RESEARCH

# Molecular genetic diversity of the selenium hyperaccumulator *Stanleya pinnata*

Aysu Başak KÖK, Sami DOĞANLAR, Anne FRARY*

Department of Molecular Biology and Genetics, Science Faculty, Izmir Institute of Technology, Izmir, 35430, Turkey

*Corresponding author e-mail: annefrary@iyte.edu.tr

## ABSTRACT

*Stanleya pinnata* is a perennial herb that can hyperaccumulate selenium. In the present study, genetic diversity of 15 accessions of *S. pinnata* originating from different locations in the USA was determined using sequence amplified polymorphism (SRAP) markers. A total of 506 alleles were identified among the 15 accessions and 98% of these alleles were polymorphic. Dice dissimilarity coefficient values between accessions varied from 0.14 to 0.87 with an average of 0.53 indicating that the plant material was highly diverse as expected given its self-incompatibility. A dendrogram was constructed to understand the genetic relationships among the accessions based on the molecular marker data. The dendrogram grouped the 15 *S. pinnata* accessions into three clusters while population structure analysis divided these accessions into five groups. Clustering was correlated with accession origin.

KEYWORDS: Brassicaceae; sequence related amplified polymorphism; variation; selenium

Citation: KÖK AB et al. (2018). Molecular genetic diversity of the selenium hyperaccumulator *Stanleya pinnata.* Polymorphism. 1; 1: 1-9.

## INTRODUCTION

*Stanleya pinnata* (*S. pinnata*) belongs to the genus Stanleya (Brassicaceae). It is a perennial, dicot herb (Feist and Parker, 2001). Its native habitat is dry, stony slopes and washes. *S. pinnata* is widely distributed in the USA and Mexico (USDA, 2017). It is a nonmodel organism and a self-incompatible selenium (Se) hyperaccumulator (Quinn et al., 2011). *S. pinnata* is known to hyperaccumulate toxic selenium to over 1000 µg Se g-1 dry weight (DW) (Cappa et al., 2014; Schiavon and Pilon-Smits, 2017). Selenium is a Group 6A element and is a metalloid. Selenium toxicity is a common problem in many countries including parts of the USA, China, Canada and India (Schiavon and Pilon-Smits, 2017). Plants and animals can accumulate selenium in greater concentrations than are present in their environment and bioaccumulation causes toxicity (Fordyce, 2005). Like excess selenium, selenium deficiency also poses problems to organisms. It is mostly seen in China, northwestern Europe, Australia, New Zealand, sub-Saharan Africa, southern Brazil and parts of the USA (Wu et al., 2015). The selenium hyperaccumulation and metabolism capacity of *S. pinnata* can be exploited to remediate both selenium toxicity and deficiency. For example, phytoremediation using *S. pinnata* is one way to clean up seleniferous polluted soil and wastewater. This process not only sequesters selenium from polluted areas but also provides other advantages. Growth of *S. pinnata* (or any other plant) increases the quality of the soil by producing organic compounds. Moreover, such plants protect soil from erosion and metal leaching (Chaudhry et al., 1998). Last but not least, the selenium extracted by the plant can be harvested and used for other purposes (Ali et al., 2013) such as amending deficient soils (Bañuelos et al., 2015).

Genetic diversity can be defined as morphological, gene and sequence level variability within and between populations. Such diversity is important as it affects the individual's and population's ability to adapt to the environment under changing circumstances. Genetic diversity studies are especially relevant because of the threat of global climate change and the observation that the allelic diversity of many species has decreased due to inbreeding and genetic drift. Examination of genetic diversity also helps us to select individuals or populations from nature that are suitable for studying a given scientific question (Fordyce, 2005).

Molecular marker analysis is one way to determine genetic diversity and is a useful method for phylogenetic studies and cultivar identification because it utilizes changes at the DNA sequence level. Sequence related amplified polymorphism (SRAP) markers are multi-allelic and multi-locus molecular markers that are powerful tools for genetic diversity analysis. SRAP markers are designed to amplify open reading frames (ORFs) by using a combination of different forward and reverse primers (Li and Quiros, 2001). The forward and reverse primers are composed of 17 and 18 bases, respectively. They have two parts: a 15 nucleotides core sequence and three selective bases. The core sequence has 10 or 11 nucleotides. This is followed by the sequence CCGG in the forward primer and AATT in the reverse primer. The three selective bases are present at the 3' ends and they determine the differences in each primer's specificity. Polymorphism in the DNA fragments amplified by SRAP primers is caused by variation in the length of introns, promoters, and spacers (Li and Quiros, 2001; Sun et al., 2006).

