

Intra-population genetic diversity of cultivated carrot (*Daucus carota* L.) assessed by analysis of microsatellite markers*

Anna Maksylewicz and Rafal Baranski[✉]

Institute of Plant Biology and Biotechnology, Faculty of Horticulture, University of Agriculture in Krakow, Kraków, Poland

Intra-population variation of 18 cultivated carrot (*Daucus carota* L. ssp. *sativus*) populations of diverse origins was evaluated using codominant microsatellite (SSR) markers. Using 27 genomic and EST-derived SSR markers, 253 alleles were identified with a mean 9.4 alleles per marker. Most of the alleles (60.5%) were rare i.e., with the frequency ≤ 0.05 while only 3.95% of alleles occurred with frequency > 0.6 . EST-derived SSR markers were less polymorphic than genomic SSR markers. Differences in allele occurrence allowed 16 out of 18 populations to be assigned to either the Western or Asian carrot gene pools with high probability. Populations could be also discriminated due to the presence of private alleles (25.3% of all alleles). Most populations had excess of alleles in the homozygous state indicating their inbreeding, although heterozygous loci were common in F1 hybrids. Genetic diversity was due to allelic variation among plants within populations (62% of total variation) and between populations (38%). Accessions originating from continental Asia and Europe had more allelic variants and higher diversity than those from Japan and USA. Also, allelic richness and variability in landraces was higher than in F1 hybrids and open-pollinated cultivars.

Key words: SSR, biodiversity, genetic structure, gene pool

Received: 15 October, 2013; **revised:** 12 December, 2013; **accepted:** 17 December, 2013; **available on-line:** 28 December, 2013

INTRODUCTION

Carrots (*Daucus carota* L. subsp. *sativus*) are one of the most important vegetables grown in the world, which production has increased recently and reached 33.7 million tons harvested from the area of 1.16 million ha in 2010 (FAO 2010). Central Asia is considered as a center of origin of cultivated carrot. Primitive purple and yellow carrots evolved to the modern edible carrot of Eastern and Western types, which differ mainly in leaf morphology, root color and shape (Bradeen & Simon, 2007; Simon *et al.*, 2008). At present, large genetic variation is observed in cultivated carrot due to fast spread of carrot ancestors from their center of origin to distant geographical regions, and additionally due to a lack of control of random cross pollination between cultivated and wild forms. Carrot domestication, early selection and then breeding programs resulted in creation of varieties differing in several root morphological traits, tolerance to diseases and pests, and chemical composition. These traits are particularly well characterized for orange-colored carrot roots, which contain a high level of beta-carotene, and being one of the main sources of pro-vitamin A in human diet. Orange carrots are thus of significant

importance for human beings and are commonly grown in Europe and USA, but carrots developing roots of other colors i.e., red, yellow or purple are also rich in nutrients and compounds of health promoting properties including carotenoids like lutein or lycopene, phenolics and polyacetylenes. They exhibit high antioxidant activity enabling scavenging of reactive oxygen species, and were shown to be valuable in prevention from cancer, heart diseases and age related dysfunctions (Alsavvar *et al.*, 2001; Simon *et al.*, 2008). Description of *Daucus* diversity has been done initially based on morphological characteristics (Drude, 1898), and then supported by a deep insight in composition of the genera specific compounds (Hegnauer, 1971), recently extended to more health oriented screening (Baranski *et al.*, 2012a; Leja *et al.*, 2013). Several DNA marker systems like RAPD, AFLP and ISSR allowed distinction between *Daucus carota* accessions represented by wild and edible carrots, thus enabled genetic analysis of biodiversity. Polymorphism of simple sequence repeats (SSR) is a marker system extensively used in the assessment of genetic diversity in many plant species (Kalia *et al.*, 2011). Recently, new SSR were identified in carrot (Niemann, 2001; Cavagnaro *et al.*, 2011) and were used for the evaluation of carrot genetic resources (Baranski *et al.*, 2012b) extending earlier assessment by Cloutault *et al.* (2010). Their use enabled diversity evaluation in genetic resources of edible carrot and lead to identification of two separate carrot gene pools. The first gene pool, being more diverse, comprised a group of varieties of Eastern type and originating mainly from Asia, and the second gene pool included accessions originated from Europe, USA and Japan of Western type. This division was even more clearly evidenced after genotyping using DaRT (Diversity Array Technology) markers, that additionally allowed separation of cultivated and wild carrots (Grzebelus *et al.*, 2013), and independently depicted by Iorizzo *et al.* (2013) by SNP (Single Nucleotide Polymorphism) analysis. So far, evaluation of genetic diversity was based on the analyses of single individuals randomly selected from the populations, thus, the conclusions drawn have not accounted for any intra-population variation. Intra-population variation is highly expected as carrot is an open-pollinating species, exhibit strong inbreeding depression after self pollination and additionally easily hybridizes with wild relatives. Early work by

[✉]e-mail: baranski@ogr.ur.krakow.pl

*Presented at the 5th Central European Congress of Life Sciences „EUROBIOTECH 2013”, Kraków, Poland.

Abbreviations: AFLP, amplified fragment length polymorphism; DaRT, diversity array technology; SNP, single nucleotide polymorphism; EST, expressed sequence tag; ISSR, inter-simple sequence repeat; OP, open pollinated cultivars; RAPD, random amplified polymorphic DNA; SSR, simple sequence repeat (microsatellite); eSSR (ESSR), EST-derived SSR; gSSR (GSSR), genomic SSR

Shim & Jørgensen (2000), who used AFLP markers, indicates that large amounts of *Daucus* genetic variation results from the differences between individuals in local carrot populations. This observation was extended to breeding materials by the analysis of polymorphisms of *DeMaster* transposon insertions (Macko & Grzebelus, 2008). Thus, random selected individual may not be representative for a population. Therefore, the aim of present studies was to characterize intra-population variability of carrot populations of various origin using SSR marker approach, and to validate previous conclusions on two gene pool structure of carrot genetic resources.

