

# Detection of Adaptive Genetic Diversity in Wild Potato Populations and Its Implications in Conservation of Potato Germplasm

Alfonso H. del Rio, John B. Bamberg

USDA/Agricultural Research Service, US Potato Genebank, Sturgeon Bay, WI, USA

Email: adelrioc@wisc.edu

**How to cite this paper:** del Rio, A.H. and Bamberg, J.B. (2020) Detection of Adaptive Genetic Diversity in Wild Potato Populations and Its Implications in Conservation of Potato Germplasm. *American Journal of Plant Sciences*, 11, 1562-1578.

<https://doi.org/10.4236/ajps.2020.1110113>

**Received:** September 17, 2020

**Accepted:** October 18, 2020

**Published:** October 21, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

A better understanding on how genetic diversity is structured at natural habitats can be helpful for exploration and acquisition of plant germplasm. Historically, studies have relied on DNA markers to elucidate potato genetic diversity. Current advances in genomics are broadening applications allowing the identification of markers linked to genomic regions under selection. Those markers, known as adaptive markers, unlock additional ways to value and organize germplasm diversity. For example, conservation priorities could be given to germplasm units containing markers associated to unique geographic identity, and/or linked to traits of tolerance to abiotic stresses. This study investigated if adaptive marker loci were possible to be identified in a large AFLP marker dataset of ninety-four populations of the wild potato species *S. fendleri*. These populations originated from six different mountain ranges in southern Arizona, USA. A total of 2094 polymorphic AFLP markers were used to conduct genetic diversity analyses of populations and mountain ranges. Adaptive markers were detected using Bayesian methods which distinguished marker loci departing significantly from frequencies expected under neutral models of genetic differentiation. This identified 16 AFLP loci that were considered to be adaptive. To contrast diversity parameters generated with each set of markers, analyses that included all the 2094 AFLP markers, and only the 16 adaptive markers were conducted. The results showed that both were efficient for establishing genetic associations among populations and mountain ranges. However, adaptive markers were better on revealing geographic patterns and identity which would suggest these markers were linked to selection at the natural sites. An additional test to determine if adaptive markers associated to climate variables found two loci associated to specific climate variables in populations from different regions but sharing similar environmental structure. The distribution of adaptive markers

---

among populations revealed that only two were needed to build a core subset able to keep all the markers. This preliminary assessment shows that adaptive genetic diversity could offer an additional way to measure diversity in potato germplasm and to set up options for conservation and research.

### Keywords

Adaptive Genetic Diversity, AFLP Markers, Plant Population Structure, Potato Germplasm, *Solanum fendleri*

---

## 1. Introduction

Use of valuable genetic diversity found in wild relatives is a traditional approach in plant breeding to add sources of resilience to pests and diseases and, adaptation to abiotic stresses. From that perspective, the implementation of *ex situ* crop germplasm repositories has played a key role in collecting, preserving, organizing, and very importantly, making that diversity quickly available to breeding and research groups. Repositories therefore require efficient work at all the phases of the conservation process [1] [2]. The most important is to implement methods that do not affect the genetic diversity preserved at genebanks. Knowing how much diversity is in the collections is a way to identify if undesirable changes are taking place and propose mitigation plans if needed [3].

Genebanks have relied on DNA or molecular markers as tools to quantify genetic variation and, to answer questions on diverse areas of the germplasm conservation process [4] [5]. Many of these studies have largely used PCR, fragment-based techniques, like RAPDs, SSRs, AFLPs, etc. These techniques assume that most of the DNA fragment sequences were neutral, that is, originated in genomic regions without gene expression or fitness advantage. In recent years however, advances in population genomics and computational methods have revealed that these markers are not totally neutral and that some of them actually are in regions experiencing selection [6]. This type of markers can theoretically identify genetic diversity associated to local selection and adaptation and thus, unlock additional options for characterization and organization of germplasm. For example, markers associated to local selection could identify germplasm units with unique environmental-geographical structure and/or with positive expression of genetic traits resulting from selective pressure to pests/diseases [7]. Previous works [8] [9] [10] [11] have indicated that robust genome scans done with dominant markers, like AFLPs, can provide enough number of markers and genome coverage to search for marker loci with adaptive signals. This uses extensive simulations to gauge population differentiation and identify marker loci with frequencies departing from those expected under neutral models of genetic differentiation. Significant changes in frequency would imply that loci were subjected to selection and therefore, potentially adaptive. Bayesian methods and  $F_{ST}$ -outlier tests have proved to be effective when assess-

