ASSESSMENT OF *LATHYRUS* SPECIES ACCESSION VARIABILITY USING VISUAL AND STATISTICAL METHODS

BOGNA ZAWIEJA[1], WOJCIECH RYBIŃSKI[2], KAMILA NOWOSAD[3] AND JAN BOCIANOWSKI[1*]

[1]*Department of Mathematical and Statistical Methods, Poznań University of Life Sciences,*
Wojska Polskiego 28, 60-637 Poznań, Poland
[2]*Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland, Strzeszyńska 34, 60-479 Poznań, Poland*
[3]*Department of Genetics, Plant Breeding and Seed Production, Wrocław University of Environmental*
and Life Sciences, Grunwaldzki 24A, 53-363 Wrocław, Poland
[*]*Corresponding author's email: jboc@up.poznan.pl*

Abstract

Underestimated lesser-known species often prove to be a very attractive object of research. For example, they adapt very well to various marginal growing conditions, e.g. highlands, arid areas, salt-affected soils, etc. Some species of the genus *Lathyrus* may be examples of such crops. In this study the species and their accessions were compared as regards to coefficients of variation. A considerable degree of variability is important to breeders of new varieties, since the chance of obtaining new cultivars significantly different from established varieties is thus increased. Thus the coefficients of variation were compared using both visual and statistical methods, with those highest in value being of greatest interest. The highest variability of 100 seeds weight was noticed in species *L. aphaca*, *L. clymenum*, *L. hirsutus*.

Key words: Andrews curves, Coefficient of variation, Linear discriminant analysis, Nonlinear kernel discriminant analysis.

Introduction

For thousands of years people have cultivated a diverse range of plant species. At least six thousand species have been used for different purposes. Most food has been produced on the basis of only a few staple crop species. Nevertheless, lesser-known species have not been disregarded. Unfortunately, most research has focused on well-known species, while relatively little attention has been given to minor (underutilized or neglected) crops. Such crops have often adapted very well to various marginal growing conditions, such as highlands, arid areas, salt-affected soils, etc. (Campbell, 1997).

Some species of the genus *Lathyrus*, which have been such lesser-known species, may be examples of such crops. *Lathyrus* species exhibit a considerable potential for crop rotation, improving soil physical conditions, reducing disease and weed infestation, thus reducing production costs (Vaz Patto *et al.*, 2006). Moreover, their seeds are characterized by high protein contents. The genus *Lathyrus* comprises approximately 200 species and subspecies found both in the Old and the New World. Representatives of this genus are found mainly in temperate climates. There are about 52 species originated from Europe, 30 from North America, 78 from Asia, 24 are native to tropical East Africa, and 24 - temperate South America (Asmussen & Liston, 1998).

Despite the presence of so many known species, only grass pea (*Lathyrus sativus* L.) has been grown for food. While this species has already been used in the Neolithic times, at present it is considered as a model crop for sustainable agriculture (Vaz Patto *et al.*, 2006). Some lesser-known species are cultivated both for animal feed and food, for example *L. cicera*, *L. clymenum*, *L. ochrus* (for grain, but mainly for forage production) and *L. tingitanus* as forage species. Other species, especially *L. hirsutus* and *L. aphaca*, have been only occasionally grown for human consumption as well as other purposes.

In this paper variability of accessions of seven *Lathyrus* species was investigated as regards five traits (pod length, pod width, number of seeds in the pod, weight of seeds from the pod and 100 seeds weight). The coefficient of variation was used as a measure of this variability. Several multidimensional visual statistical methods as well as univariate statistical methods were used to test uniformity. The aim was to evaluate uniformity of accessions mainly within a given species and a given trait, but also within all tested species and traits.

Considerable variety between representatives of one species is important to breeders of new cultivars, because it increases the chance of obtaining varieties significantly different from established ones. On the other hand, variability between plants of one variety should be as limited as possible.

Material and Methods

Data: The research material consisted of seven *Lathyrus* collection species: *L. sativus* (S), *L. cicera* (C), *L. clymenum* (CL), *L. ochrus* (O), *L. tingitanus* (T), *L. hirsutus* (HS) and *L. aphaca* (AP). The seeds were obtained from the Gene Bank in Gatersleben (Germany), except for *L. sativus*, which was derived from the collection of grain legumes at the Institute of Plant Genetics, the Polish Academy of Sciences in Poznań. The seeds of all species were sown in plots (5.2 m^2) situated in the Experimental Field of the Institute of Plant Genetics in Cerekwica (52°31'16''N, 16°41'30''E). The field trial was established in the random block design with fifteen replications. On the plots the seeds were sown in rows of 70 cm in width and 25 cm spacing between single seeds placed in rows. In the phase of full maturity 3 pods from the main stem of randomly selected plants were harvested from each plot. Species variability was analyzed in terms of five quantitative traits: pod length (cm), pod width (cm), the number of seeds per pod, weight of seeds per pod (g) and 100 seeds weight (g). The coefficients of variation are presented in Fig. 1.

Methods

1. Coefficient of variation (CV): This coefficient is the ratio of standard deviation to the mean in the tested population expressed as a percentage. It is a simple and comprehensive absolute measure of variability for traits in the researched population (Zawieja & Pilarczyk, 2005).

