https://doi.org/10.17221/68/2018-CJGPB

Phenotyping and SSR markers as a tool for identification of duplicates in lettuce germplasm

*Michal Sochor, Michaela Jemelková, Ivana Doležalová**

Centre of the Region Haná for Biotechnological and Agricultural Research, Department of Genetic Resources for Vegetables, Medicinal and Special Plants, Crop Research Institute, Olomouc, Czech Republic

**Corresponding author: dolezalova@genobanka.cz*

Citation: Sochor M., Jemelková M., Doležalová I. (2019): Phenotyping and SSR markers as a tool for identification of duplicates in lettuce germplasm. Czech J. Genet. Plant Breed., 55: 110–119.

Abstract: In total, 117 individual samples from 39 accessions of *Lactuca sativa* were selected from the Czech national collection of lettuce with the aim to quantify and compare patterns of genetic and phenotypic variability within and among lettuce accessions and to propose a rapid, reliable and inexpensive method for verification of possible duplicates. We focused on phenotypic evaluation and SSR genotyping, and studied their ability to distinguish between individual accessions. Phenotypic data revealed that no two accessions shared the exactly same phenotype and no accession exhibited variability in the characters studied. Variability in SSR markers was very low as ten of twenty scorable SSR loci exhibited no variation and the remaining ten provided 48 alleles in total. Although neither phenotypic nor SSR data alone can serve as evidence for unambiguous duplicate confirmation, their combination increases the resolution power of the method considerably. The obtained data on cultivated lettuce indicate weak, but significant correlation (R^2 = 0.34, P = 0.01) between the two data sets.

Keywords: gene banks; germplasm collections; *Lactuca sativa*; microsatellites; variation

Plant genetic resources represent a unique natural source of genetic diversity irreplaceable for the further improvement of desired traits of crops, broadening of the genetic basis in the commercial varieties, and enhancing food safety and quality. The basic working unit of conservation in the genebanks is accession. It is estimated that about 7.4 million accessions of plant germplasm are currently maintained globally (FAO 2010). The majority of these resources are stored in *ex-situ* collections of national and/or regional gene banks (ca 90%), the remaining accessions (10%) are maintained within the Consultative Group for International Agricultural Research (CGIAR) system (Richerzhagen 2010). Various analyses suggest that between 25 and 30% of the total holdings are distinct, with the remainder being duplicates held either in the same, or more frequently, in different collections (FAO 2010). Identification and elimination of redundant (identical or near identical) accessions within and among gene banks are important aspects of efficient plant genetic resource management because they do not contribute to the genetic diversity of a collection, but require resources to maintain them (Spooner *et al.* 2005).

A terminology distinguishing different types of duplicates was proposed by van Hintum and Knüpffer (1995). Identical duplication refers to genetically identical gene bank accessions, i.e., material which has not been rejuvenated, or which is homogeneous and homozygous, or is propagated vegetatively. The most frequently occurring type of duplication in gene banks

Supported by the Ministry of Agriculture of the Czech Republic, Project No. RO0419 (Sustainable Systems and Technologies, Improving Crop Production for Higher Quality of Production of Food, Feed, and Raw Materials, under Conditions of Changing Climate).

https://doi.org/10.17221/68/2018-CJGPB

is common duplicates. These are derived from the same original population and share the same alleles but with different frequencies. Partial duplicates are two accessions that have been derived from the same original population but have only a part of the alleles or genotypes in common. Compound duplication is duplication, where one accession is a selection from another one.

Identification of duplicates based on passport data is usually problematic because of their incompleteness or poor quality, and the problem is also genetic identity of accessions which are subject to change during their maintenance in the gene banks (van Hintum & Knüpffer 1995). In order to detect duplicates within and between collections, and at the same time to avoid incorrect decisions, a comprehensive approach should be used. Correctness and completeness of passport data are the first conditions to identification of duplicates (Lipman *et al.* 1997). Potential duplicates based on passport data need to be verified by morphological comparison of accessions followed by any molecular marker technique that is able to detect genetic differences between and within accessions (van Treuren *et al.* 2010).

Existence of duplicates within and among collections is well documented in the case of lettuce (*Lactuca* L.) genetic resources. Study based on passport data of accessions in four world largest collections (CGN – Centre for Genetic Resources, NL; WRPIS – Western Regional Plant Introduction Station, US; IPK – Leibniz Institute of Plant Genetics and Crop Plant Research, DE; HRI – Horticultural Research Institute, UK), which include 28% of the worlds lettuce accessions, showed that only 40% accessions are not duplicated, 30% are duplicated once (they are present in two gene banks), and the remaining accessions are duplicated twice, or are in all studied collections (van Hintum & Boukema 1999).

Cultivated lettuce (*L. sativa*) is morphologically the most diverse species of the genus *Lactuca* with considerable amount of genetic variability (Waycott *et al.* 1999) caused by its polyphyletic origin and process of domestication (Kesseli *et al.* 1991). It is considered an obligate self-fertilizing species with the possibility of pollen transmission by insect (Rao *et al.* 2006). Although up to 5% cross-pollination has been observed in some areas, lettuce is regarded as a self-pollinated crop and only a physical barrier (e.g. adjacent sections of greenhouses) or a minimum of 2 m between different cultivars is recommended to prevent cross-pollination (George 1999). Although the predominant self-fertilization reduces the capacity for genetic recombination (Frankel & Galun 1977), the total genetic variability in autogamous species including lettuce may be considerable.

