

COMPARISON OF SIMILARITY COEFFICIENTS USED FOR CLUSTER ANALYSIS BASED ON SSR MARKERS IN SISTER LINE WHEAT CULTIVARS

Srbislav DENI^{1,*}, Ron DEPAUW², Vojislava MOMILOVI¹
and Vladimir A. IN¹

¹Institute of Field and Vegetable Crops, M. Gorki St. 30, 21000 Novi Sad, Serbia,

²Semi-arid Prairie Agricultural Research Centre, AAFC, P.O. Box 1030, Swift Current, SK,
Canada, S9H 3X2.

Deni S, R. Depauw, V. Momilovic and V. A. In (2016): *Comparison of similarity coefficients used for cluster analysis based on SSR markers in sister line wheat cultivars.*- Genetika vol48, no1, 219-232.

The objective of this study was to compare fourteen different similarity coefficients and their influence in sister line wheat cultivars clustering. Seventeen sister cultivars developed from two crosses were used and fingerprinted with 19 wheat microsatellite markers. Comparisons among the similarity coefficients were made using the Sperman correlation analysis, dendrogram evaluation (visual inspection and consensus fork index - CI_c), projection efficiency in a two-dimensional space, and groups formed by the Tocher optimization procedure. The Sperman correlation coefficients among the fourteen similarity coefficients were all high showing a strong association between them. The correlation coefficient between Dice and Kulczinski and Ochiai I as well as between Hamann and Simple matching and between Kulczinski and Ochiai I was equal to 1. Although visual estimation of the dendograms shows almost identical clustering structures, CI_c indexes indicate that all coefficients are not identical.

Key words: Similarity coefficient, Cluster analysis, SSR markers, Sperman correlation, Consensus fork index, Distortion degree, Stress value

INTRODUCTION

A similarity coefficient measures the resemblance between two entities based on either or both of two distinct kinds of information pertaining to i variables. The similarity coefficients can be considered both as population parameters as well as sample statistics. So far, many different similarity coefficients have been developed (WARRENS, 2008; TODESCHINI *et al.*,

Corresponding author: Srbislav Deni, Institute of Field and Vegetable Crops, Maksima Gorkog 30, 21000 Novi Sad, Serbia; Phone: 381214898233; Fax: 381214898227 E-mail: srbislav.dencic@nsseme.com

2012). Given that, many different similarity coefficients are available the question arises as to which one (ones) are most suitable for a particular application.

In the area of population genetics and plant breeding quantifying the degree of similarity/dissimilarity among species, subspecies, populations, elite breeding progenies and varieties is one of primary concern. In the period up to 1970, measures of genetic similarity/dissimilarity between autonomous genetic entities based on pedigree analysis, morphological, physiological, or cytological traits (markers) as well as biometrical analysis of quantitative and qualitative traits (MELCHINGER, 1999). After 1980s various molecular markers were used for detection of relationship among different genotypes in seed banks and breeding programs (KOBILJSKI *et al.*, 2007), assessment of genetic diversity present in germplasm pools (LABATE *et al.*, 2003; KONDI -ŠPIKA *et al.*, 2010) as well as identification of essentially derived varieties (LOMBARD *et al.*, 2000; WARBURTON *et al.*, 2002; NOLI *et al.*, 2012; DEN I *et al.*, 2015).

In these various applications of molecular markers a proper choice of a similarity coefficients for studying similarities, dissimilarities or divergence depends on the: marker system properties, germplasm genealogy, taxonomic operational unit, study objectives and conditions required for a multivariate analysis (REIF *et al.*, 2005; BALESTRE *et al.*, 2008).

A number of similarity coefficients for binary variables have been proposed (JACKSON *et al.*, 1989; KOSMAN and LEONARD, 2005; WARRENS, 2008; TODESCHINI *et al.*, 2012). The choice of an appropriate similarity coefficient is very important and decisive point to evaluate clustering, true genetic similarity/dissimilarity between individuals because different similarity coefficients may yield conflicting results (KOSMAN and LEONARD, 2005; TODESCHINI *et al.*, 2012).

In this paper, fourteen similarity coefficients for binary variable were retrieved from literature in order to investigate the influence of the choice of them over the cluster analysis based on data taken from SSR markers of the 17 sister lines wheat cultivars.

MATERIALS AND METHODS

The comparison of similarity coefficients was based on 17 winter wheat cultivars developed in Institute of Field and Vegetable Crops in Novi Sad, Serbia. These cultivars were developed from two crosses, forming two groups of sister line cultivars (Table 1). The cultivars have been registered in former Yugoslavia in the period from 1984 to 1993 (DEN I and KOBILJSKI, 2008).

Total genomic DNA was extracted from leaves, as described in DOYLE and DOYLE (1990). For molecular assessment of wheat cultivars nineteen wheat microsatellites and one secalin-specific marker were selected in accordance with suggestions and results of RÖDER *et al.*, (1998, 2002) (Table 2). PCR products were amplified using fluorescently labeled forward and unlabeled reverse primers. Microsatellite fragments were detected by capillary electrophoresis on ABI Genetic Analyzer 3130 and analyzed using the GeneMapper Software version 4.0. Data analysis was performed using Microsoft Excel-Software, and the polymorphism information content (PIC) was calculated according to ANDERSON *et al.*, (1993).