2. Visual statistical methods: Modified Andrews curves were used to compare genotypes and species as regards their variability in all analyzed traits. Next these curves were used to compare variability in traits of tested species (Khattree & Naik, 2002). These curves are designated for the interval $[-\pi, \pi]$.

The Linear Discriminant Variables Analysis (Friedman, 1989) and Nonlinear Kernel Discriminant Analysis (Mika *et al.*, 1999) were used to visualize differences between species as regards coefficients of variation. Two kernel functions were used: sigmoid $k(x, y) = \tanh(ax'y + c)$ (Schölkopf & Smola, 2002) and Chi-Square $k(x, y) = 1 - \sum_{i=1}^{n} \left\{ (x_i - y_i)^2 / \left[\frac{1}{2} (x_i + y_i) \right] \right\}$, where $x = (x_i)$ and $y = (y_i)$ denote vectors of observations, a and c – constants (a and c are selected so that the kernel matrix was positive definite), n – the number of observations. Kernel functions allowed to present on a plane different curvature of space. In this

paper the data were normalized before analysis, thus the parameters $a = 1$ and $c = 0$.

3. Test of hypothesis on equality of coefficients of variation: In the literature many methods have been proposed for this purpose. For example, such methods were presented in studies by Bennett (1976), Miller & Felltz (1997), Forkman & Verrill (2007) and Zawieja *et al.,* (2012). In all these methods the assumption CV ≤ 33% has been fulfilled. Moreover, if the number of observations in the analyzed populations was too small (fewer than 5), then the power of Bennett's as well as Miller's methods was very small (close to zero) and then type II error (which occurs when the null hypothesis is false, but erroneously fails to be rejected) was very large as it was shown by Zawieja & Szczepańska-Álvarez (2017). In this study the Miller method (Miller & Felltz, 1997) was used.

In practice it should be verified whether populations (accessions) are uniform to a similar degree. If CV was very small, it meant that a given population was uniform to a very high degree. However, if the coefficient of variation was very high, it meant that a given population was not stable and if that result was repeated for all traits, such a population should not be taken into account in further breeding work.
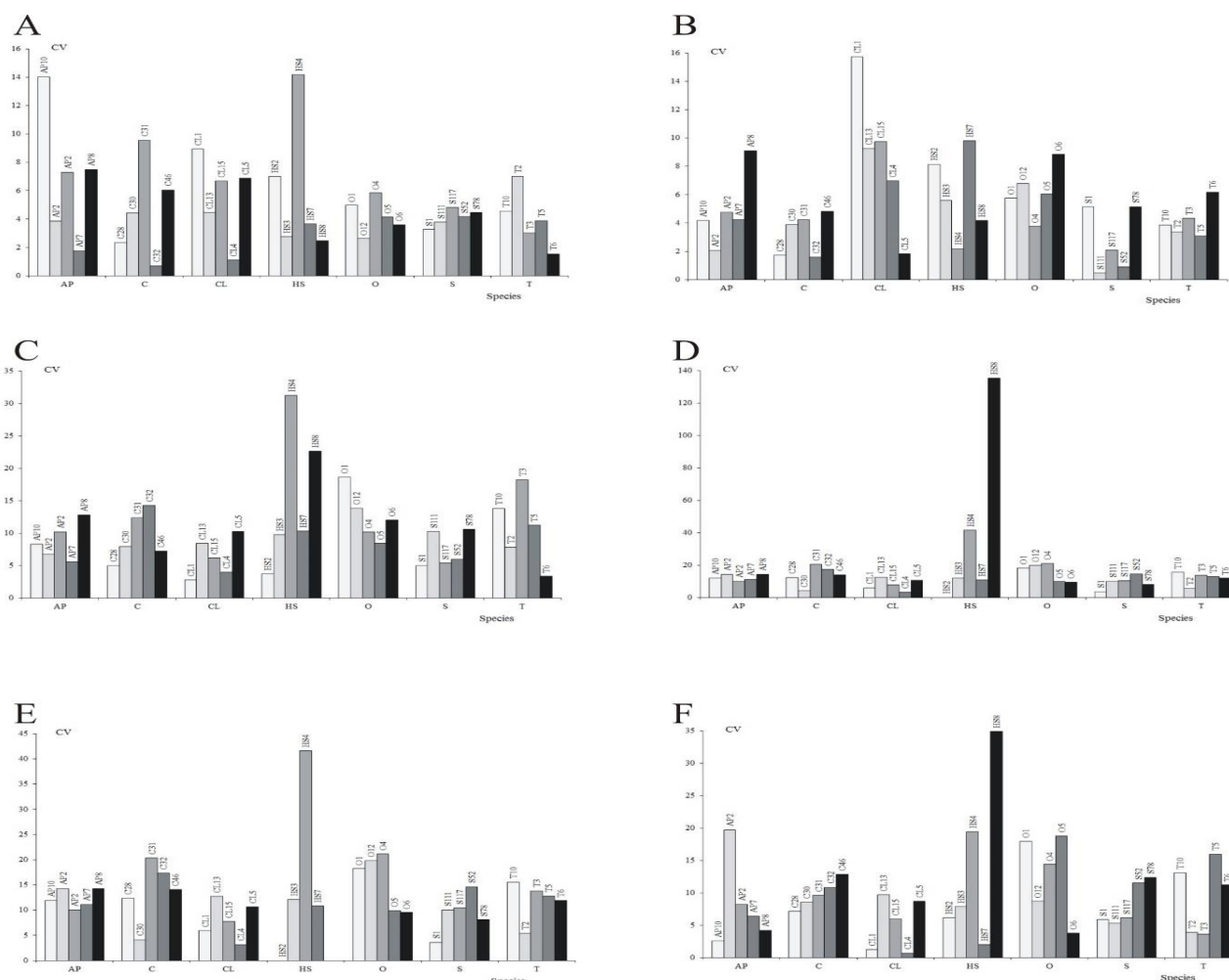