ing populations from the same taxa but adapted to different environmental conditions. Markers influenced by selection are anticipated to show unexpected high or low departures relative to the null dispersal of  $F_{ST}$  for markers not under selection. This method has then become very popular because it is easy and fast to implement [10] [12].

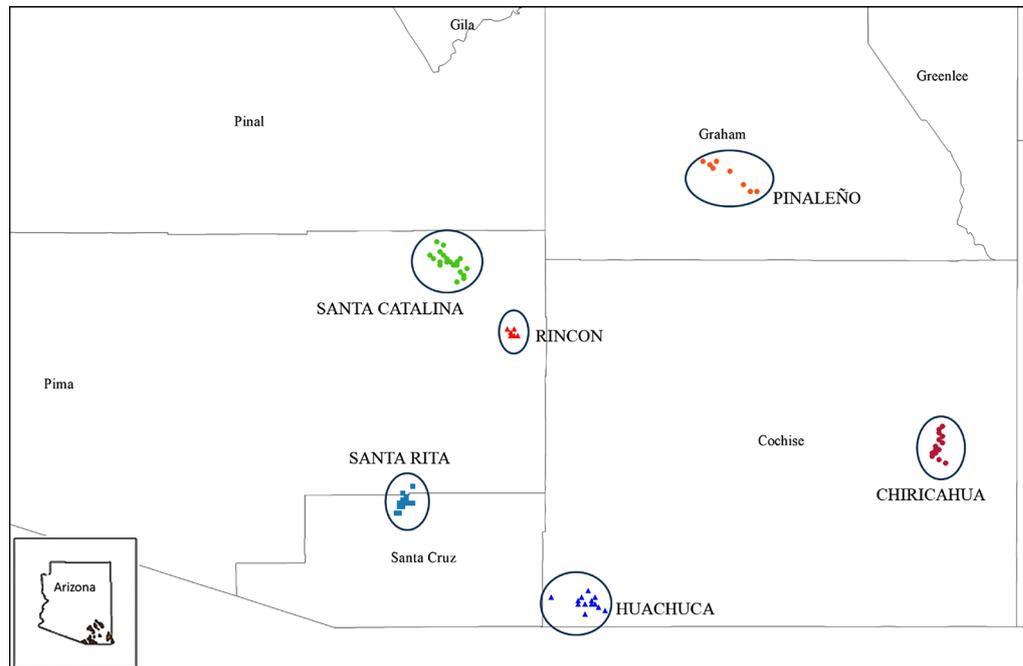
The USPG holds more than 6000 accessions that include collections of potato wild relatives. These species distribute from southern Chile to the southwest USA and, have shown to be excellent sources of genetic traits that can be transferred to enhance the crop [4] [13]. Research at the USPG has utilized molecular markers to determine quantity and quality of the genetic diversity acquired and preserved in the collections. These studies have offered scientific foundations to, for example, support strategies for plant exploration and collecting [14], validate conservation approaches [15] and, assemble core subsets [16]. The important goal at the USPG has been to create well-informed scientific knowledge to assist decisions on potato germplasm conservation [4]. Another significant result from these studies was that genetic differentiation in natural populations, estimated with neutral markers, was not proportionally correlated to geographic separation nor linked to variation in environmental conditions [17]. Moreover, additional studies reported that DNA markers also failed to predict associations of phenotypic diversity (likely originated from local evolutionary processes) with geographically divergent taxa [18].