The objective of this research was to investigate the genetic variability of 15 *S. pinnata* accessions from various locations in the USA. This study is the first to examine diversity of *S. pinnata* using molecular markers.

## MATERIALS AND METHODS

### Plant Material

*Stanleya pinnata* seeds were provided by the United States Department of Agriculture

(USDA, Beltsville MD, USA). They were collected from different locations in the USA (Table 1). A total of 15 genotypes were grown in potting soil until their leaves reached approximately 5 cm in length.

## DNA Extraction

A modified version of the CTAB method was used for DNA extraction from fresh leaf tissue (Doyle and Doyle 1987). Extracted DNA samples were re-suspended in sterile distilled water. Their concentrations were measured by a NanodropND-1000 spectrophotometer and final concentrations were adjusted to 50 ng/µl. DNA samples were stored at -20 °C.

## SRAP Analysis

Combinations of nine forward (ME) and four reverse (EM) SRAP primers were used for genetic diversity analysis (Table 2). A total of 29 combinations were tested using polymerase chain reaction (PCR) (Table 3). PCR components included 10X Tango Buffer with BSA, 0.01 mM forward and reverse primers, 25 mM MgCl2, 10 mM dNTP, 1 U Taq Polymerase, and 50 ng/µl DNA (final volume of the mixture was 25 µl in sterile dH2O). Amplification reactions were performed in BioRad Thermal Cycler with an initial step at 94 ºC for 5 min, 94 ºC for 1 min, 35 °C for 1 min and 72 ºC for 1 min for 5 cycles. The following 35 cycles continued with 94 °C for 1 min, 50 ºC for 1 min and 72 °C for 1 min. The PCR products were separated on 3% agarose gel and visualized with ethidium bromide staining and a BioRad Gel Doc XR imaging system.

## Genetic Diversity Analysis

DNA fragments were scored dominantly: presence (1), absence (0) and missing data (9). DARwin6 (Dissimilarity Analysis and Representation for Windows) software was used to process the scoring matrix (Perrier and Jacquemoud-Collet, 2006). First, genetic dissimilarities were calculated using the Dice algorithm. A dendrogram was then constructed using the unweighted neighbor joining algorithm and goodness of fit of the tree to the dissimilarity matrix was determined by the fit criterion (Mantel test).

Table 1. *Stanleya pinnata* accessions and origins used in the study.

| ID number | Accession | Species Name | Origin (Latitude, Longitude) | State |
|---|---|---|---|---|
| 1 | W6 27186 | *Stanleya pinnata* | USA (37.23806- 111.95639) | Utah |
| 2 | W6 32751 | *Stanleya pinnata* | USA (40.1648- 110.43955) | Utah |
| 3 | W6 37559 | *Stanleya pinnata* | USA (42.77902222- 115.8139556) | Idaho |
| 4 | W6 39198 | *Stanleya pinnata* | USA (42.90571111- 116.2867472) | Idaho |
| 5 | W6 39199 | *Stanleya pinnata* | USA (42.975 - 116.3401889) | Idaho |
| 6 | W6 39200 | *Stanleya pinnata* | USA (40.35311111- 109.4090556) | Utah |
| 7 | W6 39201 | *Stanleya pinnata* | USA (42.80597222- 117.7305833) | Oregon |
| 8 | W6 40612 | *Stanleya pinnata* | USA (42.80660833- 115.8988389) | Idaho |
| 9 | W6 41398 | *Stanleya pinnata* var. integrifolia | USA (40.87430556- 114.2206667) | Nevada |
| 10 | W6 42096 | *Stanleya pinnata* | USA (39.486638- 111.996444) | Utah |
| 11 | W6 42097 | *Stanleya pinnata* | USA (40.950055- 117.433777) | Nevada |
| 12 | W6 42677 | *Stanleya pinnata* | USA (39.65122- 107.06605) | Colorado |
| 13 | W6 42678 | *Stanleya pinnata* | USA (36.7933 -108.1892) | New Mexico |
| 14 | W6 48644 | *Stanleya pinnata* | USA (35.87318- 115.45582) | Nevada |
| 15 | W6 48645 | *Stanleya pinnata* | USA (36.44758- 116.43572) | Nevada |