Genetic similarity estimates (gs_{ij}) were obtained between each pair of cultivars (i, j) using fourteen similarity coefficients (Table 3). Similarities obtained from these coefficients were transformed into genetic distance measures by the following equation: $dg_{ij} = 1 - gs_{ij}$ as described by JOHNSON and WICHERN (2007). The similarity coefficients were calculated with the NTSYS-PC program (ROHLF, 1998).

Table 1. Crosses and parentage from which sister cultivar derived, their generation of separation and year of registration

Cross No.	Parentage	CVN*	Name of cultivar	Pedigree	Generation of separation	Year of Registr
I	NS 646 / Bezostaja 1 //Aurora/3/ Partizanka	1	Rudni anka	I-10-AB-5-Z	F ₆	1989
		2	Suboti anka	I-10-AB-5	F ₅	1987
		3	Lozni anka	I-12-BB-8-A	F ₇	1988
		4	Kosovka	I-12-BB-8-B	F ₇	1987
		5	Tanjugovka	I-12-BB-8-C	F ₇	1988
		6	Jednota	I-12-BB-8	F ₆	1987
		7	Rodna	I-14-CB-10-AB-4-A	F ₈	1988
		8	Partizanka niska	I-14-CB-10-A	F ₆	1984
		9	Jadranka	I-14-CB-10-B	F ₆	1988
		10	Banatka niska	I-17-DD-7	F ₅	1985
II	Talent/ Novosadsk arana 2	11	Italija	II-8-AB-2-CC	F ₇	1989
		12	Evropa	II-8-AB-2-CC-A	F ₈	1988
		13	Evropa 90	II-8-AB-2-CC-B	F ₈	1990
		14	Sremka	II-9-BB-7-Z	F ₇	1989
		15	Atina	II-9-BB-7A	F ₆	1993
		16	Novosadska 6	II-9-BB-7B	F ₆	1991
		17	Francuska	II-9-BB-7C	F ₆	1988

* CVN: cultivar number, Interpretation: CVN 1 is a selection out of CV 2; CVN 3, 4, 5 are sibs selected out of CVN 6; CVN 8, 9 are sibs selected from CVN 7; CVN 10 selected out of other cultivars; CVN 12, 13 are sibs selected out of CVN 11; CVN 14 is selected out of sibs CVN 15,16,17.

Table 2. Similarity coefficients used among 17 sister wheat cultivar based on SSR markers

Coefficient (C)	Expression	Source
Anderberg (A)	$a + 2(b + c)$	Anderberg, 1973
Dice (D)	$2a + b + c$	Dice, 1945
Gower (G)	$\frac{\sum_k W_{ijk}}{a + b + c + d}$	Gower, 1971
Hamann (H)	$a + b + c + d$	Hamann, 1961
Jaccard (J)	$a + b + c$	Jaccard, 1901
Kulczynski (KUL)	$b + c$	Kulczynski, 1927
Ochiai I (OI)	$\sqrt{(a + b)(a + c)}$	Ochiai, 1957
Ochiai II (OII)	$\sqrt{(a + b)(a + c)(d + b)(d + c)}$	Ochiai, 1957
Phi coefficient (PHI)	$\sqrt{(a + b)(c + d)(a + c)(b + d)}$	Sokal & Sneath, 1963
Russel and Rao (RR)	$a + b + c + d$	Russel and Rao, 1940
Rogers and Tanimoto (RT)	$a + d + 2(b + c)$	Rogers and Tanimoto, 1960
Simple matching (SM)	$a + b + c + d$	Sokal and Michener, 1958
Sokal & Sneath 1 (SS1)	$2(a + d) + b + c$	Sokal & Sneath, 1963
Yule (Y)	$ad + bc$	Yule, 1911

The fourteen similarity coefficients were compared using the Spermán correlation coefficient. Dendograms were constructed according to the unweighted pair-group mean arithmetic method (UPGMA) using NTSYS software. The different dendograms compared using visual inspection and the consensus fork index CI_c (ROHLF, 1982). The CI_c index provides a relative estimate of the dendogram similarities and calculated using NTSYS software.

Table 3. Description of microsatellite markers employed including the number of alleles and PIC-values

Marker	Chromosome	Variation in product size (bp)	Predominating allele (bp) and its frequency	Total no. of alleles	PIC
GWM3	3D	71-85	73 (47.1)	4	0.60
GWM18	1B	null, 186-196	186 (70.6)	3	0.46
GWM46	7B	null, 167-175	173 (41.7)	5	0.72
GWM95	2A	116-118	118 (64.7)	2	0.46
GWM155	3A	null, 138-146	138 (38.9)	6	0.76
GWM190	5D	202-214	212 (66.7)	4	0.50
GWM261	2D	175-192	192 (58.8)	2	0.48
GWM325	6D	null, 133-141	137 (41.2)	6	0.74
GWM357	1A	117-121	121 (52.9)	2	0.58
GWM389	3B	null, 98-136	136 (38.9)	7	0.76
GWM408	5B	145-177	145 (52.9)	3	0.60
GWM 437	7D	81-113	86 (35.3)	6	0.75
GWM 458	1D	null, 107-111	107 (52.9)	3	0.58
GWM513	4B	141-143	143 (64.7)	4	0.46
GWM577	7B	null, 125-161	null (52.9)	4	0.62
GWM619	2B	133-158	139 (29.4)	5	0.78
GWM631	7A	null, 189-206	198 (29.4)	7	0.80
GWM680	6B	null, 105-119	119 (81.2)	4	0.33
Taglgap	1B	null, 207-239	217 (58.8)	4	0.56
Total	19			81	
Average				4.3	0.51