Here we are expanding our previous study [19] that applied richness of unique AFLP alleles to determine value for conservation in populations of the potato species *S. fendleri*. That study revealed that populations collected in Pinaleno and Chiricahua mountain ranges captured most of the unique markers so that genebanks should prioritize those regions to add diversity. This unlocked the idea of assessing for adaptive genetic diversity since it would offer an additional way to define germplasm value for conservation priorities. Therefore, this study evaluated patterns of genetic diversity in a group of geographically and environmentally diverse populations of *S. fendleri* collected in Southern Arizona to investigate the possibility of detecting adaptive markers and testing their potential use in conservation efforts.

## 2. Materials and Methods

### *Plant materials and sampling sites*

The study was intended to identify levels of genetic differentiation in 94 populations of the wild potato species *S. fendleri* from six different mountain ranges in southern Arizona, USA (Figure 1). These populations were collected in different USPG exploration trips [19] [20]. *Solanum fendleri* is a tetraploid ( $2n = 4x = 48$ ) with disomic segregation. It is able to reproduce sexually either by selfing or outcrossing, and asexually by tuber propagation. The populations used this study inhabit a particular set of mountain ranges known as the *Sky Islands*. As the name implies these mountain ranges are isolated by large extensions of desert, valleys or grassland with distinct topographic and climatic conditions. It



**Figure 1.** Geographic distribution of the 94 different populations of *S. fendleri* used in this study in the six mountain ranges of southern Arizona where they were collected.

is believed that that setting restricts gene flow resulting in strong genetic isolation and differentiation among plant populations. **Table 1** provides additional details about these populations as germplasm units and part of the *S. fendleri* collection at the US Potato Genebank in Sturgeon Bay, Wisconsin. **Figure 1** shows the mountain range locations in Arizona and the geographic dispersal of *S. fendleri* populations used in this study.

Each mountain range was composed of a different number of populations; there were 24 populations for Santa Catalina mountains (CAT), 17 for Santa Rita (RIT), 15 for Huachuca (HUA), 20 for Chiricahua (CHI), 8 for Pinaleño (PIN) and 10 for Rincon (RIN) (**Figure 1**). Specifics of the geographic coordinates and habitat descriptions for each population at each mountain range are available at the USPG databases (<https://npgsweb.ars-grin.gov/gringlobal/search.aspx>). To assign patterns of genetic identity and variation, each mountain range was assumed to represent all the diversity of the populations it contained. That increased the odds of identifying marker loci that were uniquely expressed at each mountain range.

#### ***DNA extraction and Generation of AFLP markers***

Extraction of DNA is described in [14]. In short, genomic DNA was isolated from individual young leaves in a bulk of 27 individual plants for each population using the DNeasy Plant Mini kit (Qiagen Inc., Valencia, CA, USA). The sampling strategy and extraction methods were set to counter the effect of genetic heterogeneity and to provide an adequate representation of total diversity of the population [21]. The generation of AFLP markers and estimation of genetic diversity followed USPG methods shown to be effective for wild potato

**Table 1.** List of populations of *S. fendleri* used in this study and their geographic origins. The numbers correspond to the Plant Introduction (PI) number assigned by the US Potato Genebank. Additional information can be found at USPG website (<https://npgsweb.ars-grin.gov/gringlobal/search.aspx>).