The Czech national collection of lettuce maintained at the Department of Genetic Resources for Vegetables, Medicinal and Special Plants of the Crop Research Institute in Olomouc, includes 844 accessions of landraces, historical and advanced cultivars, and breeding materials. An essential part of the collection (approx. 92%) is represented by old Czech cultivars, cultivars that were bred in former Czechoslovakia, new Czech cultivars, and foreign cultivars (Doležalová 2014). However, 55% of the total amount of accessions has been hypothesized to be probable duplicates according to passport data (Doležalová unpublished data).

This study aims (1) to quantify and compare patterns of phenotypic and genetic variability within and among accessions in the Czech national lettuce collection, and (2) to propose a rapid, reliable and inexpensive method for verification of possible duplicates in the lettuce collections.

MATERIAL AND METHODS

Plant material. In total, 39 accessions of *L. sativa* were selected from the collection with the aim to cover wide range of different lettuce morphotypes and varieties. The set included butterhead, crisphead, cutting and leaf morphotypes, varieties with or without anthocyanin, varieties for greenhouse and field cultivation, landraces and advanced varieties (Table 1). For each accession, 16 individual plants were planted in isolation cages in order to prevent possible outcrossing among varieties.

Phenotyping. Sixteen plants per accession were phenotyped for seventeen descriptors according to the Descriptor list for *Lactuca sativa* L. (http://genbank.vurv.cz/genetic/nar_prog_rostlin/klasifikatory/Lactuca.pdf): young leaf (anthocyanin distribution, anthocyanin pattern, shape of blade, division of blade (depth of lobes from blade margin to the main vein), venation); outer adult leaf (colour, anthocyanin distribution, anthocyanin pattern, blistering); harvested part (size of head and/or rosette); leaf head (shape in vertical section, overlapping of leaves, firmness); stem length including inflorescence; fruit seed coat colour, bolting and flowering. The descriptor No.1.2.3.1. (young leaf – anthocyanin distribution) was excluded from statistical analyses due to low variability and high correlation with descriptor No. 1.2.3.3. (young leaf – anthocyanin pattern). All traits

https://doi.org/10.17221/68/2018-CJGPB

Table 1. Within-accession genetic variability in lettuce as determined by 10 SSR markers

Genebank Acc. No.[a]	Variety	Morphotype	N_a	H_o	H_e
09H5700005[b]	Liban	butterhead	1.0	0.000	0.000
09H5700009	Herm	butterhead	1.0	0.000	0.000
09H5700014	Podripan	butterhead	1.1	0.067	0.044
09H5700018	Merkur	butterhead	1.1	0.033	0.028
09H5700021	Altenbursky	butterhead	1.1	0.100	0.050
09H5700027	Kralovna maje	butterhead slightly red	1.1	0.000	0.044
09H5700028	Melnicky maj	butterhead	1.1	0.000	0.044
09H5700032	Hanacky Letni	butterhead	1.0	0.000	0.000
09H5700279	Vodnansky	butterhead	1.0	0.000	0.000
09H5700716	Kamex	butterhead	1.0	0.000	0.000
09H5700717	Karat	butterhead	1.1	0.100	0.050
09H5700718	Nefrit	butterhead	1.1	0.000	0.050
09H5700721	Tyrkys	butterhead	1.0	0.000	0.000
09H5700737	Mars	butterhead	1.1	0.033	0.028
09H5700738	Saturn	butterhead	1.0	0.000	0.000
09H5700834	Hrdelsky	butterhead	1.0	0.000	0.000
09H5701139[b]	Krystal	crisphead	1.0	0.000	0.000
09H5701224	Orion	butterhead	1.0	0.000	0.000
09H5701266[b]	Tarzan	crisphead	1.0	0.000	0.000
09H5701267	Redin	leaf red	1.0	0.000	0.000
09H5701268	Maraton	butterhead	1.0	0.000	0.000
09H5701279	Deon	butterhead	1.0	0.000	0.000
09H5701324	Faraon	butterhead	1.0	0.000	0.000
09H5701373	Lento	butterhead	1.0	0.000	0.000
09H5701374[b]	Rosela	leaf red	1.0	0.000	0.000
09H5701567	Rosemarry	butterhead	1.0	0.000	0.000
09H5701568	Traper	crisphead	1.0	0.000	0.000
09H5701569	Crimson	leaf red	1.3	0.000	0.133
09H5701570[b]	Rekord	cutting	1.2	0.000	0.089
09H5701571	Amur	butterhead	1.0	0.000	0.000
09H5701572	Merlot	cutting	1.0	0.000	0.000
09H5701573	Maugli	crisphead	1.0	0.000	0.000
09H5701576	Zeus	butterhead	1.0	0.000	0.000
09H5701578	Podivin	butterhead	1.0	0.000	0.000
09H5701580[b]	Panter	butterhead	1.1	0.000	0.044
09H5701581	Sahim	butterhead red	1.1	0.000	0.044
09H5701582	Cassini	butterhead	1.1	0.033	0.028
09H5701584	Larsen	crisphead	1.0	0.000	0.000
09H5701269 (harvest 1996)[b]	Pruhonicky cerveny	butterhead	1.4	0.033	0.183
09H5701269 (harvest 1999)[b]	Pruhonicky cerveny	butterhead	1.3	0.233	0.139

N_a – No. of alleles per locus; H_o and H_e – observed and expected heterozygosity; [a]accession No. as stated in the GRIN Czech 1.9.1. Genetic Resources System; available at (https://grinczech.vurv.cz/gringlobal/search.aspx); [b]accessions selected for pivotal screening of polymorphism at the 23 loci

https://doi.org/10.17221/68/2018-CJGPB

were evaluated as multistate categorical variables except for the bolting (descriptor No. 2.1.1.) and flowering (descriptor No. 2.1.2.) that were coded as continuous quantitative variables (number of days after sowing).