Fig. 1. Coefficients of variation for all tested objects (S – *L. sativus*, C – *L. cicera*, CL – *L. clymenum*, O – *L. ochrus*, T – *L. tingitanus*, HS – *L. hirsutus*, AP – *L. aphaca*): A – pod length, B – pod width, C – number of seeds in the pod, D – weight of seeds from the pod, E – weight of seeds from the pod except for genotype HS8, F – 100 seeds weight.

Results and Discussion

The coefficients of variation for the tested accessions ranged from 0% to 135.5% (Fig. 1). For pod length (LP) and pod width (WP) the CVs were smaller (from 0.5% to 16%) than those for the other traits (from 0% to 135.5%). In turn, a large range of CVs was shown for the AP, C, CL and HS accessions as regards pod length. Similarly, in the case of pod width for accession CL a large range of CVs was recorded. The CVs for HS was the most varied for seed traits (seed number from the pod, seed weight from the pod and 100 seeds weight).

Accessions were compared as regards the analyzed traits using the Andrews curves (Fig. 2A). These curves for accessions HS8 and HS4 differed markedly from the other curves. It may have meant that variability of these accessions was greater than that of other accessions (compare Fig. 1). Next Andrews's curves as a function of traits were considered and then each of them was the individual shape (Fig. 2B). Among them the curve for seed weight from the pod was exceptional, while the curves for the number of seeds from the pod and 100 seeds weight were the most similar (Fig. 2). The curves representing pot length and pot width are also very similar.

Moreover, analysis was carried out using Andrews's curves for traits of each tested species separately (Fig. 3). In this analysis curves representing pod length and pod width in *L. cicera*, *L. hirsutus*, *L. ochrus*, *L. sativus* and *L. tingitanus* were similar. In turn, for *L. ochrus*, *L. sativus* and *L. tingitanus* the curves for seed weight from the pod, seeds number and 100 seeds weight were similar. The trait of seed weight in the pod was highlighted in accessions *L. hirsutus* and *L. cicera*.

Finally, Andrews curves were used to demonstrate differences between accessions in terms of selected traits (Fig. 4). The curves for *L. hirsutus* were exceptional for all traits, but mainly as regards seed traits. Besides, *L. clymenum* was singled out in terms of pod width and *L. ochrus* as regards pod width, the number of seeds from the pod and 100 seeds weight.

The LDA and NKDA method were used to verify whether species could be separated in terms of CVs for all analyzed traits. The application of the LDA method yielded no satisfactory results (Fig. 5). The points representing the accessions of all species were quite mixed. On the basis of the Wilks' lambda statistics, it may be concluded that there is no reason to reject the hypothesis that all discriminant values are equal to zero (Table 1). Hence the groups could not be distinguished using linear discriminant variables, while the use of the NKDA method allowed to separate groups. Some of the species were separated. The application of the kernel sigmoid resulted in a division into

five groups (Figs. 6 and 7). Two species were classified to the first group from the left: *L. tingitanus* and *L. clymenum*. Only *L. aphaca* was classified to the second group. Two species were included in the third group: *L. cicera* and *L. sativus*. Finally, *L. ochrus* and *L. hirsutus* produced separate groups. Four groups could be visualized when the Chi-square kernel was applied. Separate groups were formed by *L. clymenum* and *L. hirsutus*, the second group comprised *L. cicera*, *L. aphaca* and *L. sativus*, whereas the last group contained *L. tingitanus* and two accessions of *L. ochrus*. A different curvature of the space resulted in the separation of *L. tingitanus* and *L. clymenum*.

The hypothesis on the equality of the coefficients of variation (CVs) was verified. Since the CVs of HS4 and HS8 in the case of seed weight from the pod as well as CV of HS8 in the case of 100 seeds weight exceeded 33%, these two accessions were not taken into account. These objects should be considered as non-uniform without testing. Most accessions of the tested species were uniform to a similar degree (Table 2).

Large variability of traits may be useful when developing a new variety. Nevertheless, new cultivars should be as uniform as known uniform varieties (that is an aspects in the criterion for varieties in the Distinctness, Uniformity and Stability studies). Thus new cultivars might be created from uniform material. In this study the variability and uniformity of accessions of *Lathyrus* species was investigated. For this purpose it was decided to apply an established statistical method (the Miller method to compare CVs) as well as visual methods (Andrews curves, LDA and NKDA), previously not used for this purpose. The Miller method was proposed by Forkman and Verrill (2007), Zawieja and Pilarczyk (2005), Zawieja *et al.,* (2012, 2016), Zawieja and Szczepańska-Álvarez (2017). The proposed visual methods made it possible to assess variability of the objects in terms of all analyzed traits and immediately exclude non-uniform cases as well as identify those that were uniform to a similar extent.

