 Identification and taxonomic classification of an individual unknown seed at the *varietas* level

It was my objective to investigate how accurate taxonomic identification can be achieved beyond the species level. Besides, it could be beneficial from the point of

practical application if the method also can be used for the identification of even individual seeds. It can be an important consideration in gene banks where in the case of inhomogeneous newly arrived samples seeds could be identified one by one, this way preventing false taxonomic identification and storage. The identification of individual seeds can also have a great importance when speaking about archaeological samples. Excavated charred samples often contain only a few einkorn grains, therefore only methods suitable for the correct identification of even one individual seed can be used for taxonomic identification beyond the species level. Since none of the parameters had data according to which the accessions could be separated without any overlapping at the sample element level, a complex method needed to be elaborated. The steps of the methodology elaborated during the analysis are the following:

Previous calculations:

1. Calculations are made in the space of measured parameters, meaning a 42-dimension sample space
2. Development of a summary table on the mean values of the 17 analysed samples in the case of the 42 parameters
3. Marking minimum and maximum values and the ratio of the interval and the maximum value
4. By specifying a limit value of 30% for the interval of the measured parameters of the different samples, the number of dimensions decreases to 25, and accuracy can be increased maximum by 10%
5. Normalization of parameter values (to 1)
6. Calculation of the quadratic deviation of the samples from the mean value

Calculations in the case of unknown samples:

1. Measuring the parameters of unknown samples
2. Performing calculations listed under points 5 and 6 of previous calculations in the case of the unknown sample
3. Calculating the distance of the unknown sample from the different classes in a 42- or 25-dimension space (sum of the quadratic difference of the parameters or its weighted value)
4. Identification of the minimum value calculated in the previous point

5. The unknown sample can be identified by the class showing the minimum value

With the help of the above mentioned method an unknown sample can be taxonomically identified with an accuracy of 85% at the *varietas* level (M5). This way an individual unknown einkorn seed can be accurately, rapidly and objectively identified by informatics methods supporting archaeological identification and gene bank storage.

After calculating the quadratic deviation of the samples from the mean value the resulting values have been also multiplied by the weighting values of the Principal Component Analysis in the case of each parameter. The lowest value of the sum of the accessions from the resulting table shows the accession belonging to the unknown sample. By testing the method with identified seeds selected from the accessions – but treated as an unknown sample – I did not get more accurate results. Those seeds wrongly identified by the original method were put in the wrong accession also by PCA weighting. Consequently, this step can be skipped since although the result may make classification more accurate, it does not create a significantly more accurate possibility for classification.

3.6 New and novel scientific results

1. During my research I elaborated a methodology based on commercially available devices for the morphometric assessment of the grain crop of *T. monococcum* L. enabling for the description and the recording of data of einkorn accessions located in gene banks or at archaeological sites.
2. By using the elaborated methodology I made an image database of *T. monococcum* L. accessions stored in the gene banks of Hungary, and belonging to 17 different *varietas* and *convarietas* categories. I recorded the data of 42 different parameters of the described 2545 einkorn seeds from which 11 were dimensional, 20 were morphological and 11 were colour and intensity related data. This database can be the basis for further identification and assessment research.
3. Based on the developed database I verified that independent of the taxonomic level it is also typical of the subspecies level that the histograms of dimensional seed data are multimodal in the case of wild populations, while they are unimodal for domesticated ones.
4. I described the morphometric impacts of domestication in the case of the species of *T. monococcum*.
5. I elaborated a nearest neighbour based taxonomic classification method in Euclid space by using the morphometric database, enabling for the *varietas* level classification of an individual *T. monococcum* seed. In order to make the method more accurate, I performed the calculations of the nearest neighbour based classification method also in weighted Euclid space, but according to the results this method did not increase accuracy significantly. Consequently, this step can be skipped since it does not create a significantly more accurate possibility for classification.

4. Conclusions and recommendations

My research revealed that the methodology of morphometry is suitable for data collection and analysis based on crops and for the taxonomic classification of unknown accessions also at the *varietas* level in the case of the species of *T. monococcum* L.

The presented method is suitable for the analysis of einkorn and other cereals, and can be adapted to the description of further foreign einkorn accessions and *Triticum* accessions stored in the gene banks of Hungary. After minimal modifications the method may be suitable also for the description of the crops and seeds of other plant species by adjusting it to their specific size and colour traits which are different from that of *Triticum* species. Recording the digital visual and morphometric data of the seeds can become an important step of gene bank storage standards, and collected data can be published through Internet databases providing a basis for several further researches in the different fields of natural sciences.