MATERIALS AND METHODS

Plant material. Plant material was grown in experimental station of the University of Agriculture in Krakow, Poland and comprised 18 carrot (*Daucus carota* L. subsp. *sativus*) populations originating from Europe (9), Asia (6), USA (1) and Japan (2). They represented two F1 hybrids, ten open pollinated cultivars and six landraces. Fourteen of those populations were used in the previous assessment of genetic diversity based on single plant analysis using SSR markers (Baranski *et al.*, 2012b) and 15 populations using DArT markers (Grzebelus *et al.*, 2013). Seeds were obtained from gene bank collections, research institutes and breeding companies. Plants were grown in a field in 3 m long rows. Random 9–20 plants were selected from each population depending on the number of growing plants available and technical limits of DNA analysis. Low number of plants in some populations resulted from poor seed germination. In total, 249 plants were used.

DNA isolation. Young and healthy leaves were collected from plants, frozen in liquid nitrogen, freeze dried and then ground in a mill (Retsch MM20) to a fine powder. DNA was isolated using a CTAB method according to Rogers & Bendich (1988). DNA concentration and purity was determined spectrophotometrically using Nano Drop 2000 (Thermo Scientific) spectrophotometer and the DNA quality was determined electrophoretically in 1% agarose gel (1h/120V). The DNA was diluted to the final concentration 50 ng/μl.

DNA amplification and capillary separation. Sequences of 27 primer pairs were taken from Niemann (2000) — hereafter marked as DCM, Rong *et al.* (2010) — hereafter marked as R-SSR and from Cavagnaro *et al.* (2011) — hereafter marked as GSSR and ESSR markers. ESSR markers were EST-derived SSR markers, while the remaining markers were genomic SSR. Fragment amplification was performed according to a modified procedure by Schuelke (2000). Amplification of one SSR marker was carried out in two stages with the use of three primers: a sequence specific forward primer extended by 18nt M13 tail at its 5' end, a sequence specific reverse primer and the universal M13 (TGT AAA ACG ACG GCC AGT) primer, which was fluorescent-labeled by one of the three WellRED (Sigma-Aldrich) dyes: D2 (NED), D3 (HEX), and D4 (6-FAM). The primer ratios of the forward to reverse to universal primers were 1/4:1:1. Polymerase chain reactions were carried separately for each marker in the 10 μl reaction mixture containing 50 ng DNA, 1× reaction buffer (Fermentas), 0.25 mM MgCl₂, 0.125 μM forward primer, 0.5 μM reverse and universal primer each, 0.5 U TrueStart polymerase (Fermentas) using the Mastercycler Gradient thermocycler (Eppendorf). Amplifications were carried out in two steps with the first 10 cycles facilitating annealing temperatures for

specific primers and the next 30 cycles in 53°C, the optimal temperature for annealing the universal fluorescent-labeled M13 primer. The amplification parameters were as follow: initial denaturation at 94°C/5 min, 10 cycles (94°C/45 s, 65–45°C/60 s (depending on a primer used), 72°C/80 s), 30 cycles (94°C/45 s, 53°C/60 s, 72°C/80 s), final extension 72°C/10 min. PCR products labeled with D4 dye were five times more diluted than PCR products labeled with D2 and D3 dyes. Dilution of PCR-D4 products was necessary to decrease their high fluorescence. Amplified products of three SSR markers obtained in independent reactions and labeled by three different dyes were pooled together and genotyped in one run using an automatic sequencer CEQ™8000 Genetic Analysis System (Beckman Coulter Ltd.). Two size standards, 400 nt or 600 nt (CEQ™ DNA Size Standard) labeled with red fluorochrome WellRED (D1-ROX) were used.

Analysis of genetic diversity. Allele frequencies were used to calculate indices of marker information content and genetic diversity implemented in GenAlEx v.6.5 (Peakall & Smouse, 2006; 2012) and CERVUS (Kalinowski *et al.*, 2007) i.e., total (A^t), effective (A^e) and mean (A^m) number of alleles per locus and per population, Polymorphic Information Content (PIC), Shannon's index (I), unbiased expected heterozygosity (H_u), observed heterozygosity (H_o), Wright's inbreeding coefficient (F), Allelic richness (R) and private allelic richness (R_p) based on the rarefaction algorithm were calculated using HP-RARE (Kalinowski, 2005). Molecular variance (AMOVA) and F statistics (F_{st} , F_{it} , F_{is}) were used for the assessment of genetic diversity in GenAlEx v.6.5 (Peakall & Smouse, 2006, 2012). The significance of differences between means was assessed after ANOVA and Newman-Keuls test ($p = 0.05$) implemented in Statistica Pl v.9.1 (StatSoft Inc. 2010). Genetic structure was assessed using Bayesian clustering approach without information on the accession origin and assuming the admixture model and correlated allele frequencies (STRUCTURE 2.3.4; Pritchard *et al.*, 2000). Three independent simulations with a burn in length 10^4 and a run length of 10^5 were used for each number of clusters K set from 2 to 6. For the most likely cluster number run parameters were increased by the factor of 10.

RESULTS

All 27 analyzed SSR markers were polymorphic and there were no duplicates in the collection. In total, 253 alleles were identified in 249 plants representing 18 populations. The number of identified alleles range from 2 to 18 depending on the locus with a mean of 9.37 ± 0.85 per locus. Length of the amplified fragments ranged from 142 to 540 bp. Most of the alleles (73.52%) had frequencies below 0.1, while only 3.95% had the frequencies above 0.6. The effective number of alleles was almost three times lower than the total number of alleles observed in the locus. This discrepancy was due to the presence of rare alleles (occurring with the frequency ≤ 0.05), which in the analyzed collection accounted for more than half of the identified alleles (60.47%). Rare alleles were present in all loci except two loci, ESSR 0028 and ESSR 0010. Sixty four private alleles were identified in 23 loci that constituted for 25.3% of all alleles (Table 1). The PIC indices ranged from 0.53 to 0.85 and from 0.46 to 0.83 for GSSR and DCM markers, respectively, and on average were higher than for ESSR and R-SSR (0.11–0.72 and 0.28–0.81, respectively) that was also reflected by Shannon's index. The expected het-