Table 1. Characteristics of the discriminant variables in LDA

Discriminant variable	The relative power of discriminant variable	*p*-value of Wilk's lambda statistic
1	54.3747	0.18927
2	26.1223	0.56039
3	16.1752	0.76617
4	2.78841	0.95768
5	0.53935	0.88194

Table 2. *p*-value of the Miller test.

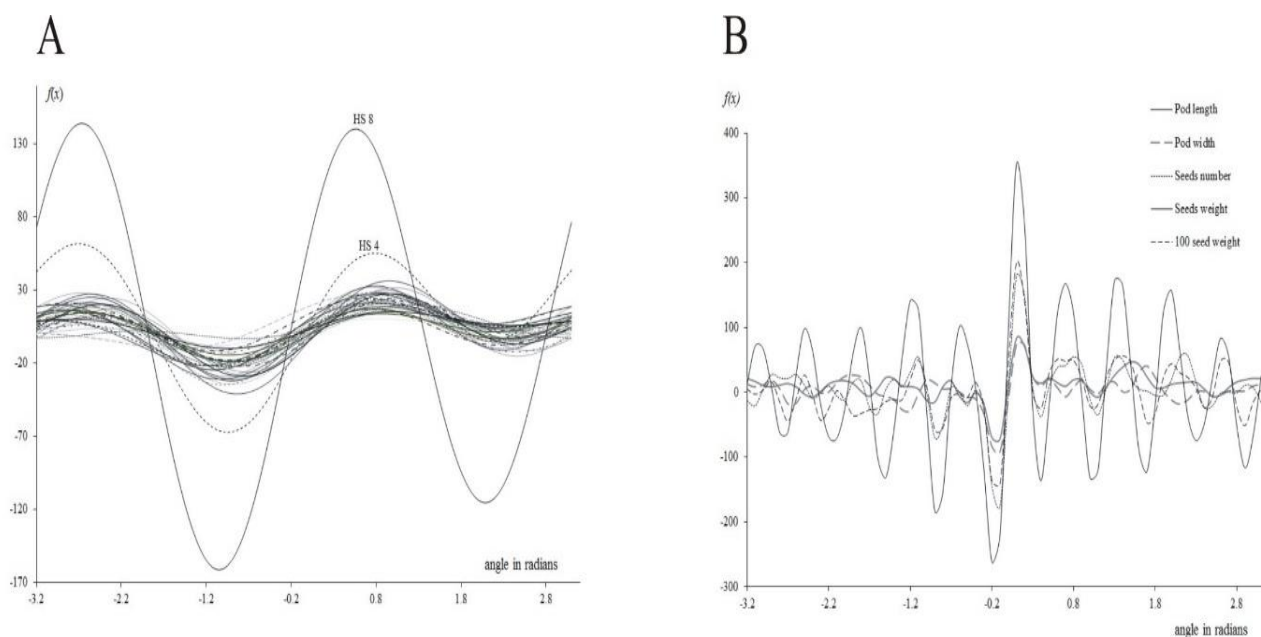| Species	Trait				
	Pod length	Pod width	Seeds number from the pod	Seeds weight from the pod	100 seeds weight
Lathyrus aphaca	0.1194	0.2545	0.7894	0.8797	0.0280
Lathyrus cicera	0.0562	0.4814	0.6174	0.4987	0.9446
Lathyrus clymenum	0.3382	0.2064	0.4432	0.4931	0.0402
Lathyrus hirsutus	0.0361	0.3053	0.0961	0.4736	0.0397
Lathyrus ochrus	0.8411	0.8499	0.8266	0.8126	0.4241
Lathyrus sativus	0.9855	0.0347	0.7109	0.5358	0.6180
Lathyrus tingitanus	0.4150	0.6494	0.3710	0.8425	0.2646

Fig. 2. Andrews curves: A – of genotypes, B – of traits for each species separately.

Conclusions

• CVs of accessions HS8 and HS4 were very high in comparison to the other accessions of *L. hirsutus* (Figs. 1 and 2). In the case of HS8 high variability was shown for 100 seeds weight and seed weight from the pod. For HS4 high variability of seed weight and seed number in the pod was recorded, while limited variability was found for pod width. Andrews curves for HS8 and HS4 were markedly different from curves of the other accessions (Fig. 2).

• Variability of accessions is very small in terms of pod traits.

• In the case of pod length the accessions of *L. ochrus*, *L. sativus* and *L. tingitanus* were characterized by low variability (CVs were max. 6%), while the accession of *L. hirsutus* was characterized by the highest variability. No uniformity was shown for *L. hirsutus* by the Miller test.

• For pod width the accessions of *L. cicera*, *L. sativus* and *L. tingitanus* were characterized by low variability (CVs were max. 6.2%), while the accession of *L. clymenum* was characterized by the highest variability (as confirmed by the shape of Andrews curves, Fig. 4). However, a lack of uniformity was detected only for *L. sativus* (Table 2). This was caused by a very small CV of accession S111 (0.79%). The uniformity of *L. clymenum* could be due to the fact that the Miller test for $n=3$ is very weak (see Methods).