Similarity expression in a two-dimensional space was done using the software GENES, in order to determine the effect of choosing the similarity coefficients (MEYER *et al.*, 2004). This technique estimates the coordinates for each entity, from a dissimilarity matrix, by statistic procedures that minimize the differences between the original distances and distances obtained in the two-dimensional space. This procedure allowed the comparison of the coefficients using the three parameters:

- Correlations between the original distances and the distances obtained by two-dimensional dispersion (r)
- Degree of distortion ($1 - \alpha$), given by:

$$\alpha = \frac{\sum_i \sum_j d_{ij}^2}{\sum_i \sum_j d_{ij}}$$

c) Stress value (s) given by:

$$s = \sqrt{\frac{\sum_{i < j} (d_{oij} - d_{gij})^2}{\sum_{i < j} d_{oij}^2}}$$

where: d_{gij} and d_{oij} are the graphic (two-dimensional space) and original distances (non two-dimensional space), respectively, for all pairs of cultivars i and j .

Representation of stress (standardized residual sum of squares) was proposed by KRUSKAL (1964) as parameter which determines the goodness-of-fit of the graphic projection. According to KRUSKAL (1964) the stress value can be classified as 40, 20, 10, 5 or 0% stress that is adjustment will be non-satisfactory, regular, good, excellent and perfect, respectively.

The establishment of the clusters was also studied by the Tocher optimization procedure (RAO, 1952), using the Gene Program (CRUZ, 2001). In this method cultivars are separated into non-empty and equally exclusive sub-groups, based on the similarity or dissimilarity matrix.

RESULTS

The sister line wheat cultivars were fingerprinted with 19 wheat microsatellite markers and the secalin-specific (*Secal*) marker. A total of 81 alleles were detected at 19 wheat microsatellite loci, resulting in an average allele number per marker of 4.3 (Table 3). The presence of the 1B-1R wheat-rye translocation was verified with the secalin-specific primers. The results have shown that PCR product of 96 bp (*Secal*) was detected in 16 out of 17 cultivars, indicating the presence of the 1B-1R translocation in most of the tested varieties.

The number of alleles for the individual markers ranged from three (Xgwm95, Xgwm261 and Xgwm357) to 7 (Xgwm389 and Xgwm631). For 10 markers, null alleles were detected. The values of polymorphism information content (PIC) ranged from 0.33 for Xgwm680 to 0.80 for Xgwm631. An average PIC value of 0.51 for all markers indicated a high level of detected polymorphism (Table 3).

Table 4. The Spearman correlation coefficient between the similarity coefficients for SSR markers

C*	A	D	G	H	J	KUL	OI	OII	PHI	RR	RT	SM	SS1
D	0.92												
G	0.91	0.98											
H	0.92	0.99	0.92										
J	0.97	0.98	0.91	0.98									
KUL	0.92	1.00	0.92	0.99	0.98								
OI	0.92	1.00	0.92	0.99	0.98	1.00							
OII	0.95	0.92	0.99	0.92	0.93	0.92	0.92						
PHI	0.92	0.99	0.92	0.99	0.98	0.99	0.99	0.92					
RR	0.92	0.97	0.90	0.98	0.98	0.99	0.99	0.91	0.99				
RT	0.96	0.99	0.92	0.99	0.99	0.99	0.99	0.92	0.99	0.98			
SM	0.92	0.99	0.92	1.00	0.98	0.99	0.99	0.92	0.99	0.98	0.99		
SS1	0.91	0.99	0.92	0.99	0.97	0.99	0.99	0.91	0.99	0.98	0.98	0.99	
Y	0.78	0.89	0.81	0.88	0.86	0.89	0.89	0.81	0.88	0.88	0.87	0.88	0.88

*abbreviations of the coefficients according to Table 2.

The Spermman correlation coefficients among the fourteen similarity coefficients were all high showing a strong association between them. The correlation coefficient between Dice and Kulczynski and Ochiai I as well as between Hamann and Simple matching and between Kulczynski and Ochiai I was equal to 1 indicating that they classify the similarity among sister cultivars exactly in the same order. The Yule coefficient had slightly lower correlation value with the all other coefficients, especially in relation Anderberg, Gower and Ochiai II. (Table 4).

A visual estimation of the dendrograms shows that all coefficients present similar clustering structures (Figure 1A; Figure 1B). It was observed that they were able to separate sister cultivars in to two groups (clusters) which not fully correspond to their pedigree. Although the general structure of the dendrograms is highly similar, it is observed that there are slight alterations in the levels in which cultivars are grouped. MEYER *et al.*, (2004), studying maize lines originating from two populations, found that the use of different similarity coefficients did not change line grouping within each group.