Molecular methods. Leaves from three plants per accession were sampled randomly for DNA extraction. Genomic DNA was extracted from fresh leaves using GenElute Plant Genomic DNA Miniprep Kit (Sigma Aldrich, St. Louis, USA), checked on 1.5% agarose gel and quantified using Nanodrop 2000 (ThermoFisher Scientific, Wilmington, USA). PCRs were performed with Kapa Taq PCR Kit (Kapa Biosystems, Wilmington, USA) in 10 µl reactions with 4 ng of template DNA. Fluorescent labelling was performed according to Blacket *et al.* (2012) in a singleplex nested PCR reaction containing three primers: a sequence-specific forward primer with one of four universal tails at its 5' end (used in concentration 0.125 µM), a sequence-specific reverse primer (concentration 0.5 µM), and a fluorescently 5'-modified universal primer (NED™, PET®, VIC™ or FAM™ modification; concentration 0.5 µM). Locus-specific annealing temperature (T_a) and the number of PCR cycles were determined in preliminary experiments for each primer pair. T_a was decreased by 0.5°C in the first six cycles, then 18–23 cycles were run with the optimal T_a (see Table S1 in Electronic Supplementary Material (ESM) and finally, it was lowered to 59°C in the last eight PCR cycles in order to facilitate annealing of the universal primers. All other parameters followed the manufacturer's protocol for the PCR kit. PCR products with different fluorescent labels or those of significantly different lengths were pooled, diluted and analysed together with the GeneScan 600LIZ® size standard on ABI 3730XL capillary sequencer at Macrogen Europe. Twenty three microsatellite loci were selected according to fragment size and annealing temperature (Simko 2009; Rauscher & Simko 2013) and screened for amplification and variability on a representative selection of eight accessions covering the largest possible phenotypic diversity (see Table 1). Ten of SSR loci were determined to be polymorphic and subsequently applied for analyses of all accessions (Table S1 in ESM). These ten marker loci are located on at least five, out of nine lettuce linkage groups.

Data analysis. Descriptive statistics, AMOVA, Mantel test were computed in GenAlEx (Ver. 6.5; Peakall & Smouse 2012), genotype assignment and matrix of pairwise genetic distances in GENOTYPE/GENODIVE software package (Ver. 2.0b23; Meirmans & van Tienderen 2004) and Gower coefficient of similarity for phenotypic data in MVSP software (Ver. 3.22; Kovach 2010). To estimate a power and ability to distinguish between different autogamic lineages, probability of identity sensu Peatkau and Strobeck (1994) was computed in GIMLET (Ver. 1.3.3; Valière 2002). Because the GIMLET algorithm was designed for molecular data, the morphological data were recoded so that every character state was considered an allele (for categorical variables) and each of the two quantitative variables was changed to a categorical variable with two categories based on their bimodal distribution (100 and 140 days being the boundary lines for bolting and flowering, respectively). This matrix was also used in GENOTYPE/GENODIVE.

RESULTS

Overall variability in SSR markers was very low. Out of the 23 SSR screened primer pairs 20 provided specific interpretable products, but only ten loci exhibited polymorphism, as tested on selection of 8 accessions (Table S1 in ESM). Average number of detected alleles (± SE) of 4.8 ± 0.8 and expected heterozygosity of 0.533 ± 0.066 across the ten polymorphic loci and the whole dataset were also rather low. Most of the variation (90% based on AMOVA) was observed among accessions. Within accessions, some degree of expected heterozygosity was detected in 14 accessions of which heterozygotes were observed in 7 accessions (Table 1). One accession (with two seed sets from different harvest years; 09H5701269, Pruhonicky cerveny) was extraordinarily variable in microsatellites and unstable in morphology (plants within this accession exhibited different states of characters, e.g. shape of leaves, leaf colour, blistering, presence/absence of anthocyanin), and was therefore excluded from further analyses both for technical and pragmatic reasons. Except the high phenotypic variability in 09H5701269, no phenotypic variation was observed within other accessions.