• For the seed traits the accession of *L. hirsutus* was characterized by the highest variability. This conclusion was confirmed by Andrews curves (Fig. 4).

• Two pod traits and three seed traits were clearly different (Fig. 3) in most of the species. For example, for *L. cicera*, *L. hirsutus*, *L. ochrus*, *L. sativus* and *L. tingitanus* the pod traits were similar in shape, while for *L. cicera*, *L. ochrus*, *L. sativus* and *L. tingitanus* those for the seed traits

were similar in shape (see the first conclusion). Moreover, it was shown that seeds weight from the pod was distinguished in the case of all species (Fig. 2) it was especially visible for *L. aphaca* and *L. hirsutus*.

• Despite the observed large differences between CVs of seed traits, no significant differences were found for the number of seeds in the pod and seed weight from the pod using the Miller test; only in the case of 100 seeds weight the CVs differed significantly for *L. aphaca*, *L. clymenum* and *L. hirsutus*.

• The application of the multidimensional NKDA method demonstrated differences in variability between the accessions of *L. hirsutus* (characterized by the greatest differences between CVs) and *L. tingitanus* as well as some accessions of *L. ochrus* (relatively small differences between CVs).

• The NKDA method showed that variability in accessions of *L. sativus*, *L. cicera* and *L. aphaca* was comparable.

• Uniformity of a given accession in respect of all considered traits may be comparable to that of known uniform accessions. Insufficiently uniform accessions could be detected using the Andrews curves (Fig. 2A).

• The significant differences among genotypes of *L. aphaca*, *L. clymenum* and *L. hirsutus* indicate potential to obtain new cultivars differing in terms of 100 seeds weight from established genotypes.

Generally: Uniformity of genotypes may be pre-tested using Andrews curves. Genotypes, which curves have a clearly greater amplitude, are characterized by high CV values. The lines of a similar amplitude and waveform indicate the genotypes with similar variability for all the analyzed traits.