Table 1. Characteristics of 27 SSR loci in 18 carrot populations represented by 9–20 plants per population

Locus	N	A ^t	A ^e	A ^p	A ^r	R	R ^p	I	PIC	H _o	H _e	F	F(Null)	HW
R-SSR-1-3C	224	10	1.4	0	7	1.57	0.22	0.74	0.28	0.20	0.28	0.29	0.19	***
R-SSR-1-9C	233	11	2.16	2	8	1.62	0.15	1.21	0.51	0.30	0.54	0.45	0.29	***
R-SSR-2-10C	236	6	2.32	3	4	1.75	0.10	0.98	0.48	0.41	0.57	0.29	0.17	***
R-SSR-2-10H	196	8	3.03	2	4	1.95	0.07	1.34	0.61	0.46	0.67	0.31	0.19	***
R-SSR-2-7A	178	12	6.04	5	7	2.14	0.02	1.95	0.81	0.40	0.84	0.52	0.35	***
DCM 12	231	5	2.25	1	3	1.60	0.06	0.93	0.46	0.07	0.56	0.88	0.78	ND
DCM 2	228	18	6.5	4	13	2.24	0.31	2.12	0.83	0.54	0.85	0.36	0.22	***
DCM 3	238	13	3.43	6	8	2.02	0.21	1.60	0.68	0.49	0.71	0.31	0.20	***
DCM 5	235	12	3.64	3	8	2.13	0.18	1.68	0.70	0.63	0.73	0.14	0.08	***
DCM 9	181	11	3.92	1	8	1.90	0.18	1.65	0.71	0.33	0.75	0.55	0.40	***
DCM17	227	8	5.08	3	3	2.16	0.13	1.77	0.78	0.56	0.80	0.30	0.17	***
ESSR 0009	223	9	2.64	4	6	1.9	0.12	1.28	0.57	0.51	0.62	0.18	0.1	***
ESSR 0028	188	3	2.33	0	0	1.71	0.00	0.95	0.50	0.22	0.57	0.61	0.44	***
ESSR 0010	244	2	1.13	0	0	1.12	0.01	0.23	0.11	0.10	0.12	0.15	0.08	ND
ESSR 0011	200	8	2.7	3	5	1.82	0.12	1.20	0.56	0.38	0.63	0.40	0.26	***
ESSR 0027	240	5	1.17	2	3	1.23	0.06	0.34	0.14	0.09	0.15	0.37	0.23	ND
ESSR 0032	237	4	1.19	1	2	1.08	0.09	0.35	0.15	0.05	0.16	0.68	0.49	ND
ESSR 0042	238	9	3.94	2	4	2.33	0.08	1.64	0.72	0.67	0.75	0.10	0.05	***
ESSR 0053	226	7	1.31	3	5	1.32	0.11	0.55	0.23	0.16	0.24	0.30	0.17	ND
GSSR 101	214	16	6.09	6	9	2.12	0.26	2.15	0.82	0.31	0.84	0.63	0.45	***
GSSR 107	247	11	3.79	2	6	2.16	0.18	1.68	0.70	0.57	0.74	0.22	0.12	**
GSSR 134	213	15	4.44	4	10	2.29	0.22	1.85	0.75	0.64	0.77	0.18	0.09	NS
GSSR 14	243	18	7	4	13	2.21	0.31	2.30	0.85	0.55	0.86	0.35	0.21	***
GSSR 17	246	6	3.8	1	1	1.78	0.01	1.47	0.70	0.42	0.74	0.44	0.27	***
GSSR 4	224	14	4.25	0	10	2.39	0.26	1.94	0.74	0.69	0.77	0.10	0.04	**
GSSR 6	238	5	3.87	1	1	2.01	0.05	1.43	0.70	0.58	0.74	0.22	0.12	***
GSSR154	205	7	2.48	2	4	1.74	0.12	1.13	0.53	0.23	0.6	0.61	0.44	***
mean		9.37	3.4	2.41	5.63	1.86	0.14	1.35	0.56	0.39	0.61	0.37	0.24	
se		0.85	0.32	0.33	0.69	0.07	0.02	0.11	0.04	0.04	0.04	0.04	0.03	
R-SSR	mean	9.4ab	3.0ab	2.4a	6.0a	1.8ab	0.1a	1.2ab	0.5ab	0.4a	0.6ab	0.4a	0.2a	
se		0.96	0.72	0.73	0.75	0.03	0.1	0.18	0.08	0.04	0.08	0.04	0.03	
DCM	mean	11.2a	4.1a	3.0a	7.2a	2.0a	0.2a	1.6a	0.7a	0.4a	0.7a	0.4a	0.3a	
se		1.66	0.55	0.71	1.4	0.03	0.09	0.15	0.05	0.08	0.04	0.1	0.09	
ESSR	mean	5.9b	2.1b	1.9a	3.1a	1.6b	0.1a	0.8b	0.4b	0.3a	0.4b	0.3a	0.2a	
se		0.91	0.34	0.48	0.76	0.02	0.15	0.17	0.08	0.07	0.09	0.07	0.05	
GSSR	mean	11.5a	4.5a	2.5a	6.8a	2.1a	0.2a	1.7a	0.7a	0.5a	0.8a	0.3a	0.2a	
se		1.65	0.47	0.66	1.48	0.04	0.08	0.13	0.03	0.05	0.03	0.07	0.05	

N, number of plants, number of alleles: A^t – total, A^e – effective, A^p – private, A^r – rare, R – allelic richness, R^p – private allelic richness, PIC – polymorphic information content, I – Shannon's index, H_o – observed heterozygosity, H_e – expected heterozygosity, F – fixation index, F(Null) – fixation index without null alleles, significant deviation from Hardy-Weinberg equilibrium at $p = 0.01$ (**), and 0.001 (***); ns, non-significant; nd, not determined; se, standard error

erogosity of the markers was usually higher than observed heterozygosity. However, out of 27 loci only one GSSR 134 was in Hardy-Weinberg equilibrium that could affect meaning of H_e and Wright's inbreeding coefficients. The mean value of Wright's inbreeding coef-

ficient for all 27 loci was above zero (0.37) and ranged between 0.1 and 0.8. About 30% lower inbreeding coefficients were obtained after taking into account the presence of possible null alleles, which hypothetical presence was estimated based on the observed and expect-