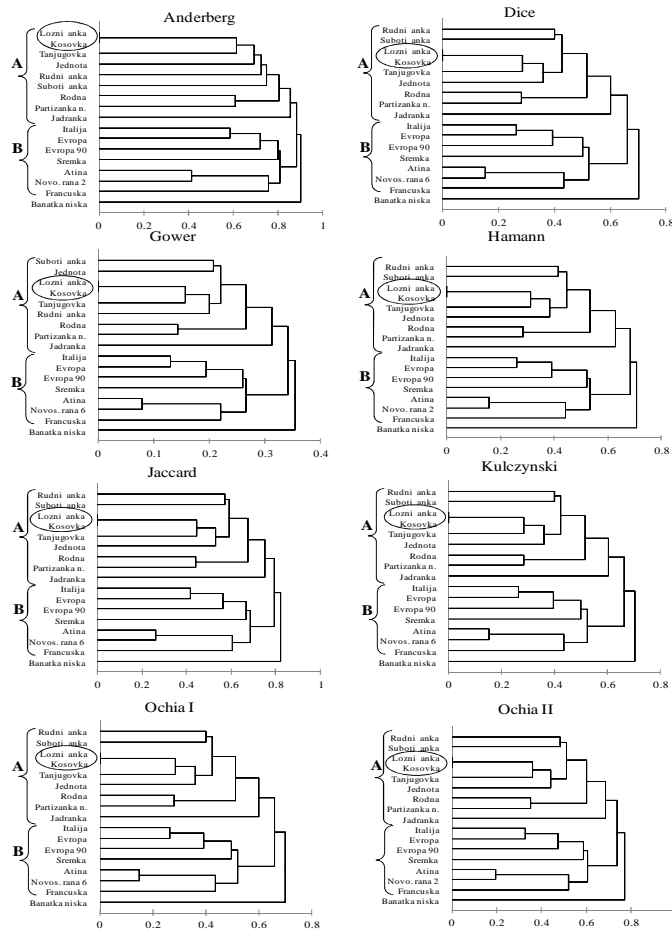


Figure 1A. Dendrograms constructed from matrices of genetic distances obtained by the complement of the similarity coefficients

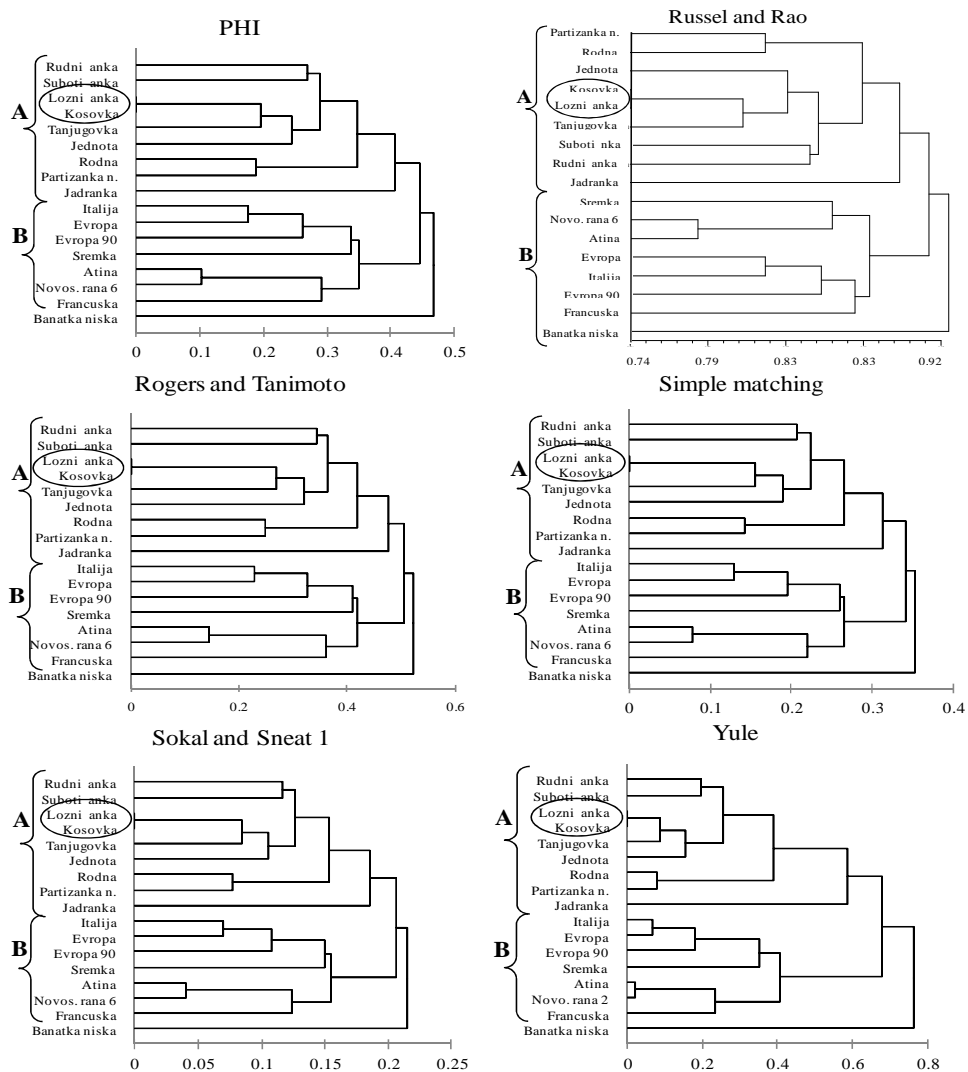


Figure 1 B. Dendrograms constructed from matrices of genetic distances obtained by the complement of the similarity coefficients

The comparison of the generated dendograms, using the consensus fork index CI_c , allows a refining of what was observed through visual estimation. This index ranges from 0 to 1, where the dendograms are considered identical when the index value is 1. Although visual estimation of the dendograms shows almost identical clustering structures, CI_c indexes indicate that all coefficients are not identical. Anderberg, Gower, Ochiai II and Russel and Rao are not identical to any other coefficients (Table 5).