Probability of identity, computed from allelic or character state frequencies, was 5.5×10^{-7} for SSR markers, and 3.0×10^{-8} for phenotypic traits. Mean coefficient of similarity among accessions (± SD) was 0.31 ± 0.17 for SSR and 0.55 ± 0.20 for phenotypic traits. Hypothetically, both approaches should therefore be able to distinguish differences between individuals (or autogamic lineages) sufficiently. Nevertheless, neither phenotypic (when

https://doi.org/10.17221/68/2018-CJGPB

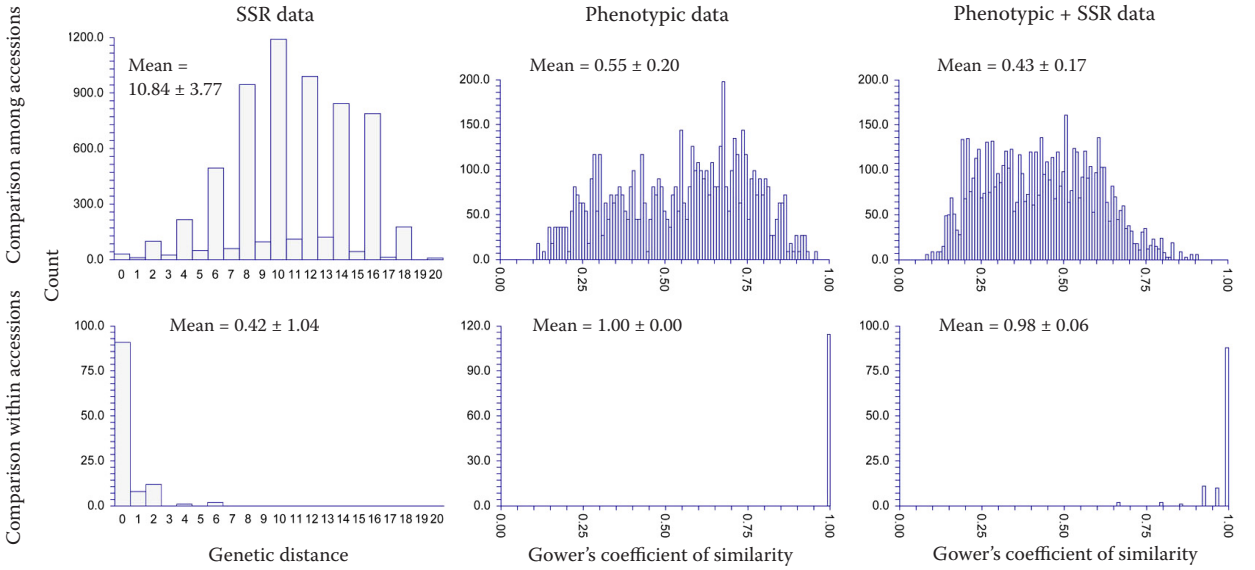

Figure 1. Histograms of pairwise genetic distances or Gower's coefficient of similarity among individual lettuce plants based on SSR, phenotypic and mixed data; arithmetic mean ± standard deviation shown above histograms; note that the discontinuous distribution for SSR data is shaped by diploidy and high observed homozygosity

coded as categorical variables) nor SSR data alone could distinguish among all different accessions, as seen from genotype assignment even at the threshold of zero (i.e. no variation within accessions allowed; Table S2 in ESM). Also histograms of pairwise genetic distances or phenotypic similarities did not reveal any apparent threshold delimiting intra- and inter-accession variability (Figure 1, Figure S1 in ESM). By combining both data types, the mean coefficient of similarity among accessions (± SD) shifted to 0.43 ± 0.17 implying that ca 95% of the comparisons lay in the interval 0.09–0.77, while most intra-accession comparisons had coefficient of similarity above 0.86 (Figure 1).

Mantel test comparing matrices of genetic distances and Gower's coefficient of similarity in phenotypic characters indicated significant correlation ($P = 0.01$) among SSR and phenotypic data both on the level of individual plants (Figure 2) and on the level of accessions (Figure S1 in ESM).

DISCUSSION

Combination of phenotypic descriptors and microsatellites distinguishes most accessions. Duplications in gene bank collections represent a significant burden for curators because of higher demands on budget, extensive storage/regeneration

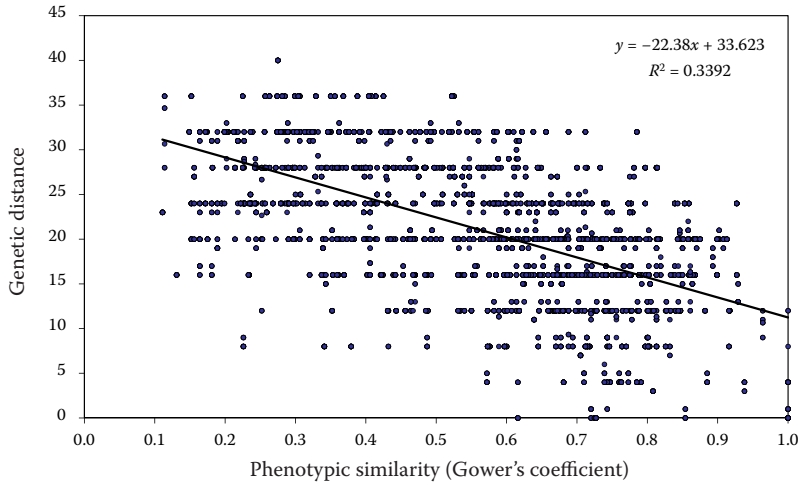

Figure 2. Mantel test for correlation between genetic distance and morphological similarity matrices on the level of individual plants
$P = 0.01$

https://doi.org/10.17221/68/2018-CJGPB

capacities, as well as labour. Unlike the previous years, when the priority was expanding the collections to include new materials from collection expeditions and obsolete varieties for breeding needs, gene banks are currently targeting on removing redundant accessions from their collections in order to increase efficiency of collections management. The highest priority for the Vegetables Network of European Cooperative Programme for Plant Genetic Resources (ECPGR) is sharing responsibilities for the *ex-situ* conservation of European vegetable crops genetic resources within A European Genebank Integrated System (AEGIS) for plant genetic resources for food and agriculture (Daunay *et al.* 2011). The challenge is currently to identify genetically unique and important accessions among potential duplicate materials. The Most Appropriate Accessions (MAA) for each crop will subsequently be included into the decentrally managed European Collection. In the case of lettuce, comprehensive approach includes: (1) obtaining basic missing passport data on accessions, (2) morphological comparison of accessions within and among collections, and (3) use of biochemical and molecular techniques for exact genetic distinction of genetic resources (Lebeda *et al.* 2007).