Besides using them in the presented analyses studying the impacts of domestication and evolution, collected data are also suitable for extensive population genetics research. In the case of controlled experiments the morphological trait related impacts of certain environmental factors, controlled crossings and genetic interventions can be also studied. In addition, by using other species domestication related and evolutionary changes can be further researched. By analysing other species, varieties and accessions of the *Triticum* genus located at different levels of domestication, the extent of the various impacts could be also defined in the case of each domestication level.

In addition, the calculation system enabling for the individual identification of the seeds could be further extended and specified. By extending the databases the method could become the basis for developing an artificial intelligence based automated taxonomic identification system, which would provide unique possibilities for recording the visual and numeric data of large size samples and for their taxonomic classification in the case of gene bank accessions and archaeological findings.

5. Scientific publications covering the topic of the thesis

Journal articles

Journal articles with impact factor

GYULAI G., I. ROVNER, S. VINOGRADOV, B. KERTI, A. EMŐDI, E. CSÁKVÁRI, A. KERÉKES, Z. MRVACSIK, F. GYULAI (2015): Digital seed morphometry of wild and crop dioecious plants - development and usefulness of the seed diversity index. *Seed Science and Technology*, 43, pp. 492-506 (ISSN 0251-0952)

Proofread journal articles without impact factor

In foreign language

EMŐDI A., TIRCZKA I., HARTMAN M., DOBOLYI CS., SEBŐK F., GYULAI F. (2014): Two Ancient Wheat Species, Possibilities for the Production of Einkorn and Emmer in Organic Farming. *Annals of Faculty Engineering Hunedoara – International Journal of Engineering* 12 (2), pp. 43-50 (ISSN: 1584-2665).

MRAVCSIK Z., GYULAI F., VINOGRADOV S., EMŐDI A., ROVNER I., GYULAI G. (2015): Digital seed morphometry for genotype identification – Case study of seeds of excavated (15th century Hungary) and current vinegrape-*Acta Botanica Hungarica* 57:(1-2) pp. 169-182. (ISBN 1588-2578)

In Hungarian

GYULAI F., EMŐDI A., MRAVCSIK Z., PÓSA P. (2013): Reconstruction of the agricultural cultivated environment of modern times by using the example of the archaeological findings of Sárospatak, *GESTA*, 2013. XII.: 67-71.

EMŐDI A., GYULAI F., MRAVCSIK Z., GYULAI G., SZ. VINOGRADOV, SZABÓ T. A., I. ROVNER (2014): Digital seed morphometry I. Analysis of produced einkorn varieties and landraces (*T. m. ssp. monococcum*) – *Plant production* 63: pp. 61-70.

MRAVCSIK Z., GYULAI G., EMŐDI A., GYULAI F., SZ. VINOGRADOV, I. ROVNER (2014): Seed morphometric analysis of archaeological and recent grape seed samples – *Kertgazdaság: Professional journal of the University of Horticulture and Food Industry, the Hungarian Horticultural Council and the Hungarian Society of Horticultural Science* 46:(4) pp. 27-33.

PÓSA P., EMŐDI A., SCHELLENBERGER J., HAJDÚ M., MRAVCSIK Z., GYULAI F. (2014): Preliminary report on the processing of plant remains found in a well near Miskolc-Hejő from the Scythian era. *GESTA MISKOLC* 13: pp. 3-18. (ISBN 1417-2569)

EMŐDI A., GYULAI G., SZ. VINOGRADOV, MRAVCSIK Z., GYULAI F., I. ROVNER (2015): Digital seed morphometry II. Seed morphometric characterization of the two subspecies (*T. m. ssp. aegilopoides*, *T. m. ssp. monococcum*) of einkorn (*Triticum monococcum* L.) – *Plant production* 64: pp. 23-37.

Conference publications

Hungarian (complete)

EMŐDI A., GYULAI F., MRAVCSIK Z., KERTI B., HIDVÉGI N., VINOGRADOV S., SZABÓ T. A., ROVNER I., GYULAI G. (2014): Molecular, production and digital seed morphometric analysis of einkorn (*Triticum m. monococcum*) varieties. In: Veisz O. (ed.): 20th Plant Breeding Scientific Days, Budapest, 18th March 2014, 125-129. ISBN: 978-963-8351-42-5.

GYULAI G., KERTI B., VINOGRADOV S., EMŐDI A., MRAVCSIK Z., GYULAI F., ROVNER I. (2014): Seed morphometric analysis of dioecious plants. In: Veisz O. (ed.): 20th Plant Breeding Scientific Days, Budapest, 18th March 2014, 165-169. ISBN: 978-963-8351-42-5.

MRAVCSIK Z., GYULAI F., EMŐDI A., KERTI B., VINOGRADOV S., HIDVÉGI N., ROVNER I., GYULAI G. (2014): Molecular and digital seed morphometric identification of archaeological grape (*Vitis vinifera*) seed findings. In: Veisz O. (ed.): 20th Plant Breeding Scientific Days, Budapest, 18th March 2014, 319-324. ISBN: 978-963-8351-42-5.