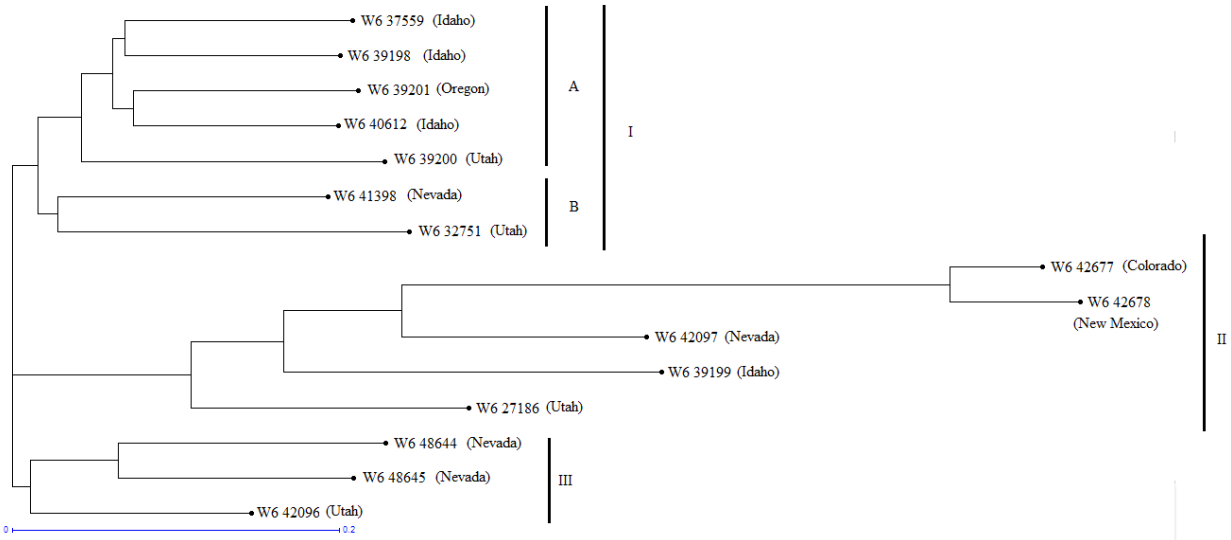


Figure 1. Unweighted neighbor joining tree representing cluster analysis of 15 *S. pinnata* accessions using the Dice coefficient. The tree consists of three major clusters: cluster I with subclusters A and B, cluster II and cluster III.

In addition, average, maximum and minimum values of gene diversity (GD) for each marker and marker alleles were calculated using Gene Diversity Software (GDdom) (Abuzayed et al., 2017). Population structure was analyzed with the Structure program. Analysis was run with burn in period of 50,000 and 300,000 MCMC replications. Subpopulation numbers (K) from 1 to 10 were tested with 10 iterations for each value of K (Pritchard et al., 2000). Structure output was visualized using Structure Harvester (Earl and VonHoldt, 2012).

## RESULTS

Combinations of nine forward (ME) and four reverse (EM) SRAP primers were used for genetic diversity analysis. Thus, a total of 29 combinations were tested yielding 506 fragments, 496 (98%) of which were polymorphic. The SRAP marker combinations gave an average of 17.4 alleles per locus and the number of alleles per marker ranged from six to 27. The percentage of polymorphic bands (PPB) for different primer combinations ranged from 81.3 to 100% with an average of 97.9%. A total of 24 primer combinations out of 29 had 100% polymorphism. The maximum gene diversity was 0.38 and the minimum gene diversity was 0.23 using primer combinations EM3 - ME6 and EM3 - ME7, respectively (Table 2). The average gene diversity of the SRAP markers was 0.30.

Dice dissimilarity values were calculated between all pairs of accessions. They varied from 0.14 to 0.87 with an average of 0.56 (data not shown). According to these values, the lowest genetic distance was between accessions W642677 (from Colorado) and W642678 (New Mexico) and the highest genetic distance was between W639201 (Oregon) and W642678 (New Mexico).