In order to distinguish samples from distinct accessions and/or detect the level of intra-accession diversity in germplasm collections, different molecular methods and classes of DNA markers have been used (Barcaccia 2009). Among them, simple sequence repeats (SSRs) have been widely applied in many crops for the purpose of identification and validation of duplicate accessions (*e.g.* del Rio *et al.* 2006; Irish *et al.* 2010; van Treuren *et al.* 2010). SSR are generally known as highly variable markers. This fact stems from the high mutation rate caused by multiple repetition of simple short motives that are prone to replication errors, and from their independency on phenotype (with some exceptions) and thus selective neutrality of the sequences in many cases (Selkoe & Toonen 2006). The probability of detection of variation among genetically different individuals is therefore considerable for the SSR markers. Another commonly used marker system is amplified fragment length polymorphism (AFLP), which usually exhibit comparable variation to SSR (Garcia *et al.* 2004; Majeský *et al.* 2015). A great advantage of microsatellites over AFLP or other commonly used dominant markers is their co-dominant nature, i.e. the ability to detect heterozygosity at individual loci. At the same time, SSR method is relatively low-cost, demands only little optimization and is also highly reproducible with the possibility to compare data from different experiments or laboratories (Jones *et al.* 1997). Moreover, with knowledge of primer sequences for the particular plant group it can be performed in basic molecular laboratory, as compared e.g. to modern methods based on microarrays, SNPlex system or next-generation sequencing. Such approaches also exhibit many advantages, e.g. ability to provide rapid high density genome scans, robustness of data and cost-effectiveness per data point when genotyping large numbers of loci and samples. Nevertheless, the overall cost is still quite high, which makes them inaccessible for many users (Rasheed *et al.* 2017). Moreover, SSR's perform comparably or even better than SNP's and microarrays in many aspects of genotyping of crops (Simko *et al.* 2012). The drawback of SSR markers can be their availability. Some of these markers have been published, successfully applied in lettuce research, and are publicly accessible (Simko 2009; Hong *et al.* 2013, 2015; Rauscher & Simko 2013; Wang *et al*. 2017). On the other hand, Van Treuren *et al.* (2010) used in their extensive study SSR markers with primers whose sequences are not freely available. This complicates their use in practice and e.g. development of a panel of SSR markers that should be used for duplicate differentiation of lettuce accessions in germplasm collections. Thus, SSR markers used in this study might be adopted for such purpose.

Considering the above mentioned, we studied patterns of variability in the Czech national collection of lettuce with the aim of developing a method of duplicate identification that would be relatively inexpensive, fast, reproducible and as accurate as possible. The use of many independent methods is the most reliable approach (e.g., Donelli *et al.* 2013; van Treuren *et al.* 2010), but on the other hand, every additional method increase the cost and time needed for the analysis. We therefore focused on two methods, phenotypic evaluation and SSR genotyping, and studied their ability to distinguish between individual accessions. Variability in SSR markers was very low as ten of twenty scorable SSR loci exhibited no variation and the remaining ten provided 48 alleles in total. This number is, nevertheless, comparable to that reported by van Treuren *et al.* (2010) who detected 150 alleles at ten SSR loci in the complete CGN collection of cultivated lettuce (1540 accessions) and crop wild relatives (1031 accessions). Despite high potential ability to distinguish among

https://doi.org/10.17221/68/2018-CJGPB

accessions (probability of identity 5.5×10^{-7}), SSR data alone cannot distinguish every single accession as some accessions share identical SSR genotype (Table S2 in ESM). Moreover, due to the SSR variability within accessions, high mutation rate, and scoring errors, a threshold between intra- and inter-accession variation needs to be determined. If not determined accurately, it can affect the frequency of false positive or false negative identifications of duplicates (van Treuren *et al.* 2001, 2010; Lund *et al.* 2003). Based on inclusion of two individuals per accession, maximum of a single SSR difference was regarded an acceptable level for inter-accession variation by van Treuren *et al.* (2010). Both their and our data reveal that intra- and inter-accession variation components overlap (the overlap in our data is approximately 2% of all pair-wise comparisons, given the data have normal distribution; Figure 1, Figure S1 in ESM), thus setting a reliable threshold is problematic. False positive identification of duplicate accessions (when the threshold is too high) would lead to highly undesirable elimination of closely related material, while false negative identifications (when threshold is too low) would lead to preservation of duplicate accessions in the gene bank (van Treuren & van Hintum 2003). For this reason (and despite the fact that two or more differences were detected within nine accessions; data not shown), the rather low threshold of one allele difference seems to be appropriate. Alternatively, because the two-allele difference was always caused by a single homozygote locus, the threshold may be set to the difference of one locus, rather than one allele.