The projection efficiency of the genetic distances of the similarity coefficients in a bidimensional space considering three evaluation parameters (distortion, correlation between original and estimated distances and stress) presented certain variation (Table 6). The distorted value are coherent with correlation value and especially with stress value. The Yule coefficient showed the smallest distorted value (27.96%) as well as stress value (39.49%), while Russel and Rao's the biggest (Table 6). The same tendency, in which Russel and Rao's similarity coefficient had the highest stress value among other coefficients reported by MAYER *et al.*, (2004) and BALESTRE *et al.*, (2008).

Table 5. Consensus fork index (CI_c) among the dendograms (UPGMA) produced by the complements of the similarity coefficients, among sister wheat cultivars

C*	A	D	G	H	J	KUL	OI	OII	PHI	RR	RT	SM	SS1
D	0.93												
G	0.80	0.80											
H	0.93	1.00	0.80										
J	0.93	1.00	0.80	1.00									
KUL	0.93	1.00	0.80	1.00	1.00								
OI	0.93	1.00	0.80	1.00	1.00	1.00							
OII	0.73	0.80	0.73	0.80	0.80	0.80	0.80						
PHI	0.93	1.00	0.80	1.00	1.00	1.00	1.00	0.80					
RR	0.80	0.87	0.73	0.87	0.87	0.87	0.87	0.80	0.87				
RT	0.93	1.00	0.80	1.00	1.00	1.00	1.00	0.80	1.00	0.87			
SM	0.93	1.00	0.80	1.00	1.00	1.00	1.00	0.80	1.00	0.87	1.00		
SS1	0.93	1.00	0.80	1.00	1.00	1.00	1.00	0.80	1.00	0.87	1.00	1.00	
Y	0.93	1.00	0.80	0.80	1.00	1.00	1.00	0.80	1.00	0.87	1.00	1.00	1.00

*abbreviations of the coefficients according to Table 2.

The Tocher optimization procedure as an individual clustering method does not form the same groups as the dendograms. Nevertheless, there is no information about the similarity of the cultivars inside each group or similarity among the groups. This is the major disadvantage of the method. In the present study, in all coefficients, except Russel and Rao, the number and structure of the formed groups were identical. Russel and Rao coefficient form 2 clusters while all other coefficients form 9 clusters (Table 7).

Table 6. Distortion degree, correlation between original and estimated distances and stress value obtained by the projection of the genetic distances in a two-dimensional space

C*	Distortion (%)	Correlation (r)	Stress value (%)
A	61.82	0.3634	66.57
D	40.93	0.5739	47.92
G	41.49	0.5511	48.52
H	41.46	0.5510	48.50
J	52.74	0.4886	57.79
KUL	40.88	0.5754	47.86
OI	40.87	0.5753	47.85
OII	46.41	0.5443	52.08
PHI	40.87	0.5699	47.87
RR	71.85	0.6624	76.40
RT	48.30	0.5159	53.80
SM	41.49	0.5520	48.51
SS1	37.32	0.5636	46.01
Y	27.96	0.6612	39.49

*abbreviations of the coefficients according to Table 2.

Table 7. Clustering of 17 wheat sister cultivars by means of Tocher's optimization method considering different similarity coefficients

Cluster	Coefficients**													
	A	D	G	H	J	KUL	OI	OII	PHI	RR	RT	SM	SS1	Y
I	12	12	12	12	12	12	12	12	12	13 5 7 9 11 13 15 17	12	12	12	12
II	3 5	3 5	3 5	3 5	3 5	3 5	3 5	3 5	3 5	2 4 6 8 10 12 14 16	3 5	3 5	3 5	3 5
III	4 6	4 6	4 6	4 6	4 6	4 6	4 6	4 6	4 6		4 6	4 6	4 6	4 6
IV	7 9	7 9	7 9	7 9	7 9	7 9	7 9	7 9	7 9		7 9	7 9	7 9	7 9
V	8 10	8 10	8 10	8 10	8 10	8 10	8 10	8 10	8 10		8 10	8 10	8 10	8 10
VI	11	11	11	11	11	11	11	11	11		11	11	11	11
	12	12	12	12	12	12	12	12	12		12	12	12	12
VII	13	13	13	13	13	13	13	13	13		13	13	13	13
	14	14	14	14	14	14	14	14	14		14	14	14	14
VIII	15	15	15	15	15	15	15	15	15		15	15	15	15
	17	17	17	17	17	17	17	17	17		17	17	17	17
IX	16	16	16	16	16	16	16	16	16		16	16	16	16

*enumeration of the cultivars according to Table 1; ** abbreviations of the coefficients according to Table 2.