A dendrogram was drawn to illustrate the genetic relationships among the *S. pinnata* accessions. The fit criterion for the tree (r = 0.99) suggested a strong correlation between the dissimilarity matrix and the dendrogram. The dendrogram grouped the 15 *S. pinnata* accessions into three main clusters. Cluster I was the largest cluster with seven genotypes; cluster II had five genotypes and cluster III was the smallest with three genotypes. Cluster I had two subclusters: A and B (Figure 1).

Table 2. Polymorphism of SRAP markers in *S. pinnata*.

| Primer combination | Average GD valuea | TNFb | NPBc | PPBd (%) |
|---|---|---|---|---|
| EM1 - ME2 | 0.26 | 6 | 5 | 83.3 |
| EM1 - ME3 | 0.29 | 14 | 14 | 100 |
| EM1 - ME4 | 0.33 | 8 | 8 | 100 |
| EM1 - ME5 | 0.32 | 16 | 13 | 81.3 |
| EM1 - ME7 | 0.36 | 11 | 11 | 100 |
| EM1 - ME9 | 0.28 | 21 | 21 | 100 |
| EM1 - ME10 | 0.30 | 19 | 19 | 100 |
| EM2 - ME1 | 0.32 | 12 | 12 | 100 |
| EM2 - ME2 | 0.29 | 17 | 17 | 100 |
| EM2 - ME3 | 0.27 | 8 | 8 | 100 |
| EM2 - ME4 | 0.32 | 15 | 15 | 100 |
| EM2 - ME5 | 0.25 | 22 | 21 | 95.5 |
| EM2 - ME6 | 0.28 | 25 | 24 | 96 |
| EM2 - ME7 | 0.30 | 5 | 5 | 100 |
| EM2 - ME9 | 0.36 | 18 | 18 | 100 |
| EM2 - ME10 | 0.25 | 12 | 11 | 91.7 |
| EM3 - ME1 | 0.31 | 17 | 17 | 100 |
| EM3 - ME2 | 0.37 | 15 | 15 | 100 |
| EM3 - ME3 | 0.33 | 15 | 15 | 100 |
| EM3 - ME4 | 0.29 | 19 | 19 | 100 |
| EM3 - ME5 | 0.33 | 15 | 15 | 100 |
| EM3 - ME6 | 0.23 | 24 | 24 | 100 |
| EM3 - ME7 | 0.38 | 14 | 14 | 100 |
| EM3 - ME9 | 0.35 | 15 | 15 | 100 |
| EM3 - ME10 | 0.26 | 18 | 17 | 100 |
| EM4 - ME1 | 0.33 | 16 | 16 | 100 |
| EM4 - ME2 | 0.30 | 27 | 27 | 100 |
| EM4 - ME3 | 0.35 | 20 | 20 | 100 |
| EM4 - ME4 | 0.35 | 18 | 18 | 100 |
| EM4 - ME5 | 0.26 | 17 | 15 | 88.2 |
| EM4 - ME6 | 0.30 | 27 | 27 | 100 |
| Mean | 0.30 | 16.3 | 16 | 97.9 |
| Max | 0.38 | 27 | 27 | 100 |
| Min | 0.23 | 5 | 5 | 81.3 |
| Total | | 506 | 496 | |

a GD genetic diversity, b TNF total number of fragments, c NPB number of polymorphic bands, d PPB percentage of polymorphic bands

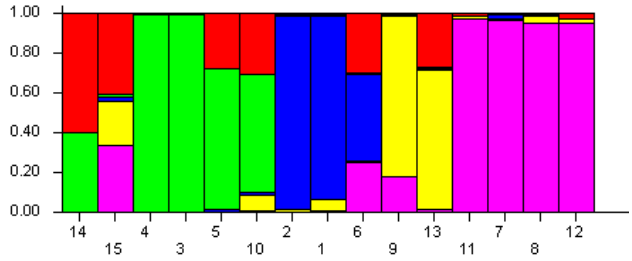


Figure 2. The bar plot for K = 5. Red, green, blue, yellow and magenta colors represent subpopulations 1, 2, 3, 4 and 5. X-axis displays the identity numbers of each accession and the y-axis shows the probability of membership in the corresponding subpopulation.