Compared to SSR's, which could not differentiate among accessions in 5 cases and revealed within-acccession genotypic variability in 12 cases (Table S2 in ESM), standard morphological and phenological descriptors performed better in our analysis as no two accessions shared the exactly same phenotype (the highest coefficient of similarity was 0.964, although one pair of accessions – 09H5700014 and 09H57000718, differed only in phenology and this difference was erased during data transformation to categorical variables) and no accession exhibited variability, except the highly unstable 09H5701269. One reason that this variety is not uniform can be the fact, that it is an old landrace, that dates from the 1950s, the second is that it could be a mixture of two or more accessions. Alternatively, the acccession was obtained from foreign gene bank, and the possibility that it is a mixture of two accessions cannot be excluded. Nevertheless, the better performance of phenotypic data could have been caused by limited sampling of accessions and similar but non-identical accessions would probably remain undistinguished in wider data set. Another drawback of using phenotypic traits is the variability due to different growing environments and subjectivity of human evaluators. Comparison among different seasons, regions or cultivation conditions may therefore be difficult, possibly resulting in false duplicate identifications. Although neither phenotypic nor SSR data alone can serve as evidence for unambiguous duplicate confirmation, their combination increases the resolution power of the method considerably, as seen on decrease of mean similarity coefficient from 0.55 (phenotypic data) to 0.43 (combined data) and thus smaller overlap of intra- and inter-accession variation (Figure 1). The use of two independent lines of evidence was formerly recommended for confirmation of potential duplicate accessions defined according to the passport data and may be considered sufficient (van Treuren *et al.* 2010).

Structuring of genetic variation mostly corresponds to autogamy and phenotypes. The use of molecular markers for duplicates identification can often bring valuable results that are not in the primary focus of the researchers, as has been demonstrated in several works (van Treuren & van Hintum 2003 and references therein). In this study, we focused on characterization of intra-accession variation across selected morphotypes of lettuce, which is generally expected to be low in autogamous plants. In cultivated lettuce, only around 1–5% of seeds are derived from cross-pollination under field conditions (Thomson *et al.* 1958). Rejuvenation from limited number of individuals under isolation further decreases probability of long-term survival of more than one autogamous (and increasingly homozygous) line per accession in gene banks (van Treuren & van Hintum 2001). Detection of any intra-accession variability should therefore be very rare. Contrary to this expectation, non-zero level of expected heterozygosity was detected in 14 (36%) of the studied accessions and heterozygous individuals were observed in 7 (18%) of them, always at a single SSR locus (Table 1). Comparable or even higher variation was detected also in the CGN cultivated and wild lettuce collection (more than 20% of accessions were variable in SSR and more than 30% in AFLP; van Treuren *et al.* 2010), in strictly self-fertilizing barley (van Treuren & van Hintum

https://doi.org/10.17221/68/2018-CJGPB

2001) or flax (van Treuren & van Hintum 2003; Smýkal *et al.* 2011). Our study therefore confirms previously reported intra-accession variability in (usually neutral) molecular markers in many lettuce accessions, although variation in phenotypic (phenological and qualitative morphological) traits was not detected. This contradiction, together with the patterns of observed variability, may imply that the intra-accession diversification is driven by ongoing (mostly neutral) mutations rather than rare cross-pollination events, mistakes in rejuvenation (insufficient isolation, seeds contamination or mishandling, etc.) or residual variation.

SSR markers are usually difficult for inferences of evolutionary relationships due to high mutation rate and high degree of homoplasy (Šarhanová *et al.* 2017). On the other hand, morphological traits are not free of homoplasy either, particularly in crops where strong artificial selection may lead to convergence of traits and thus blur or even alter the phylogenetic/genealogical signal (Bortiri *et al.* 2006; Gaut 2015; Washburn *et al.* 2016). As a result, data from molecular and morphological markers may be only weakly correlated. Triangular relationship (named after the typical structure of the correlation plot) was repeatedly observed in maize, suggesting that the weakly and moderately related genotypes, based on molecular markers assessment, may exhibit similar or significantly different phenotypes (Rebourg *et al.* 2001; Babic *et al.* 2016). Our data indicate weak (R^2 = 0.34), but significant (P = 0.01) correlation between both of the data types (Figure 2) which implies that evolution of phenotypic traits and DNA markers is concerted in cultivated lettuce, or in other words, homoplasy appears to be low in both SSR and phenotypic traits. This correlation, although not explicitly studied, was observed in lettuce also by van Treuren *et al.* (2010) who noted increasing similarity detected by DNA markers (both SSR and AFLP) with decreasing organizational level (i.e. from crop type through cultivar type to accession).

Implementation of results. The Czech national collection of lettuce consists of 844 accessions, of which 55% are potential duplicates. In the near future, we plan to test accessions in individual groups of potential duplicates as formulated according to their passport data. We will use the phenotypic data (including quantitative morphological) that are available for some of the accessions from field trials from previous years along with microsatellite fingerprinting of individual accessions. Combination of phenotyping and SSR genotyping would allow us to confirm duplicate accessions and subsequently to identify MAA for inclusion to AEGIS. Using this approach and considering the autogamous breeding system of lettuce, we will decide which accession should be selected as authentic and which can be excluded from the basic and active collections. The redundant accessions will remain a part of the working collection and will not be included in the regeneration process. Using of this approach would allow for more efficient management of The Czech national lettuce collection.