Table 2 Comparative parameters of genetic diversity for 18 carrot populations

AS	Popula- tion	C	R	S	GP	N	A ^t	A ^m	A ^e	A ^p	A ^r	R	R ^p	I	PIC	H _o	H _e	F
12	Persia No, 242	IRN	A	LR	A	14	96	3.56	2.23	4	17	2.08	0.09	0.88	0.43	0.44	0.51	0.11
22	Pusa Ke- sar	IND	A	LR	A	11	94	3.48	2.46	7	11	2.19	0.25	0.93	0.46	0.48	0.55	0.08
23	Panipat Special	IND	A	LR	A	12	95	3.52	2.33	6	20	2.15	0.14	0.91	0.45	0.44	0.54	0.12
37	Gajar (10146)	PAK	A	LR	A	13	64	2.37	1.73	0	6	1.71	0.08	0.57	0.29	0.33	0.36	0.00
38	Mestnaya (13403)	RUS	A	LR	na	9	69	2.56	2.01	3	0	1.92	0.18	0.70	0.36	0.38	0.44	0.06
49	Gajar (10197)	IND	A	LR	na	11	76	2.81	2.07	0	10	1.94	0.07	0.73	0.29	0.41	0.45	0.08
01	White Satin	NLD	E	F1	W	14	67	2.48	1.74	2	15	1.71	0.09	0.57	0.28	0.49	0.36	-0.34
02	Yellow- stone	NLD	E	OP	W	15	105	3.89	2.63	12	17	2.22	0.35	1.00	0.48	0.48	0.55	0.09
51	Red Ele- phant	GBR	E	OP	W	15	87	3.22	2.12	2	9	2.00	0.13	0.82	0.41	0.41	0.48	0.14
07	Deep Purple	NLD	E	F1	A	20	90	3.33	2.21	3	14	1.97	0.13	0.79	0.39	0.5	0.45	-0.14
10	Kuettiger	CHE	E	OP	W	15	78	2.89	2.02	9	7	1.90	0.34	0.73	0.37	0.43	0.43	0.00
11	Blanche 1/2 lon- gue des vosges	FRA	E	OP	W	14	62	2.3	1.76	3	4	1.72	0.17	0.57	0.30	0.36	0.36	-0.04
20	Nantejska Polana	POL	E	OP	W	14	90	3.33	2.29	2	18	2.02	0.07	0.83	0.41	0.45	0.47	0.03
33	Amster- dam 3	POL	E	OP	W	15	79	3.00	2.09	5	7	1.94	0.13	0.76	0.24	0.37	0.44	0.14
47	Yellow Belgian	NLD	E	OP	W	16	41	1.52	1.43	0	0	1.43	0.03	0.33	0.18	0.3	0.24	-0.28
19	Himuro Fuyugosi Gosun No, 2	JPN	J	OP	W	16	54	2.00	1.48	1	8	1.50	0.09	0.4	0.21	0.22	0.27	0.21
31	Shima Ninjin	JPN	J	OP	A	12	82	2.26	1.51	5	12	1.52	0.12	0.43	0.28	0.18	0.25	0.15
04	Nutried	USA	USA	OP	A	13	62	2.30	1.66	1	6	1.65	0.08	0.52	0.27	0.34	0.33	-0.11

AS, AS No. used in references Baranski *et al.* (2012b) and Grzebelus *et al.* (2013); C, country of origin; R, region of origin (A — Asia, E — Europe, J — Japan), S — Biological status (LR — landrace, OP — open-pollinated cultivar, F1–F1 hybrid); GP, Gene pool assigned based on Bayesian approach: A — Asian, W — Western, na — not assigned. N, number of plants; number of alleles: A^t — total, A^m — mean number of alleles, A^e — effective, A^p — private, A^r — rare, R — allelic richness, R^p — private allelic richness, I — Shannon's index, PIC — polymorphic information content, H_o — observed heterozygosity, H_e — expected heterozygosity, F — fixation index

ed frequencies (Table 1). In general, DCM and GSSR markers were more informative and heterozygous than R-SSR markers and the least were ESSR markers. The first two marker types identified about 20% more allele variants than ESSR markers based on the estimation of allelic richness R. Similar tendency was observed for the presence of private alleles (private allelic richness R^p) although the difference between the marker types was not confirmed statistically. DCM and GSSR markers had also higher PIC, I and H_e values.

A Bayesian clustering method was used to indicate the most probable structure of the whole collection of 18 populations. The results showed the material is the most probably divided in two clusters (K=2) of seven and nine populations. Thus 16 out of 18 populations were

assigned to one of the two clusters with probability above 0.9. In four populations, single plants had lower assignment probability (>0.6). Only two individuals, one in 'Panipat Special' and one in 'Shirma Ninjin' failed to be assigned to the same cluster as the remaining plants of these populations. Only 2 out of 18 populations, landraces 'Mestnaya' (AS 50) and 'Gajar' (AS 49), could not be assigned to any cluster as their mean assignment probability ranged between 0.4–0.5 and individual plants showed affinity to either cluster 1 or cluster 2 (Fig. 1). Assignment of most population of Eastern type to one cluster, and most populations of Western type to another with high probability and additionally high congruence with results obtained for the assignment of the same accessions by Grzebelus *et al.*, (2013) indicates that cluster

Table 3. Comparative parameters of genetic diversity for 18 carrot populations grouped according to their biological status, origin and gene pool