DISCUSSION

The visual analysis of Figure 1A and 1B indicated that all the dendograms expressed the following characteristics: i) not capable to estimate the distance/distinctness between two sister cultivars Lozni anka and Kosovka; ii) separated all cultivars into two main clusters: A (sister cultivars from cross NS 646/Bezostaja 1//Aurora/3/Partizanka) and B (sister cultivars from cross Talent/ Novosadska rana2) except cultivar Banatka niska which derived from cross NS 646/Bezostaja 1//Aurora/3/Partizanka exhibited the great distances to the both clusters.

Only one pair of sister cultivars (Lozni anka and Kosovka) derived from the same population I could not be distinguished. That cannot be addressed to nature of used set of similarity coefficients but rather to limited number of genomic regions in which the cultivars differ and such differences cannot be identified with a small number of markers. The same observation was described by PLASCHKE *et al.*, (1995) who tested 40 wheat genotypes using 23 microsatellite markers and found two sister cultivars could not be distinguished.

All coefficients demonstrated that cultivar Banatska niska was clearly separated from both clusters. There is no doubt that this cultivar is genetically distant from its sister varieties. The reason for this is probably its segregation from basic population in the early stage of selection (F_5).

The dendograms were not affected by the nature of the coefficient i.e., the coefficients Anderberg, Dice, Jaccard, Kulczynski and Ochiai I which do not consider the negative matches had the same intergroup distribution pattern such as other coefficients which included the negative matches. BALESTRE *et al.*, (2008), studying maize inbred lines, also found that use of different similarity coefficients did not change distribution structures of line grouping in the dendograms. Similar results were found in potatoes (ROCHA *et al.*, 2001).

The discussion of inclusion or exclusion of negative matches was under long debate. SOKAL and SNEATH (1963) argued that the negative matches do not mean necessarily any similarity between two entities. This is because an almost infinite number of attributes is possible lacking in two entities. In cases where the two binary states are not equally important, such as in the asymmetric type of binary data, the positive matches are usually more significant than negative matches (BARONI-URBANI and BUSER, 1976; FAITH, 1983). For that kind of asymmetric data FAITH *et al.*, (1987) suggested to include the negative matches but only gave the half weight while giving the full weight for the positive matches.

In general, all correlations among coefficients were high. It was observed that some pairs of coefficients such as Dice and Kulczynski; Dice and Ochiai I; Kulczynski and Ochiai I as well as Hamann and Simple matching were perfectly correlated (1). These perfectly correlated coefficients belong to same type i.e., the coefficients Dice, Kulczynski and Ochiai I are asymmetric coefficients which ignore double-zero attributes in the similarity calculation, while coefficients Hamann and Simple matching are symmetric type which use double-zero state (d). The high correlations between eight similarity coefficient (some of them were equal to 1) reported by MEYER *et al.*, (2004) and BALESTRE *et al.*, (2008) both in maize. In very detailed paper which reports an analysis and comparison of the use of 51 different similarity coefficients TODESCHINI *et al.*, (2012) emphasized that some pairs of coefficients were always perfectly correlated. These are: Simple matching, Rogers and Tanimoto, Jaccard, Ochiai I, Dice, Yule and Adenberg.

Considering the values of CI_c indexes most of the similarity coefficient were identical. The exceptional cases are the Anderberg, Gower, Ochiai II and Russel and Rao coefficients which have CI_c value less than 1 in relation to all other coefficients. The dendograms obtained by

Anderberg's and Russel and Rao's coefficients showed a little alterations in intra-grouping structure compare to all other coefficients (Fig.1). Different dendograms obtained by Russel and Rao's coefficients as well as lower value of Cl_c index in compare to other similarity coefficients have been reported by MEYER *et al.*, (2004), BALESTRE *et al.*, (2008) and SESLI and YEGENOGLU (2010).

The two-dimensional projection efficiency, analyzed according to the Kruskal's classification showed, that the stress value were unsatisfactory for all coefficients except the Yule coefficient whose stress value (39,49%) was on the border of unsatisfactory level. The Yule coefficient also had the significantly lowest distortion degree (27,96%) in compare to all other coefficients (Tab. 6). According to Kruskal's classification only Yule coefficient is appropriate to use for the data that we have in the analysis. This suggests that correlations based coefficients (Yule and Phi coefficients) are better suited to the analysis of these data in comparing to the symmetric (Grower, Hamann, Ochiai II, Roger and Tanimoto, Simple matching, and Sokal and Sneath) and asymmetric (Anderberg, Dice, Jaccard, Kulczynski, Ochiai I, and Russel and Rao) coefficients. According to the projection efficiency, i.e., distortion degree and stress value, Russel and Rao's coefficient was the worst. Similar results refer to Russel and Rao's coefficient were reported by other authors in maize inbred lines (MEYER *et al.*, 2004; BALESTRE *et al.*, 2008).

In the Tocher procedure Russel and Rao's coefficient once again was the most discriminatory (Tab. 7). Russel and Rao's coefficient almost always showed different results from the others (MEYER *et al.*, 2004). Reason for that was explained by SOKAL and SNEATH (1963). They reported that this coefficient is a special type so-called "hybrid" coefficient which excluding negative matches from numerator but not from the denominator.