Acknowledgements. We thank Prof. A. Lebeda (Palacký University) for valuable comments when selecting lettuce varieties, and our technical crew for assistance with growing plant material.

References

Babic V., Nikolic A., Andjelkovic V., Kovacevic D., Filipovic M., Vasic V., Mladenovic-Drinic S. (2016): UPOV morphological versus molecular markers for maize inbred lines variability determination. Chilean Journal of Agricultural Research, 76: 417–426.

Barcaccia G. (2009): Molecular markers for characterizing and conserving crop plant germplasm. In: Jain S.M., Brar D.S. (eds.): Molecular Techniques in Crop Improvement. 2nd Ed. Dordrecht, Springer Netherlands: 231–254.

Blacket M.J., Robin C., Good R.T., Lee S.F., Miller A.D. (2012): Universal primers for fluorescent labelling of PCR fragments – an efficient and cost-effective approach to genotyping by fluorescence. Molecular Ecology Resources, 12: 456–463.

Bortiri E., Vanden Heuvel B., Potter D. (2006): Phylogenetic analysis of morphology in *Prunus* reveals extensive homoplasy. Plant Systematics and Evolution, 259: 53–71.

Daunay M.C., Allender C., Astley D., Bas N., van Dooijeweert W., van Treuren R., Branca F., Díez M.J., Geoffriau E., Keller E.R.J., Kotlińska T., Smékalová K., Engels J., Maggioni L. (2011): More efficient conservation and use of vegetable genetic resources in Europe: ECPGR achievements and perspectives. Acta Horticulturae, 918: 405–417.

del Rio A.H., Bamberg J.B., Huaman Z. (2006): Genetic equivalence of putative duplicate germplasm collections held at CIP and US potato genebanks. American Journal of Potato Research 83: 279–285.

Doležalová I. (2014): Genetic resources of leafy vegetables with respect to the genus *Lactuca* L. In: Petrželová I., Doležalová I., Sapáková E., Horák M. (eds.): Genetic Re-

https://doi.org/10.17221/68/2018-CJGPB

sources of Vegetable and Special Crops. Folia Universitas Agriculturae et Silviculturae Mendelianae Brunensis, 7: 32–45.

Donelli G., Vuotto C., Mastromarino P. (2013): Phenotyping and genotyping are both essential to identify and classify a probiotic microorganism. Microbial Ecology in Health and Disease, 24: 20105.

FAO (2010): The state of *ex situ* conservation. Chapter 3. In: The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome, FAO: 54–90.

Frankel R., Galun E. (1977): Pollination Mechanisms, Reproduction and Plant Breeding. Berlin, Heidelberg, New York, Springer-Verlag.

Garcia A.A.F., Benchimol L.L., Barbosa A.M.M., Geraldi I.O., Souza C.L., de Souza A.P. (2004): Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. Genetics and Molecular Biology, 27: 579–588.

Gaut B.S. (2015): Evolution is an experiment: Assessing parallelism in crop domestication and experimental evolution. Molecular Biology and Evolution, 32: 1661–1671.

George R.A.T. (1999): Vegetable Seed Production. 2nd Ed. Wallingford, CABI Publishing.

Hong J.H., Kwon Y.S., Choi K.J., Mishra R.K., Kim D.H. (2013): Identification of lettuce germplasms and commercial cultivars using SSR markers developed from EST. Korean Journal of Horticultural Science and Technology, 31: 772–781.

Hong J.H., Kwon Y.S., Mishra R.K., Kim D.H. (2015): Construction of EST-SSR databases for effective cultivar identification and their applicability to complement for lettuce (*Lactuca sativa* L.) distinctness test. American Journal of Plant Sciences, 6: 113–125.

Irish B.M., Goenaga R., Zhang D., Schnell R., Brown J.S., Motamayor J.C. (2010): Microsatellite fingerprinting of the USDA-ARS tropical agriculture research station cacao (*Theobroma cacao* L.) germplasm collection. Crop Science, 50: 656–667.

Jones C.J., Edwards K.J., Castaglione S., Winfield M.O., Sala F., van de Wiel C., Bredemeijer G., Vosman B., Matthes M., Daly A., Brettschneider R., Bettini P., Buiatti M., Maestri E., Malcevschi A., Marmiroli N., Aert R., Volckaert G., Rueda J., Linacero R., Vazquez A., Karp A. (1997): Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular Breeding, 3: 381–390.

Kesseli R.V., Ochoa O., Michelmore R.W. (1991): Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). Genome, 34: 430–436.

Kovach W.L. (2010): MVSP – A MultiVariate Statistical Package for Windows, Ver. 3.2. Pentraeth, Wales, Kovach Computing Services.

Lebeda A., Ryder E.J., Grube R., Doležalová I., Křistková E. (2007): Lettuce (Asteraceae; *Lactuca* spp.). In: Singh R.J. (ed.): Genetic Resources, Chromosome Engineering, and Crop Improvement. Vol. 3. Vegetable Crops. Boca Raton, CRC Press, Taylor and Francis Group: 377–472.

Lipman E., Jongen M.W.M, van Hintum T.J.L., Gass T., Maggioni L. (1997): Central Crop Databases: Tools for Plant Genetic Resources Management. Rome, Wageningen, International Plant Genetic Resources Institute.

Lund B., Ortiz R., Skovgaard I.M., Waugh R., Andersen S.B. (2003): Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. Theoretical and Applied Genetics, 106: 1129–1138.