A similar pre-analysis may be performed in terms of traits, variability of genotypes is the largest in terms of the trait of the largest amplitude.
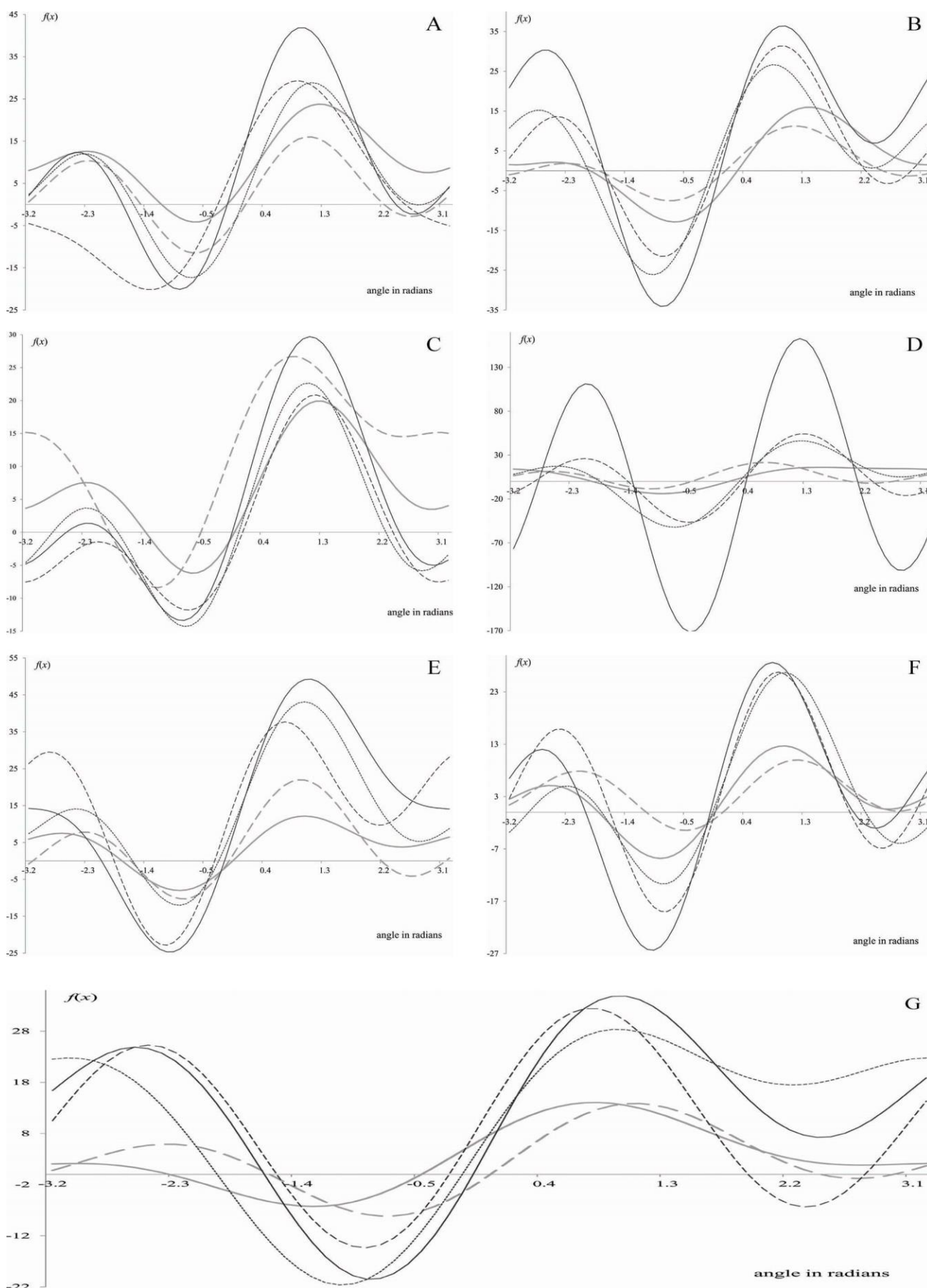
Fig. 3. Andrews curves of traits for each species separately (black solid line – seed weight; black dashed line – 100 pods weight; black dotted line – seed number in the pod; grey solid line – pod length; grey dashed line – pod width): A – *L. aphaca*, B – *L. cicera*, C – *L. clymenum*, D – *L. hirsutus*, E – *L. ochrus*, F – *L. sativus*, G – *L. tingitanus*.
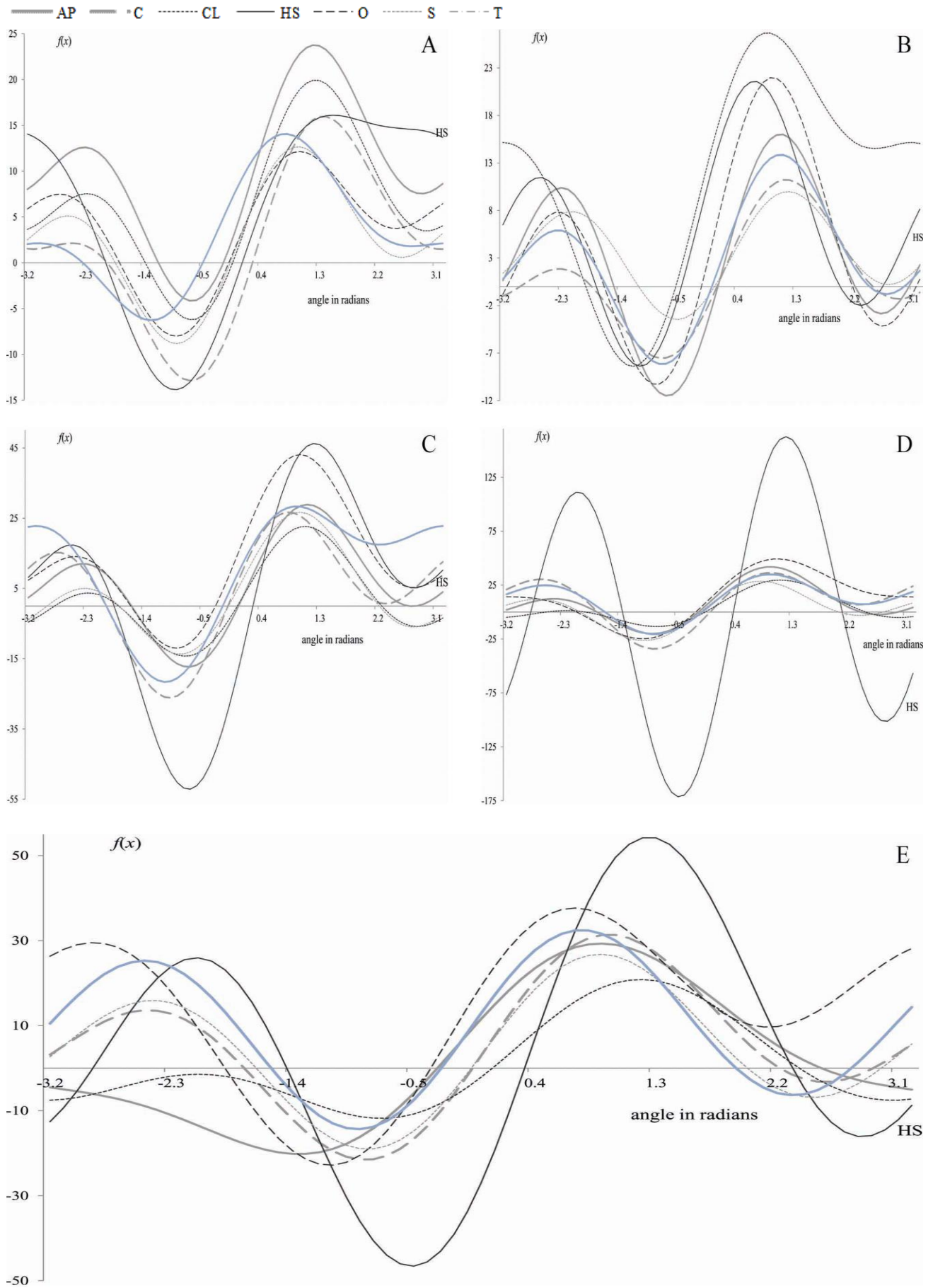
Fig. 4. Andrews curves for species (A – pod length, B – pod width, C – seed number from the pod, D – seed weight, E – 100 seeds weight).
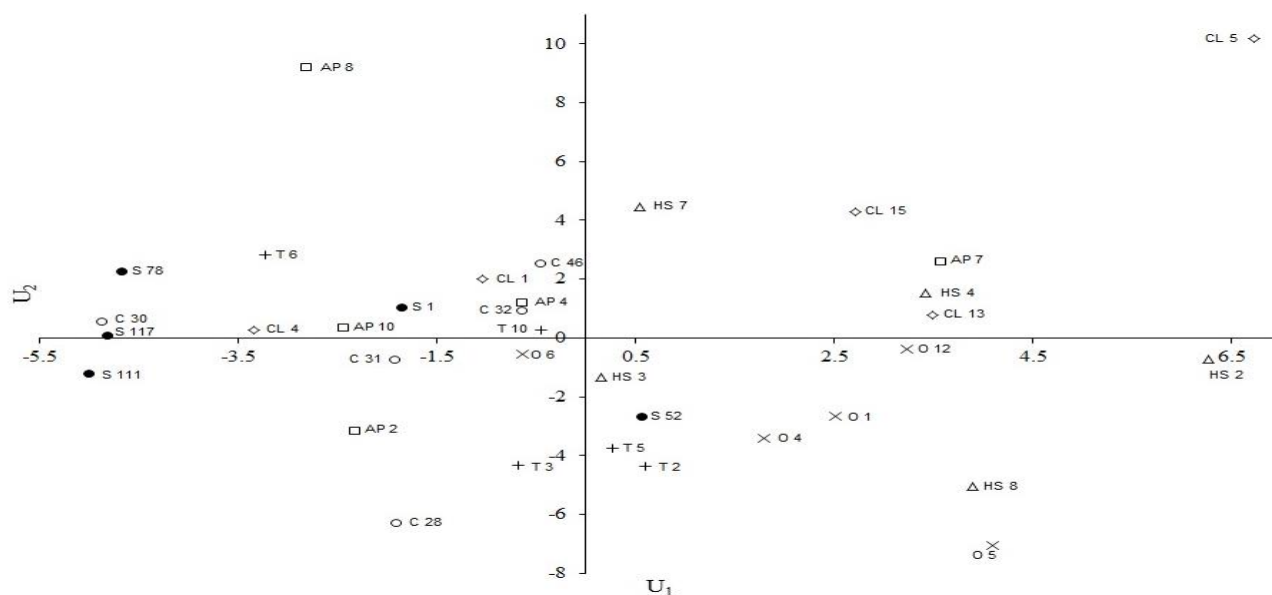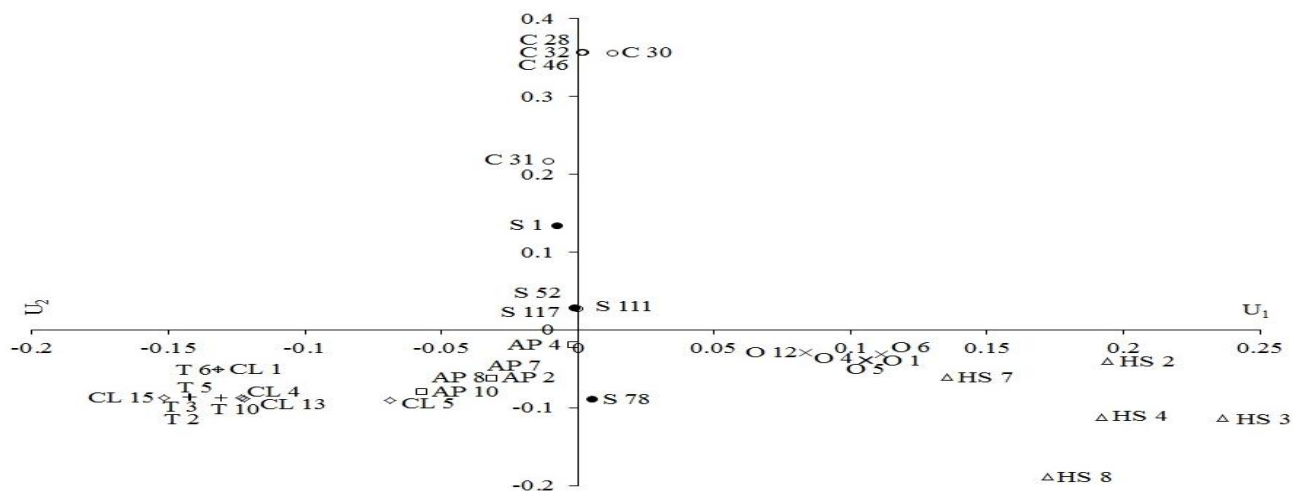
Fig. 5. The linear discriminant analysis (LDA).