ACKNOWLEDGMENT

This work was supported by the Serbian Ministry of Education, Science and Technological Development (Project No. TR 31066).

Received October 19st, 2015

Accepted January 25th, 2016

REFERENCES

- ANDERBERG, M.R. (1973): Clustering analysis for applications. London, Academic Press, UK.
- ANDERSON, J.A., G.A. CHURCHILL, J.E. AUTRIQUE, S.D. TANKSLEY AND M.E. SORRELLS (1993): Optimizing parental selection for genetic linkage maps. *Genome*, 36: 181-186.
- BALESTRE, M., R.G. VON PINHO, J.C. SOUYA and J.L. LIMA (2008): Comparison of maize similarity and dissimilarity genetic coefficients based on microsatellite markers. *Genet. Mol. Res.*, 7: 695-705.
- CRUZ, C.D. (2001): Programa Genes: versão Windows; aplicativo computacional em genética e estatística. Universidade Federal de Viçosa, Viçosa.
- DEN I , S. and B. KOBILJSKI (2008): Results of half a century of wheat breeding at Institute of Field and Vegetable crops in Novi Sad. Proc. Inter. Conference „Conventional and Molecular Breeding of Field and Vegetable Crops“, 24-27 November, Novi Sad, Serbia: 377-384.
- DEN I , S., R. DEPAUW, V. MOM ILOVI , A. KONDI -ŠPIKA (2015): Efficiency of the different marker systems for estimation of distinctness between sister line wheat cultivars. *Genetika*, 47 (1): 219-232
- DICE, L.R. (1945): Measures of the amount of ecologic association between species. *Ecology*, 26: 297-302.
- DOYLE, J.J. and J.L. DOYLE (1990): Isolation of plant DNA from fresh tissue. *Focus*, 12: 13-15.

- FAITH, D.P. (1983): Asymmetric binary similarity measures. *Oecologia*, 57: 287-290.
- FAITH, D.P., P.R. MINCHIN and L. BELBIN (1987): Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69: 57-68.
- HAMANN, U. (1961): Merkmalbestand und Verwandtschaftsbeziehungen der Farinosae. Ein Beitrag zum System der Monokotyledonen. *Willdenowia*, 2: 639-768.
- JACCARD, P. (1901): Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société vaudoise des sciences naturelles*, 37: 547-579.
- JACKSON, D.J., K.M., SOMERS and H.H. HARVEY (1989): Similarity Coefficients: Measures of Co-Occurrence and Association or Simply Measures of Occurrence. *Am. Nat.*, 133: 436-453.
- JOHNSON, R.A. and D.W. WICHERN (2007): *Applied multivariate statistical analysis*. 6th ed. Prentice Hall, New York.
- KOBILJSKI, B., S. DEN I , N. HRISTOV, N. MLADENOV, S. QUARRIE, P. STEPHENSON and J. KIRBY (2007): The potential of microsatellites for Marker-Assisted Breeding of improved grain yield in wheat. In: Buck, H.T., Nisi, J.E and Salomon, N. (eds.) *Wheat Production in Stress Environments*. Pub. Springer, Dordrecht, The Netherlands: 729-737.
- KONDIC-SPIKA, A., B. KOBILJSKI, S. DENCIC, M. MLADENOV, N. HRISTOV, LJ. BRBAKLIC and D. TRKULJA (2010): Genetic Diversity in a Collection of Serbian Wheat (*Triticum aestivum* L.) Cultivars Released in Twentieth Century as Revealed by Microsatellite Markers. *In vitro cellular & developmental biology-animal*, 46: 134-134
- KOSMAN, E. and K.J. LEONARD (2005): Similarity coefficients for molecular markers in studies of genetic relationship between individuals for haploid, diploid, and polyploid species. *Mol. Ecol.*, 14: 415-425.
- KRUSKAL, J.B. (1964): Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, 29: 1-27.
- KULCZYNSKI, S. (1927): Die Pflanzenassoziationen der Pieninen. *Bulletin International de l'Academie Polonaise des Sciences et des Lettres, Classe des Sciences Mathematiques et Naturelles*, B: 57-203.
- LABATE, J.A., K.R. LAMKEY, S.H. MITCHELL, H. KRESOVICH, H. SULLIVAN and J.S.C. SMITH (2003): Molecular and Historical Aspects of Corn Belt Dent Diversity. *Crop Sci.*, 43: 80-91.
- LOMBARD, V., C.P. BARIL, P. DOBREUIL, F. BLOUET and D. YHANG (2000): Genetic relationships and fingerprinting of rapeseed cultivars by AFLP. Consequences for varietal registration. *Crop Sci.*, 40: 1417-1425.
- MELCHINGER, A.E. (1999): Genetic diversity and heterosis. In Coors J.G., Pandey S. (ed.): *The genetics and exploitation of heterosis in crops*. ASA, CSSA and SSSA, Madison WI, USA: 99-118.
- MEYER, A.S., A.A.F. GARCIA, A.P. DE SOUZA and JR.C.L. DE SOUZA (2004): Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays* L). *Genet. Mol. Biol.*, 27: 83-91.
- NOLI, E., M.S. TERIACA and S. CONTI (2012): Identification of a threshold level to assess essential derivation in durum wheat. *Mol. Breeding*, 29: 687-698.
- OCHIAI, A. (1957): Zoogeographic studies on the soleoid fishes found in Japan and its neighboring regions. *Bull. Jpn. Soc. Sci. Fish.*, 22: 526-530.
- PLASCHKE, J., M.W. GANAL and M.S. RÖDER (1995): Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor. Appl. Genet.*, 91: 1001-1007.
- RAO, R.C. (1952): *Advanced statistical methods in biometric research*. John Wiley and Sons, New York.
- REIF, J.C., A.E. MELCHINGER and M. FRISCH (2005): Genetical and Mathematical Properties of Similarity and Dissimilarity Coefficients Applied in Plant Breeding and Seed Bank Management. *Crop Sci.*, 45: 1-7.
- ROCHA, B.H.G., E. AUGUSTIN, J.B. DA SILVA and J. VIEGAS (2001): Isoenzymatic variability in wild potatoes. *Pesq. Agropec. Bras.*, 36: 781-791.
- RÖDER, M.S., K. WENDEHAKE, V. KORZUN, G. BREDEMEIJER, D. LABORIE, D. BERTRAND, P. ISAAC, S. RENDELL, J. JACKSON, R.J. COOKE, B. VOSMAN and M.W. GANAL (2002): Construction and analysis of a microsatellite-based database of European wheat varieties. *Theor. Appl. Genet.*, 106: 67-73.
- ROGERS, J.S. and T.T. TANIMOTO (1960): A computer program for classing plants. *Science*, 132:1115-1118.