Table 3. Primer sequences used in SRAP analysis.

| Name | Sequence ( 5' - 3') |
|---|---|
| ME1 | 5'TGAGTCCAAACCGGATA3' |
| ME2 | 5'TGAGTCCAAACCGGAGC3' |
| ME3 | 5'TGAGTCCAAACCGGAAT3' |
| ME4 | 5'TGAGTCCAAACCGGACC3' |
| ME5 | 5'TGAGTCCAAACCGGAAG3' |
| ME6 | 5'TGAGTCCAAACCGGTAG3' |
| ME7 | 5'TGAGTCCAAACCGGTTG3' |
| ME9 | 5'GACTGCGTACGAATTAGG3' |
| ME10 | 5'GACTGCGTACGAATTTAG3' |
| EM1 | 5'GACTGCGTACGAATTAAT3' |
| EM2 | 5'GACTGCGTACGAATTGAC3' |
| EM3 | 5'GACTGCGTACGAATTGAC3' |
| EM4 | 5'GACTGCGTACGAATTTGA3' |

Population structure analysis indicated that when the estimated best number of subpopulations was five (K = 5), the standard deviation in the mean estimated Ln probability of the data was minimum. Therefore, five subpopulations were indicated to be the best fit for the structure analysis. A cut-off identity value of 60% was used to group accessions. Accessions with an identity value under 60% were considered as admixed. Among the 15 accessions, four were admixed (26.7% of all accessions). Subpopulation I had one (6.7% of all accessions), subpopulation II had three (20%), subpopulation III had two (13.3%), subpopulation IV had two (13.3%) and subpopulation V had four accessions (26.7% of all accessions) (Figure 2).

## DISCUSSION

*S. pinnata* is a non-model organism which is of scientific interest because it is a selenium hyperaccumulator. To date, 114 genomic DNA/RNA sequences are available for the species (https://www.ncbi.nlm.nih.gov) and the entire genome of *S. pinnata* has not been sequenced. Therefore, SRAP markers, which do not require species sequence information, were used in this study to examine genetic variation among 15 *S. pinnata* accessions. The results indicated very high polymorphism at the DNA level with an average band polymorphism of 97.9% and 83% of primer combinations with 100% polymorphism (Table 2). High polymorphism was expected as *S. pinnata* is a self-incompatible plant, therefore it reproduces by cross-pollination. These results also confirmed that the SRAP markers were highly polymorphic and, therefore, effective molecular markers to analyze the genetic diversity and population structure of *S. pinnata*.

Average genetic distance of the accession pairs varied up to 0.87 with an average of 0.56 indicating high genetic diversity in *S. pinnata*. Such high diversity is expected from self-incompatible plants. For example, Koelling et al. (2011) compared genetic diversity of two species of Leavenworthia (Brassicaceae) with self-incompatible and self-compatible populations. They showed that self-incompatible individuals had two to three times greater mean genetic diversity than self-compatible ones. Although no other research has been performed with *S. pinnata*, Tanhuanpää et al. (2016) studied genetic diversity in the related self-incompatible species Brassica rapa using SNP markers. They found a mean gene diversity of only 0.34. The disparity between *S. pinnata* and B. rapa is most probably due to the fact that B. rapa is a crop plant and genetic diversity may have been lost during domestication and breeding of this species.

The least genetic diversity was observed between W642677 and W642678 which were collected in

the contiguous states of Colorado and New Mexico, respectively. The greatest genetic distance was between W639201 and W642678. This result was consistent with our expectations because they were the two most distantly located accessions, originating from Oregon and New Mexico, respectively (Figure 3a).

Dendrogram analysis divided the 15 genotypes into three clusters (Figure 1). Cluster I was the largest with seven members and included accessions mostly originating from Idaho (three of seven) and Utah (two of seven). Cluster I was further divided into two subclusters: A and B. W637559, W639198 and W640612 originating from Idaho were grouped in the same subcluster (A) (Figure 3). Cluster II was more variable and included accessions with different origins (Figure 1) which were, nevertheless, genetically similar. This similarity might have been the result of self-incompatibility because it is known that a wide range of gene flow over large geographical areas due to outcrossing can cause genetic similarity among populations (Schaal, 1980; Charlesworth and Charlesworth, 1987; Quinn et al., 2011). Cluster III was the smallest one with three members which were mostly from Nevada (two of three) (Figure 3). W639200, W632751, W627186 and W642096 coming from Utah were distributed in different clusters (Figure 1, Figure 3a). This disparity might be due to self-incompatibility (Quinn et al., 2011) in combination with geographical isolation due to mountain ranges.