Grouping factor			A ^t	A ^m	A ^e	A ^p	A ^r	R	R ^p	I	PIC	H _o	H _e	F
Status	F1 hybrids	mean	78.50a	2.91a	1.97a	2.50a	14.50a	1.84b	0.11a	0.68ab	0.34b	0.50a	0.40b	-0.24b
		se	8.13	0.30	0.17	0.35	0.35	0.09	0.01	0.08	0.04	0.00	0.03	0.07
	Landraces	mean	82.33a	3.05a	2.14a	3.33a	10.67	2.00a	0.14a	0.79a	0.38a	0.41b	0.47a	0.07a
		se	5.37	0.20	0.10	1.10	2.70	0.07	0.03	0.05	0.03	0.02	0.03	0.02
	OP cultivars	mean	74.00a	2.67a	1.90a	4.00a	8.80a	1.79b	0.15a	0.64b	0.31b	0.35b	0.38b	0.03a
		se	5.72	0.21	0.12	1.16	1.67	0.08	0.03	0.07	0.03	0.03	0.03	0.04
Region	Asia	mean	82.33a	3.05a	2.14a	3.33a	10.67	2.00a	0.14a	0.79a	0.38a	0.41a	0.47a	0.07a
		se	5.37	0.20	0.10	1.10	2.70	0.07	0.03	0.05	0.03	0.02	0.03	0.02
	Europe	mean	77.67a	2.88a	2.03a	4.22a	10.11a	1.88a	0.16a	0.71a	0.34a	0.42a	0.42a	-0.04a
		se	5.93	0.22	0.11	1.21	1.95	0.07	0.04	0.06	0.03	0.02	0.03	0.05
	Japan	mean	68.00b	2.13b	1.50b	3.00a	10.00a	1.51b	0.11a	0.42b	0.25b	0.20b	0.26b	0.18a
		se	9.90	0.09	0.01	1.41	1.41	0.01	0.01	0.01	0.03	0.01	0.01	0.02
	USA	mean	62.00b	2.30b	1.66b	1.00a	6.00a	1.65b	0.08a	0.52b	0.27b	0.34a	0.33b	-0.11a
		se	-	-	-	-	-	-	-	-	-	-	-	-
Gene pool	Asian	mean	83.29a	2.97a	2.01a	3.71a	12.30	1.89a	0.12a	0.72a	0.36a	0.38a	0.42a	0.03a
		se	5.12	0.22	0.13	0.9	1.84a	0.09	0.02	0.07	0.03	0.04	0.04	0.04
	Western	mean	73.67a	2.73a	1.95a	4.00a	9.44a	1.82a	0.15a	0.66a	0.32a	0.39a	0.40a	-0.01a
		se	6.19	0.23	0.12	1.26	1.91	0.08	0.04	0.07	0.03	0.03	0.03	0.06

A^t — total, A^m — mean number of alleles, A^e — effective, A^p — private, A^r — rare, R — allelic richness, R^p — private allelic richness, I — Shannon's index, PIC — polymorphic information content, H_o — observed heterozygosity, H_e — expected heterozygosity, F — fixation index

1 represents Asian gene pool and cluster 2 represents Western gene pool. Thus, the Western gene pool comprised mainly open-pollinated (7) cultivars, modern hybrid 'White Satin' originated from Europe and open pollinated 'Himuro Fuyugosi Gosun No.2' from Japan. The Asian gene pool comprised Asian landraces (4), open-pollinated cultivars from Japan ('Sirma Ninjin') and USA ('Nutrired'), and one hybrid ('Deep Purple') of European origin (Table 2).

AMOVA indicated that 62% of total variation resulted from differences between accessions within populations (intra-population), while the remaining 38% resulted from differences between 18 populations (inter-population). The overall F_{it} statistics (0.47) indicated that the collection was highly diverse. This diversity was partially due to differences between individuals within populations (F_{is} = 0.26), and partially to differences between populations (F_{st} = 0.28). The populations differed in private allelic richness that ranged from 0.07 to 0.35. Two European populations 'Yellowstone' and 'Kuettiger' had the higher number of private alleles that constituted 11–12% of all alleles identified in these populations. Three populations had no private alleles (Table 2). Additionally, 6–22% of alleles were rare depending on the population. Only two populations had no rare alleles. Most populations had low fixation indexes indicating low inbreeding and the excess of heterozygotes was noticed in four populations. The highest intra-population diversity was identified in Asian landraces 'Pusa Kesar' (AS 22) and 'Panipat Special' (AS23) developing red storage roots,

and in yellow rooted Asian landrace 'Pesia No.242' (AS 12) and European OP 'Yellowstone' (AS 02). Asian and Western gene pools did not differ significantly with regard to allelic and diversity parameters. However, populations from Asia and Europe were richer in allelic variants than those from Japan and USA, and tended to have higher private allelic richness. They were also more diverse as both PIC and H_e were higher by about 40% than those for Japan and USA populations. In contrast, population from Japan had higher share of rare alleles (15%) than populations from Europe and Asia (12% and 14%, respectively). However, parameters for populations from Japan and USA might be underestimated as the used number of accessions were five times lower than those counted for European and Asian origin. The total and mean number of alleles, and the number of effective and private alleles did not differ between landraces and advanced cultivars. However, landraces had higher allelic richness, PIC and H_e than open-pollinated and F1 hybrid cultivars indicating their higher diversity. Both F1 hybrid and OP cultivars had higher percentage of rare as well as frequent alleles in comparison to landraces. F1 hybrids had also highly negative fixation index resulting from the excess of heterozygous loci.

DISCUSSION

Microsatellite markers are powerful tools to characterize population structure as they may show high pol-