Majeský Ľ., Vašut R.J., Kitner M. (2015): Genotypic diversity of apomictic microspecies of the *Taraxacum scanicum* group (*Taraxacum* sect. *Erythrosperma*). Plant Systematics and Evolution, 301: 2105–2124.

Meirmans P.G., van Tienderen P.H. (2004): GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes, 4: 792–794.

Michelmore R.W. (2014): Genetic variation in lettuce. In: California Leafy Greens Research Program, Davis, The Genome Center and The Department of Plant Sciences, University of California, Apr 1, 2013–Mar 31, 2014.

Paetkau D., Strobeck C. (1994): Microsatellite analysis of genetic variation in black bear populations. Molecular Ecology, 3: 489–495.

Peakall R., Smouse P.E. (2012): GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics, 28: 2537–2539.

Rao N.K., Hanson J., Dulloo M.E., Ghosh K., Nowell D., Larinde M. (2006): Manual of Seed Handling in Genebanks. Handbooks for Genebanks No. 8. Rome, Bioversity International.

Rasheed A., Hao Y., Xia X., Khan A., Xu Y., Varshney R.K., He Z. (2017): Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. Molecular Plant, 10: 1047–1064.

Rauscher G., Simko I. (2013): Development of genomic SSR markers for fingerprinting lettuce (*Lactuca sativa* L.) cultivars and mapping genes. BMC Plant Biology, 13: 1–11.

Rebourg C., Gouesnard B., Charcosset A. (2001): Large scale molecular analysis of traditional European maize populations. Relationships with morphological variation. Heredity, 86: 574–587.

Richerzhagen C. (2010): Protecting Biological Diversity. The Effectiveness of Access and Benefit-sharing Regimes. New York, Abingdon, Taylor & Francis.

Šarhanová P., Sharbel T.F., Sochor M., Vašut R.J., Dančák M., Trávníček B. (2017): Hybridization drives evolution of

https://doi.org/10.17221/68/2018-CJGPB

apomicts in *Rubus* subgenus *Rubus*: evidence from microsatellite markers. Annals of Botany, 120: 317–328.

Selkoe K.A., Toonen R.J. (2006): Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters, 9: 615–629.

Simko I. (2009): Development of EST-SSR markers for the study of population structure in lettuce *(Lactuca sativa* L.). Journal of Heredity 100: 256–262.

Simko I., Eujayl I., van Hintum T.J.L (2012): Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. Plant Science, 184: 54–62.

Smýkal P., Bačová-Kerteszová N., Kalendar R., Corander J., Schulman A.H., Pavelek M. (2011): Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers. Theoretical and Applied Genetics, 122: 1385–1397.

Spooner D., van Treuren R., de Vicente M.C. (2005): Molecular Markers for Genebank Management. IPGRI Technical No. 10. Rome, International Plant Genetic Resources Institute.

Thomson R.C., Whitaker T.W., Bohn G.W., van Horn C.W. (1958): Natural cross-pollination in lettuce. American Society for Horticultural Science, 72: 403–409.

Valière N. (2002): GIMLET: a computer program for analysing genetic individual identification data. Molecular Ecology Notes, 2: 377–379.

van Hintum T.J.L., Knüpffer H. (1995): Duplication within and between germplasm collections: I. Identifying duplication on the basis of passport data. Genetic Resources and Crop Evolution, 42: 127–133.

van Hintum T.J.L., Boukema I.W. (1999): Genetic resources of leafy vegetables, In: Lebeda A., Křístková E. (eds.): Eucarpia Leafy Vegetables '99, Proc. Eucarpia Meeting on Leafy Vegetables Genetics and Breeding, Olomouc, June 8–10, 1999: 59–72.

van Treuren R., van Hintum T.J.L. (2001): Identification of intra-accession genetic diversity in selfing crops using AFLP markers: Implications for collection management. Genetic Resources and Crop Evolution, 48: 287–295.

van Treuren R., van Hintum T. J. L. (2003): Marker-assisted reduction of redundancy in germplasm collections: Genetic and economic aspects. Acta Horticulturae, 623: 139–149.

van Treuren R., van Soest L.J.M., van Hintum Th.J.L. (2001): Marker-assisted rationalisation of genetic resources collections: a case study in flax using AFLPs. Theoretical and Applied Genetics, 103: 144–152.

van Treuren R., de Groot E.C., Boukema I.W., van de Wiel C.C.M., van Hintum T.J.L. (2010): Marker-assisted reduction of redundancy in a genebank collection of cultivated lettuce. Plant Genetic Resources: Characterization and Utilization, 8: 95–105.

Wang S., Wang B., Liu J., Ren J., Huang X., Zhou G., Wang A. (2017): Novel polymorphic EST-based microsatellite markers characterized in lettuce (*Lactuca sativa*). Biologia, 72: 1300–1305.

Washburn J.D., Bird K.A., Conant G.C., Pires J.C. (2016): Convergent evolution and the origin of complex phenotypes in the age of systems biology. International Journal of Plant Sciences, 177: 305–318.

Waycott W., Fort S.B., Ryder E.J., Michelmore R.W. (1999): Mapping morphological genes relative to molecular markers in lettuce (*Lactuca sativa* L.). Heredity, 82: 245–251.

Received for publication May 9, 2018
Accepted after corrections September 19, 2018
Published online January 6, 2019