- ROHLF, F.J. (1982): Consensus indices for comparing classifications. *Math. Biosci.*, 59: 131-144.
- ROHLF, F.J. (1998): NTSY Spc: Numerical Taxonomy and Multivariate Analysis System version 2.02 Exeter Software, Setauket, NY, USA.
- RÖDER, M.S., V. KORZUN, K. WENDEHAKE, J. PLASCHKE, M.H. TIXIER, P. LEROY and M.W. GANAL (1998): A microsatellite map of wheat. *Genetics*, 149: 2007-2023.
- RUSSEL, P.F. and T.R. RAO (1940): On habitat and association of species of anophelinae larvae in south-eastern Madras. *J. Malaria Inst. India*, 3: 153-178.
- SESLI, M. and E.D. YEGENOGLU (2010): Comparison of similarity coefficient used for cluster analysis based on RAPD markers in wild olives. *Genet. Mol. Res.*, 9: 2248-2253.
- SOKAL, R.R. and C.D. MICHENER (1958): A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.*, 38: 1409-1438.
- SOKAL, R.R. and P.H.A. SNEATH (1963): Principles of numeric taxonomy. W. H. Freeman and Company, San Francisco, CA, USA.
- TODESCHINI, R., V. CONSONNI, H. HIANG, J. HOLLIDAY, M. BUSCEMA AND P. WILLETT (2012): Similarity Coefficients for Binary Chemoinformatics Data: Overview and Extended Comparison Using Simulated and Data Sets. *J. Chem. Inf. Model.*, 52: 2884-2901.
- WARRENS, M.J. (2008): Similarity Coefficients for Binary Data. [PhD. Thesis.] Institute of Psychology, Unit Methodology and Statistics, Leiden University, Germany.
- WARBURTON, M.L., B. SKOVMAND and A. MUJEEB-KAZI (2002): The molecular genetics characterization of the 'Bobwhite' bread wheat family using AFLPs and the effect of the T1BL.1RS translocation. *Theor. Appl. Genet.*, 104: 868-873.

**POREĐENJE KOEFICIJENATA SLIČNOSTI KORIŠĆENIH ZA KLASTER ANALIZU
NA OSNOVU SSR MARKERA U SESTRINSKIM LINIJAMA PŠENICE**

Srbislav DENI^{1,*}, Ron DEPAUW², Vojislava MOMILOVI¹
i Vladimir A. IČIĆ¹

¹Institut za ratarstvo i povrtarstvo, Novi Sad, Srbija, ²Semiarid Prairie Agricultural Research
Centre, AAFC, P.O. Box 1030, Swift Current, SK, Canada, S9H 3X2

Izvod

Cilj ovoga rada je bio da se uporede rezultati etnaest različitih koeficijenata sličnosti u grupisanju sestrinskih sorti pšenice. Urađena je fingerprint sa 19 mikrosatelitskih markera na sedamnaest sestrinskih sorti stvorenim iz dve kombinacije ukrštanja. Poređena koeficijenata sličnosti su urađena koristeći: Spermans korelacijone analize, dendograma evaluacije (vizuelni indeks inspekcije i Consensus fork index - C_{IC}), projekcione efikasnost u dvodimenzionalnom prostoru, i grupa koje su formirane na bazi Tocher optimizacione procedure. Prema Spermans korelacionim koeficijentima sve korelacije između etnaest koeficijenata sličnosti su veoma visoke te pokazuju snažnu vezu između njih. Koeficijent korelacije između Dice i Kulczinski i Ochiai I, kao i između Hamann i Simple matching te između Kulczinski i Ochiai I bio je jednak 1. Iako vizuelna procena dendograma pokazuje gotovo identične klaster strukture, C_{IC} indeksi ukazuju da svi koeficijenti nisu identični.

Primljeno 19.X.2015.

Odobreno 25.I. 2016.