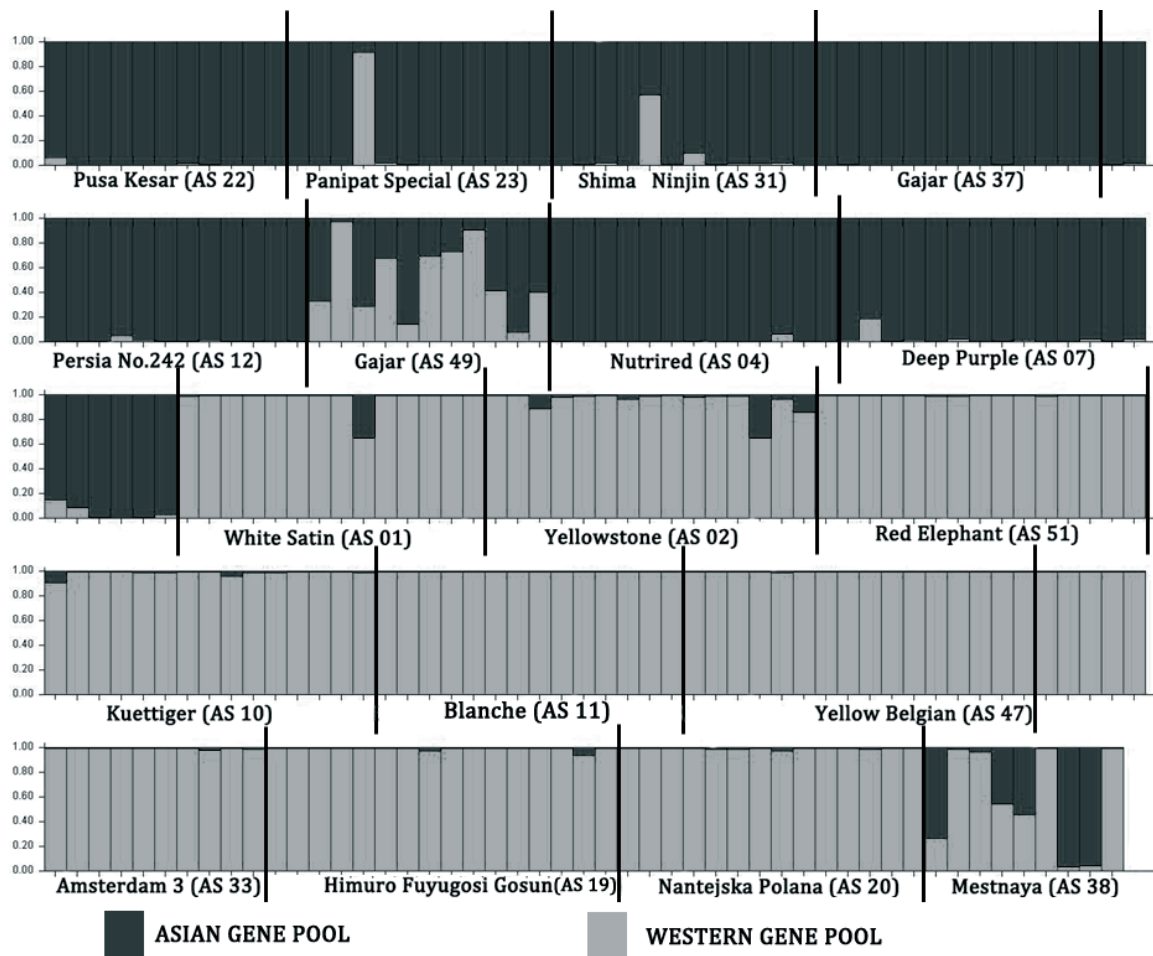


Figure 1. Structure of the genetic diversity of 18 carrot populations represented by 9–20 individuals based on a Bayesian approach and assuming two gene pools.

ymorphic information content due to multiallelic and highly variable status (Powell *et al.*, 1996). The mean PIC value obtained for all loci in the present study was high (PIC=0.56) and similar to PIC value (0.6) obtained previously by Baranski *et al.* (2012b) and Clotault *et al.* (2010). Lower discriminatory power was observed for dominant markers such as DArT (PIC=0.34) used by Grzebelus *et al.*, (2013), however, the higher number of DArT markers, which is generated during the analysis, finally resulted in clearer division of carrot resources to two cultivated forms that additionally differ from wild relatives. In the present study, the highest discriminatory power was observed for GSSR markers developed by Cavagnaro *et al.* (2011) and DCM developed by Niemann (2000). In contrast, the least informative were ESSR markers (Cavagnaro *et al.*, 2011) that might be expected as ESSR markers were developed based on the screening of coding sequences (EST-derived SSR markers) while the former two SSR marker types (GSSR and DCM) were obtained by screening the whole genomic DNA. Lower values of several parameters, i.e., Shannon's index, heterozygosity and allelic richness obtained for ESSR confirm lower discriminating power of these markers in this diversity study.

Molecular evidence for the existence of two carrot gene pools was postulated by Clotault *et al.* (2010), who screened 47 accessions. That was further extended and supported by SSR analyses of a larger collections of 88 accessions (Baranski *et al.*, 2012b) as well as by DArT (Grzebelus *et al.*, 2013) and SNP analyses (Iorizzo *et al.*,

2013). Those reports enabled the assignment of populations used in this work to either the Asian or Western gene pools, although the assignment done were previously based on the analysis of only one plant representing each population. In this work, we extended analyses to 9–20 individuals per population to verify the previous assignments and to quantify genetic intra-population diversity. The assignment of populations to gene pools fully agrees with the assignment based on SSR markers by Baranski *et al.* (2012b) taking into account the correction implemented in Grzebelus *et al.* (2013). The assignment of two landraces, 'Gajar' (AS 49) previously assigned to the Asian gene pool and 'Mestnaya' (AS 38) previously assigned to the Western gene pool could not be confirmed. That resulted from a high diversity of both populations, which shared alleles common for both gene pools. Thus, their assignment in previous studies have to be verified, and both populations cannot be unequivocally assigned to either gene pool. It is also evident from the presented results that the remaining populations evaluated here consisted of plants sharing the same sets of alleles allowing their assignment to a given gene pool with a very high probability. Occasionally, individual plants were assigned to other gene pool than the majority of a given population. Such ambiguity was found regardless of population origin or biological status. The frequency of identified outliers (probability <0.9) ranged from 7% to 13% and determines uncertainty of an assignment to a gene pool when allelic composition of a single plant from a population is considered. The as-