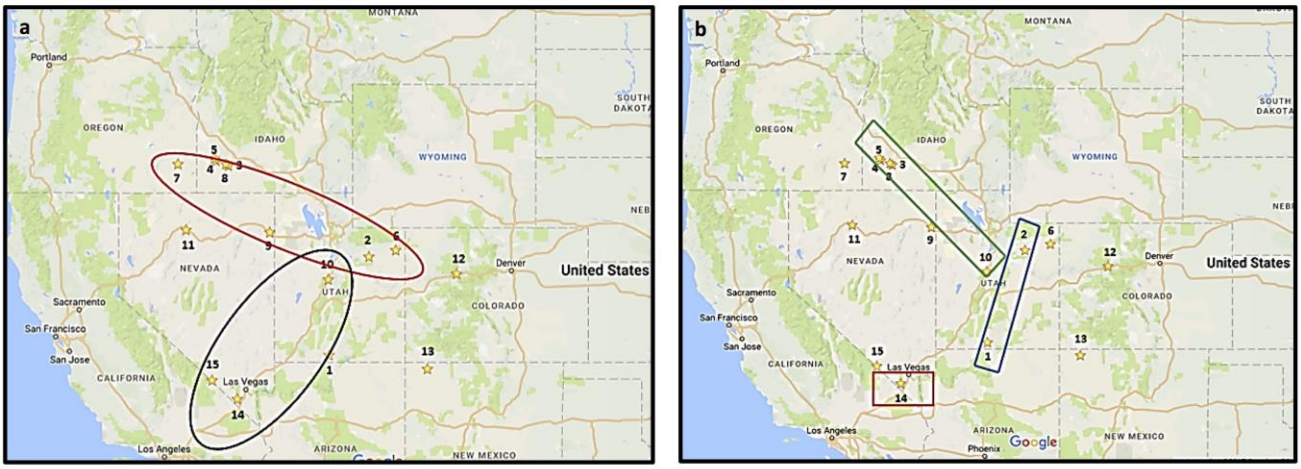


Figure 3. Geographic distribution of 15 accessions of *S. pinnata* in the USA. *S. pinnata* accessions (stars) are numbered 1-15 (see Table 1). a Red and black circles indicate Cluster I and III based on unweighted neighbor joining tree using the Dice coefficient, respectively. b Red, green and blue boxes represent subpopulations 1, 2 and 3, respectively, based on population structure analysis.

Population structure analysis divided the 15 *S. pinnata* accessions into 5 subpopulations (Figure 2). The majority (75%) of accessions originating from Idaho were in subpopulation 2 and all of the accessions from Utah were in subpopulation 3 (Figure 3b). Although population structure analysis results had some similarity with the dendrogram results, they did not completely agree. This discrepancy is probably due to the fact that the analyses are based on different clustering methods. The dendrogram was constructed with a distance matrix-based method that uses the neighbor joining algorithm (Saitou and Nei, 1987). On the other hand, structure analysis is based on a model and Bayesian clustering which uses allele frequency as its input (Pritchard et al., 2000).

This study showed that SRAP markers are a convenient system to study genetic diversity and population structure in *S. pinnata*. *S. pinnata* is a hyperaccumulator of selenium and plays an

important role in the elucidation of Se hyperaccumulation mechanism (Freeman et al., 2010). Future studies can be carried out to understand the evolution of Se hyperaccumulation and to associate individual genotypes with their hyperaccumulation capacity. Selected individuals can potentially be used to clean up excess Se from selenium-polluted soils and wastewater. Additionally, selenium concentration of food crops can be increased through transfer of gene(s) playing role(s) in selenium hyperaccumulation from these selected individuals. Then, farm animals could be fed with *S. pinnata* or other forage with enhanced Se content. In this way, *S. pinnata* can directly or indirectly provide sufficient dietary Se to individuals in areas where Se deficiency is prevalent.