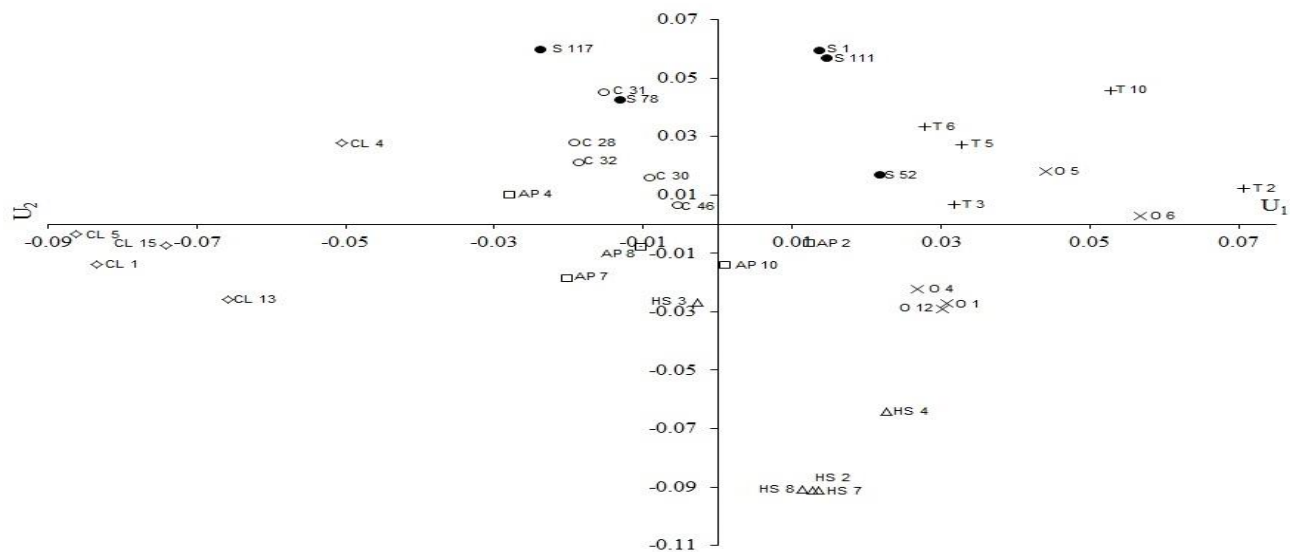Fig. 6. Nonlinear kernel discriminant analysis (NKDA) with the sigmoid kernel.