Hungarian (abstract)

PÓSA P., MRAVCSIK Z., GYULAI G., EMŐDI A., GYULAI F. (2013): Archaeological findings in plant breeding – Remains of cultivated plant species and varieties at the excavation sites of Sárospatak from the early modern times (16-17th centuries), 19th Plant Breeding Scientific Days, Abstracts, 131., ISBN 978-963-9639-50-8, (07.03.2013)

EMŐDI A., PÓSA P., MRAVCSIK Z., SZOLNOKI L., DARÓCZI-SZABÓ L., GYULAI F. (2013): Remains of the plant diversity of the Hungarian Ishmaelites from the middle ages, Abstracts of the presentations of the 4th SzaKKKör Conference, 30.p., In: Takács Márton (editor), Gödöllő, Hungary, 22.04.2013, ISBN:978-963-269-346-0

PÓSA P., EMŐDI A., MRAVCSIK Z., PÓSA B., GYULAI F. (2014): The latest results of plant diversity analyses at the excavation sites of Sárospatak from the early modern times, 19th Bolyai Conference. Budapest, 22-23th March 2014 Abstracts, 46-47.

HAJDÚ M., PÓSA P., EMŐDI A., SCHELLENBERGER J., MRAVCSIK Z., GYULAI F. (2014): Preliminary report on the processing of plant remains found in a well near Miskolc-Hejő from the Scythian era. Archeometria, cognitive and socio-archaeological conference, Miskolc, 3-4th April 2014. Abstracts, 18-19.

In foreign language

GYULAI F., SZOLNOKI L., DARÓCZI-SZABÓ L., EMŐDI A. (2013): Plant remains of Ismaeliti from Medieval Hungary, 16th Conference of the International

Work Group for Palaeoethnobotany, Abstracts book. pp. 125-126., 17-22th June 2013, Thessaloniki, Greece.

EMŐDI A., GYULAI F., MRAVCSIK Z., HIDVÉGI N., KERTI B., VINOGRADOV S., GYULAI G., ROVNER I. (2013): Morphometry of einkorn (*Triticum m. monococcum*) seed populations for providing digital tools for cultivation and variety registration, 2nd Conference of Cereal Biotechnology and Breeding (CBB2), 5-7th November, Budapest, Hungary. 66.p.

MRAVCSIK Z., GYULAI F., EMŐDI A., KERTI B., HIDVÉGI N., VINOGRADOV S., GYULAI G., ROVNER I. (2013): Measuring diversity: morphometric monitoring of variation in wheat (*Triticum ssp.*) seed populations, 2nd Conference of Cereal Biotechnology and Breeding (CBB2), 5-7th November, Budapest, Hungary. 67.p.

MRAVCSIK Z., F. GYULAI, A. EMŐDI, G. GYULAI, S. VINOGRADOV, I. ROVNER (2014): Genotype identification of ancient vinegrape (*Vitis v. vinifera*) seed remains (15th CENT. Hungary) by digital morphometry, Advances in Plant Breeding and Biotechnology Techniques, Book of abstracts, 91-92. p., (Pannonian Plant Biotechnology Association) 28-29th April 2014, Mosonmagyaróvár, Hungary, ISBN: 978-963-89129-5-4

EMŐDI A., F. GYULAI, Z. MRAVCSIK, B. KERTI, N. HIDVÉGI, S. VINOGRADOV, A. T. SZABÓ, I. ROVNER AND G. GYULAI (2014): Variametric analysis of einkorn (*Triticum monococcum ssp. monococcum*) seed populations measuring of genetic distances using phenovariation, Advances in Plant Breeding and Biotechnology Techniques, Book of abstracts, 82-84. p., 28-29th April 2014, Mosonmagyaróvár, Hungary

EMŐDI A., VINOGRADOV S., GYULAI G., PÓSA P., MRAVCSIK Z., ROVNER I., GYULAI F. (2015): Digital seed morphometry of the two subspecies of ancient einkorn (*T. m. aegilopoides* and *T. m. monococcum*), Archéométrie, Programme et résumés, 171-172. p., 20^e Colloque d'archéométrie du GMPCA, 27-30 avril 2015., Besancon, France.

MRAVCSIK Z., GYULAI F., VINOGRADOV S., PÓSA P., EMŐDI A., GYULAI G., ROVNER I., (2015): Morphometrical identification of excavated (15th century Hungary) and current vinegrape (*Vitis v. vinifera*) varieties, Archéométrie, Programme et résumés, 176. p., 20^e Colloque d'archéométrie du GMPCA, 27-30 avril 2015., Besancon, France.