Mountain Range	PI Number
CAT	458421, 564024, 578234, 585112, 632327, 641039, 641040, 658169, 658170, 658171, 658172, 658173, 658174, 658175, 658176, 658177, 658178, 658179, 658180, 658181, 658182, 658183, 658184, 660273
CHI	275162, 275163, 564026, 564027, 564028, 592405, 592406, 636399, 636400, 636402, 636404, 636405, 636406, 636407, 636408, 636409, 636410, 636411, 636412, 636414
HUA	275165, 275167, 283100, 564025, 641027, 641028, 641030, 641031, 641032, 641033, 641034, 641035, 641036, 641037, 641038
PIN	585115, 655247, 655249, 655250, 655251, 658166, 655257, 658167
RIN	643993, 643994, 643995, 643996, 643997, 643998, 643999, 644000, 644002, 644003
RIT	660257, 660258, 660259, 660262, 660263, 660264, 660265, 660266, 660267, 660268, 660269, 660274, 660276, 660277, 660279, 660280, 660281

species. Only polymorphic loci within *S. fendleri* populations were included in the analysis as they were true informative of genetic variation across genotypes and populations.

#### ***Estimation of genetic diversity and identification of adaptive AFLP markers***

To determine levels of genetic diversity in the populations, polymorphic AFLP markers were used to calculate genetic relationships among populations as well as other standard estimates of population genetic structure using GenAlex version 6.5 [22]. Constellation plots [23] based on hierarchical and multivariate cluster analysis and Principal Component (PC) analysis were used to present visual representations of genetic diversity and differentiation among populations and mountain ranges.

To detect adaptive genetic diversity, an analysis to identify *signatures* of selection in the AFLP marker dataset was used. The approach assumed that frequencies at some marker loci were not random and exhibited unusual patterns of differentiation not explained by chance and probably reflecting selection. A statistical multinomial-Dirichlet model to identify these loci was applied using the program BayeScan [24]. BayeScan is presented as a hierarchical Bayesian method less susceptible to false positive detection because it accepts for different demographic scenarios and amounts of genetic drift in the studied populations. It does not make any assumptions about the genotypic structure of the populations under consideration which fits well the dynamics and genetic structures observed in wild potato population. The program calculates the odds of a locus being under selection using two models, one that includes the effect of selection and another that excludes selection using a reversible jump Markov chain Monte-Carlo method. The default parameters used in this study were the ones

indicated by BayeScan recommendations (20 pilot runs of 5000 iterations, 100,000 total iterations). The analysis was run three times to validate the outcome. The significance of adaptive loci detection was validated using posterior odds (*i.e.*, the probability that an event will happen after all evidence has been taken into account) and used Jeffrey's scale with a false discovery rate ( $FDR \leq 0.05$ ) [24] which allowed identifying marker loci that can be reliably regarded as adaptive across populations.

#### ***Estimation of climate variables at collecting sites***

To determine associations of DNA marker variation with climate variables observed at the sites, climate data was collected based on the geographical location and elevation for each population site. The climate profiles were extracted from *Worldclim* databases <https://www.worldclim.org/>. These databases contain archives of global historical weather and climate data. The variables used included monthly averages of maximum, minimum, mean temperatures and rainfalls. In addition, variables known as bioclimatic variables were added. They consist of monthly temperature and rainfall that are supposed to produce more biologically meaningful parameters. Bioclimatic variables involve annual trends (*e.g.*, mean annual temperature, and annual precipitation), seasonality (*e.g.*, annual range in temperature and precipitation), and extreme or limiting environmental factors (*e.g.*, temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A total of 55 climate variables were applied in multivariate cluster analysis to establish climate-based associations and use them to test for adaptive marker associations.

### **3. Results and Discussion**

#### ***Genetic diversity among populations***

All of the ninety-four populations of *S. fendleri* were effectively genotyped with 2094 AFLP markers generated using the USPG standard protocol. When all of these markers were utilized for genetic diversity assessments, it was found that the average percentage of polymorphic AFLP loci across populations was 69.5%. The highest and lowest percentages of polymorphic loci were found in CHI (77.08%) and RIN (60.17%) (**Table 2**).