Fig. 7. Nonlinear kernel discriminant analysis (NKDA) with the Chi-squared kernel.

References

Asmussen, C.B. and A. Liston. 1998. Chloroplast DNA characters, phylogeny, and classification of *Lathyrus* (Fabaceae). *Amer. J. Bot.,* 85(3): 387-392.

Bennett, B.M. 1976. On an approximate test for homogeneity of coefficients of variation. In: Ziegler, W.J. (Ed.), *Contributions to Applied Statistics Dedicated to A. Linder. Experientia Suppl.*, 22: 169-171.

Campbell, C.G. 1997. *Grass pea (Lathyrus sativus L.). Promoting the conservation and use of underutilized and neglected crops*. 18. Institute of Plant Genetics and Crop Plant Research. Gatersleben/International Plant Genetic Resources Institute, Rome, Italy.

Forkman, J. and S. Verrill. 2007. The distribution of McKay's approximation for the coefficient of variation. *Stat Probabil Lett.*, 78: 10-14.

Friedman, J.H. 1989. Regularized discriminant analysis. *J. Amer. Stat. Assoc.*, 84: 165-175.

Khattree, R. and D.N. Naik. 2002. Association in contingency tables, correspondence analysis and (modified) Andrews plots. In Huber-Carol, C., N. Balakrishnan, M.S. Nikulin and M. Mesbah (Eds.). *Goodness-of-fit Tests and Model Validity. Birkhäuser*, Boston: pp. 311-326.

Mika, S., G. Rätsch, J. Weston, B. Schölkopf and K.R. Müller. 1999. Fisher discriminant analysis with kernels. In: Hu, Y.H., J. Larsen, E. Wilson and S. Douglas (Eds.), *Neural Networks for Signal Processing*, IX: 41-48.

Miller, G.E. and C.J. Feltz. 1997. Asymptotic inference for coefficients of variation. *Commun. Stat. Theory*, 26: 715-726.

Schölkopf, B. and A.J. Smola. 2002. *Learning with kernels*. MIT Press.

Vaz Patto, M.C., B. Skiba, E.C.K. Pang, S.J. Ochatt, F. Lambein and D. Rubiales. 2006. *Lathyrus* improvement for resistance against biotic and abiotic stresses: From classical breeding to marker assisted selection. *Euphytica*, 147: 133-147.

Zawieja, B. and A. Szczepańska-Álvarez. 2017. A simulation study of the Bennett test and Miller test for coefficients of variation. *Commun. Stat. Simulat.*, 46(5):3701-3711.

Zawieja, B. and W. Pilarczyk. 2005. The comparison of traditional UPOV uniformity criterion and new approach based on Bennett's test for coefficients of variation. *Colloq. Biometr.*, 35: 155-163.

Zawieja, B., J. Bocianowski and W. Rybiński. 2012. Testing uniformity of mutants of the *Lathyrus sativus* L. (grasspea) using Bennett methods. *Russ. J. Genet.*, 48(2): 224-229.

Zawieja, B., W. Rybiński, K. Nowosad, D. Piesik and J. Bocianowski. 2016. Testing of uniformity of seven *Lathyrus* species using Bennett's and Miller's methods. *Euphytica*, 208: 123-128.