signment of some populations to gene pools may seem awkward without detailed insight. Two cultivars from Japan were assigned to different gene pools, 'Shirma Ninjin' producing yellow roots to the Asian gene pool and 'Himuro Fuyugosi Gosun No.2' with orange roots to the Western gene pool. It is documented that in breeding programs in Japan carrot of various origin were frequently used and crossed. Carrot of the Eastern type was introduced to Japan from continental Asia first, but European and American varieties were later used in crossing programs to enhance carotenoid levels and improve root morphological traits in new cultivars (Simon *et al.* (2008). Thus, both gene pools have contributed to carrot breeding in Japan. This seems to be an argument in favor of Cloutault *et al.*, (2010), who postulated that in spite of intensive breeding since 19th century the background genetic structure resulting from demographic history and early cultivation still persist in cultivated carrot germplasm nowadays. On the other hand, novelty traits are demanded by European and American consumers. They are willing to have carrots of high quality and nutritional value that can be assured, for example by elevating lutein, lycopene or anthocyanins contents through breeding programs (Arscott & Tanumihardjo, 2010). Yellow, red or purple/black colored roots not only attract visually, but may also be valuable source of pigments for industrial use (Kammerer *et al.*, 2004). Thus, in recent years several cultivars introduced to western market exhibit root color other than orange. As the selection of European carrots was carried out to fix orange color, Western type carrots unlikely contain anthocyanins or lycopene. The genetic determinants responsible for lycopene accumulation and anthocyanins biosynthesis in roots, and existing in the Eastern carrots, have to be transferred to Western carrots by hybridization. Consequently, a modern F1 hybrid 'Deep Purple' of the Western type and rich in anthocyanins was assigned here to the Asian gene pool. An analogous case concerns the assignment of the lycopene rich, red 'Nutrired' (Bradeen & Simon, 2007; Simon *et al.*, 2008).

Two third of total variation was attributed to intra-population variation although the collection used contained accessions of diverse origin. High number of identified alleles confirms a high level of intra-populations variation. Because the populations were represented by different number of plants, the parameter of allelic richness is more accurate for comparison purposes. Landraces had higher allelic richness than F1 hybrids and OP cultivars. This is congruent with the fact of less intense selection pressure during landrace reproduction in comparison to the stringent uniformity check in commercial cultivars to assure registration requirements. Surprisingly, we have not detected higher diversity in OP cultivars in comparison to F1 hybrids, but only two hybrids were considered in the evaluation so the mean parameters might be highly biased. Also, accessions originating from continental Asia and Europe were more diverse than those of Japanese and USA origin. It should be, however, noted that accessions of Asian origin were landraces and those from other regions were OP cultivars or F1 hybrids. Thus, it is not clear whether differences in diversity parameters are due to the fact that some populations originated from Asia or that they were landraces. Finally, probably due to a low number of populations, we have not found significant differences in genetic diversity parameters between Asian and Western gene pools as was shown previously using both SSR and DArT markers (Baranski *et al.*, 2012b; Grzebeul *et al.*, 2013). Most of the analyzed populations had

lower values of observed heterozygosity than expected heterozygosity, and positive values of inbreeding coefficients, indicating excess of alleles in the homozygous state. The excess of homozygous loci resulted most likely from repeated selfing during breeding programs, but the calculated parameters might be overestimated due to potential presence of null alleles that increase score of homozygotes even though a locus is heterozygous (Table 1). Much lower, hence negative inbreeding coefficients were observed for F1 hybrids that confirmed a predicted excess of heterozygotes.

The results obtained support earlier opinion that cultivated carrot is largely genetically non-structured Bradeen *et al.* (2002), and as an out-crossing species shows less intense population differentiation and more uniform distribution of genetic diversity than inbreeding species (Brown, 1989). High intra-population variation is observed in other allogamous species, too. The use of 56 eSSR and 4 gSSR markers allowed discrimination of 89 inbred and open pollinated bulb onion populations according to their heterozygosity, which was higher for landraces and cultivars in comparison to inbred lines. They also differentiated accessions originating from temperate climate regions from those typical for tropical regions (McCallum *et al.*, 2008). In maize genetic diversity within 63 populations was significantly higher than diversity between them, indicating that the populations were highly heterogeneous. Also, inbred populations showed a high level of genetic diversity and only lines closely related by pedigree clustered together (Warburton *et al.*, 2002).

An evaluation of intra-population diversity indicates that plant to plant variation in carrot is substantial and might be a source of different alleles even if a phenotype variability is small as suggested by Simon *et al.* (2008). High genetic variation may have also substantial impact on preservation strategies carried out in gene bank collections as they are supposed to preserve both common and rare alleles (Bradeen & Simon, 2007). The number of individuals required for the replenishment of seed stock of an accession partly depend on the variability of the accession (Le Clerc *et al.*, 2005). The higher homogeneity the fewer parental plants are required to maintain the allelic diversity (Gilbert *et al.*, 1999). Conservation of rare alleles requires, however, larger mating population, particularly in allogamous species like carrot (Le Clerc *et al.*, 2003). Thus, information on genetic diversity helps curators to rationalize the use of resources required for seed reproduction and to preserve allelic variants (Gilbert *et al.*, 1999). Our results indicate that carrot intra-population genetic diversity varies depending on the accession, thus, identification and quantification of rare and private alleles in populations is of high importance for preservation strategy. Higher number of plants may be necessary to maintain gene diversity in highly heterogeneous populations like landraces 'Gajar' and 'Mestnaya'.

Acknowledgement

The research was supported by the Ministry of Science and Higher Education Grant No. NN310782440.

REFERENCES

- Alsavar C, Grigor JM, Zhang D, Quantick PC, Shahidi F (2001) Comparison of volatiles, phenolics, sugars, antioxidant, vitamins and sensory quality of different colored carrot varieties. *J Agric Food Chem* 49: 1410–1416.