## Acknowledgements

Authors acknowledge the support of their host institute.

## Conflict of interest

The authors have declared that no competing or conflict of interests exist. The funders had no role in study design, writing of the manuscript and decision to publish.

## Authors' contributions

ABK performed the research, analyzed the data and drafted the paper. SD designed the study and helped revise the draft. AF designed the study, participated in data analysis and revised the draft.

## REFERENCES

Abuzayed M, El-Dabba N, Frary A, Doganlar S. GDdom: An online tool for calculation of dominant marker gene diversity. Biochem Genet. 2017; 55: 155.

Ali H, Khan E, Sajad MA. Phytoremediation of heavy metals—concepts and applications. Chemosphere. 2013; 91: 869-881.

Bañuelos GS, Arroyo I, Pickering IJ, Yang SI, Freeman JL. Selenium biofortification of broccoli and carrots grown in soil amended with Se-enriched hyperaccumulator *Stanleya pinnata*. Food Chem. 2015; 166: 603-608.

Cappa JJ, Cappa PJ, Mehdawi AFE, McAleer JM, Simmons MP, Pilon-Smits EAH. Characterization of selenium and sulfur accumulation across the genus Stanleya (Brassicaceae): a field survey and common-garden experiment. Am J Bot. 2014; 5: 830-839.

Charlesworth D, Charlesworth B. Inbreeding depression and itsevolutionary consequences. Ann Rev Ecol Sys. 1987; 18: 237-268.

Chaudhry TM, Hayes WJ, Khan AG, Khoo CS. Phytoremediation-focusing on accumulator plants that remediate metal-contaminated soils. Aust J Ecotoxicol. 1998; 4: 37-51.

Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987; 19: 11-15.

Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet Resour. 2012; 4: 359-361.

Feist LJ, Parker DR. Ecotypic variation in selenium accumulation among populations of *Stanleya pinnata*. New Phytol. 2001; 149: 61-69.

Fordyce FM (2005). Selenium deficiency and toxicity in the environment. In: Selinus O, editor. Essentials of Medical Geology, Amsterdam, Netherlands: Elsevier, 2005; 373-415.

Freeman JL, Tamaoki M, Stushnoff C, Quinn CF, Cappa JJ, Devonshire J, et al. Molecular mechanisms of selenium tolerance and hyperaccumulation in *Stanleya pinnata*. Plant Physiol. 2010; 153: 1630-1652.

Koelling VA, Hamrick JL, Mauricio R. Genetic diversity and structure in two species of Leavenworthia with self-incompatible and self-compatible populations. Heredity. 2011; 106: 310-318.

Li G, Quiros CF. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theor Appl Genet. 2001; 103: 455-461.

Perrier X, Jacquemoud-Collet JP. DARwin software. 2006. Website http://darwin.cirad.fr/ [Accessed 2016].

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945-959.

Quinn CF, Prins CN, Freeman JL, Gross AM, Hantzis LJ, Reynolds RJB, et al. Selenium accumulation in flowers and its effects on pollination. New Phytol. 2011; 192: 727-737.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4: 406-425.

Schaal BA . Measurement of gene flow in Lupinus texensis. Nature. 1980; 284: 450-451.

Schiavon M, Pilon-Smits EAH . The fascinating facets of plant selenium accumulation-biochemistry, physiology, evolution and ecology. New Phytol. 2017; 213: 1582-1596.

Sun SJ, Gao W, Lin SQ, Zhu J, Xie BG, Lin ZB. Analysis of genetic diversity in Ganoderma population with a novel molecular marker SRAP. Appl Microbiol Biotechnol. 2006; 72: 537-543.

Tanhuanpää P, Erkkilä M, Tenhola-Roininen T, Tanskanen J, Manninen O. SNP diversity within and among Brassica rapa accessions reveals no geographic differentiation. Genome. 2016; 59: 11-21.

USDA, NRCS. The PLANTS Database. National Plant Data Team, Greensboro, NC 27401-4901 USA. 2017. Website http://plants.usda.gov/ [Accessed 17 April 2017]

Wu Z, Bañuelos GS, Lin ZQ, Liu Y, Yuan L, Yin X et al. Biofortification and phytoremediation of selenium in China. Front Plant Sci. 2015; 6: 136.