- Arcott SA, Tanumihardjo SA (2010) Carrot of many colors provide basic nutrition and bioavailable phytochemicals acting as a functional food. *CRFSFS* **9**: 223–239.
- Baranski R, Allender C, Klimek-Chodacka M (2012a) Toward better tasting and more nutritious carrot: carotenoid and sugar content variation in carrot genetic resources. *Food Res Int* **47**: 182–187.
- Baranski R, Maksylewicz-Kaul A, Nothnagel T, Cavagnaro PF, Simon PW, Grzebelus D (2012b) Genetic diversity of carrot (*Daucus carota* L.) cultivars revealed by analysis of SSR loci. *Genet Resour Crop Evol* **59**: 163–170.
- Bradeen JM, Simon PW (2007) Carrot. In: *Genome mapping and molecular breeding in plants*. Kole C, eds, pp 161–184. Springer-Verlag, Heidelberg, Berlin.
- Bradeen JM, Bach IC, Briard M, Le Clerc V, Grzebelus D, Senalik DA, Simon PW (2002) Molecular diversity analysis of cultivated carrot (*Daucus carota* L.) and wild *Daucus* population reveals a genetically nonstructured composition. *J Amer Soc Hort Sci* **127**: 383–391.
- Brown AHD (1989) Core collections: a practical approach to genetic resources management. *Genome* **31**: 818–824.
- Cavagnaro PF, Chung S-M, Manin S, Yildiz M, Ali A, Alessandro MS, Iorizzo M, Senalik DA, Simon PW (2011) Microsatellite isolation and marker development in carrot — genomic distribution, linkage mapping, genetic diversity analysis and marker transferability across *Apiaceae*. *BMC Genomics* **12**: 386.
- Clotault L, Geoffriau E, Lionneton E, Briard M, Peltier D (2010) Carotenoid biosynthesis genes provide evidence of geographical subdivision and extensive linkage disequilibrium in the carrot. *Theor Appl Genet* **121**: 650–672.
- Drude O (1898) *Umbelliferae*. In *Die natürlichen Pflanzenfamilien*. Engler A, Prantl K, eds, pp 63–250. Wilhelm Engemann, Leipzig (in German).
- Food and Agriculture Organization of the United Nations (FAO, 2010) <http://faostat.fao.org>.
- Gilbert JE, Lewis RV, Wilkinson MJ, Caligari PDS (1999) Developing an appropriate strategy to assess genetic variability in plant germplasm collections. *Theor Appl Genet* **98**: 1125–1131.
- Grzebelus D, Iorizzo M, Senalik D, Ellison S, Cavagnaro P, Macko-Podgorni A, Heller-Uszynska K, Kilian A, Nothnagel T, Simon PW, Baranski R (2013) Diversity, genetic mapping and signatures of domestication in the carrot (*Daucus carota* L.) genome as revealed by Diversity Arrays Technology (DArT) markers. *Mol Breed* **10.1007/s11032-013-9979-9**.
- Hegnauer R (1971) Chemical patterns and relationships of *Umbelliferae*. In *The biology and chemistry of the Umbelliferae*. Heywood VH, eds, pp 267–277. Acad. Press, London.
- Iorizzo M, Senalik D, Ellison S, Grzebelus D, Cavagnaro PF, Allender C, Brunet J, Spooner D, Van Deynze A, Simon PW (2013) Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (*Apiaceae*). *Am J Bot* **100**: 930938.
- Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* **177**: 309–334.
- Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.
- Kalinowski ST, Taper ML, Marschall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* **16**: 1099–1106.
- Kammerer D, Carle R, Schieber A (2004) Quantification of anthocyanins in black carrot extracts (*Daucus carota* ssp. *sativus* var. *atrorubens* Alef.) and evaluation of their color properties. *Eur Food Res Technol* **219**: 479–486.
- Le Clerc V, Briard M, Granger J, Delettre J (2003) Genebank biodiversity assessments regarding optimal sample size and seed harvesting technique for the regeneration of carrot accessions. *Biodiversity and Conservation* **12**: 2227–2336.
- Le Clerc V, Suel A, Briard M (2005) Identification of duplicates for the optimization of carrot collection management. *Biodiversity and Conservation* **14**: 1211–1223.
- Leja M, Kaminska I, Kramer M, Maksylewicz-Kaul A, Kammerer D, Carle R, Baranski R (2013) The content of phenolic compounds and radical scavenging activity varies with carrot origin and root color. *Plant Foods Hum Nutr* **68**: 163–170.
- Macko A, Grzebelus D (2008) *DeMaster* Transposon Display as a tool for diversity evaluation of carrot breeding materials and testing hybrid seed purity. *J Appl Genet* **49**: 33–39.
- McCallum J, Thomson S, Pither-Joyce M, Kenel F, Clarke A, Havey MJ (2008) Genetic diversity analysis and single-nucleotide polymorphism marker development in cultivated bulb onion based on expressed sequences tag-simple sequence repeat markers. *J Amer Soc Hort Sci* **6**: 810–818.
- Niemann M (2001) Entwicklung von Mikrosatelliten – Marker bei der Möhre (*Daucus carota* L.) und die Markierung eines Alternaria-Resistenzgens. Berichte aus der Agrarwissenschaft. Shaker Verlag, Aachen.
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.
- Peakall R and Smouse PE (2012) GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* **28**: 2537–2539.
- Powell W, Marchray GC, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed* **2**: 225–238.
- Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. *Plant Molecular Biology Manual* **A6**: 1–10.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Rong J, Janson S, Umehara M, Ono M, Vrieling K (2010) Historical and contemporary gene dispersal in wild carrot (*Daucus carota* ssp. *carota*) population. *Ann Bot* **106**: 285–296.
- Shim SI, Jørgensen RB (2000) Genetic structure in cultivated and wild carrot (*Daucus carota* L.) revealed by AFLP analysis. *Theor Appl Genet* **101**: 227–233.
- Simon PW, Freeman RE, Vieira JV, Boiteux LS, Briard M, Nothnagel T, Michalik B, Kwon Y-S (2008) Carrot: Genetic resources, chromosome engineering and crop improvement series. In *Vegetables crops*. Singh RJ eds. Vol **3** pp 497–517. CRC Press, New York.
- Schuelke M (2000) An economic methods for the fluorescent labeling of PCR fragments. *Nat Biotechnol* **18**: 233–234.
- Stat Soft. Inc. (2010). STATISTICA (data analysis software system), version 9.1. www.statsoft.com.
- Wartburton ML, Xianchun X, Crossa J, Franc J, Melchinger AL, Frisch M, Bohn M, Hoisington D (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci* **42**: 1832–1840.