Genetic distance pairwise associations among populations showed that the range between two populations went from a high of 42% to a low of 9%, with an average of 23.5%. A hierarchical cluster analysis presented in a Constellation plot shows that all the populations were able to be distinguished individually and that no duplicate populations were found (**Figure 2**). The plot also shows that, most times, populations from the same mountain range do not cluster together. In fact, only a few clusters of populations sharing the same geographic origin are generated; *i.e.*, some CAT populations are clustering in one unique large group. (**Figure 2**). From a practical standpoint, however, this is an important result—it shows that AFLP markers can be used to differentiate the populations of *S. fendleri* used in the study as separate and unique germplasm units.

**Table 2.** Genetic diversity and genetic distance parameters estimated with 2094 AFLP markers for populations of *S. fendleri* from six different mountain ranges of southern Arizona, USA.

Mountain Range	Genetic Diversity Analysis			Pairwise Population Matrix of Nei's Genetic Distance				
	<i>n</i>	<i>h</i>	% Polymorphic loci	CAT	CHI	HUA	PIN	RIN
CAT	24	0.182	73.45%	0.000				
CHI	20	0.216	77.08%	0.027	0.000			
HUA	15	0.180	66.19%	0.026	0.031	0.000		
PIN	8.	0.236	68.05%	0.109	0.090	0.113	0.000	
RIN	10	0.195	60.17%	0.029	0.030	0.038	0.115	0.000
RIT	17	0.217	72.25%	0.030	0.023	0.035	0.097	0.025

*n* = number of populations. *h* = genetic diversity at the mountain range.



**Figure 2.** A hierarchical constellation plot built using pairwise genetic distances of the 94 populations of *S. fendleri* used in this study. The genetic distance coefficients were estimated using 2094 AFLP markers. The numbers identifying each individual population are the PI numbers indicated in **Table 1**.

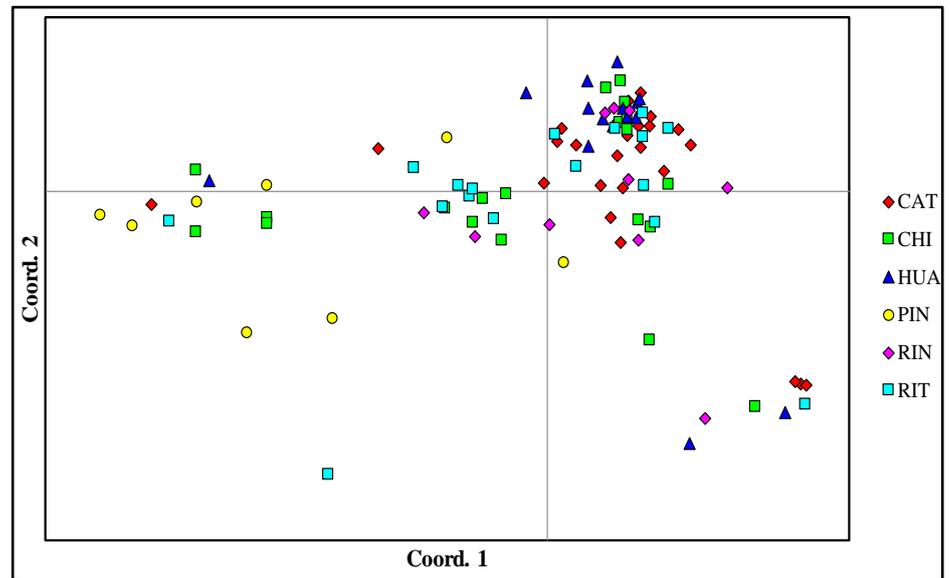
### ***Genetic diversity and structure among mountain ranges***

The levels of genetic diversity ( $h$ ) within-population at each mountain range ranged from 0.180 to 0.236 with an average of 0.204. The populations from PIN mountain range exhibited the highest genetic diversity, while HUA and CAT revealed the lowest (**Table 2**). This was particularly interesting, PIN mountain range contained the smallest number of populations among all the ranges, yet it showed the highest variation. A previous study [19] reported that PIN populations were also set apart because of their highest number of unique AFLP markers when compared with other mountain ranges. The additional number of unique loci could have been responsible here for increasing  $h$  levels.

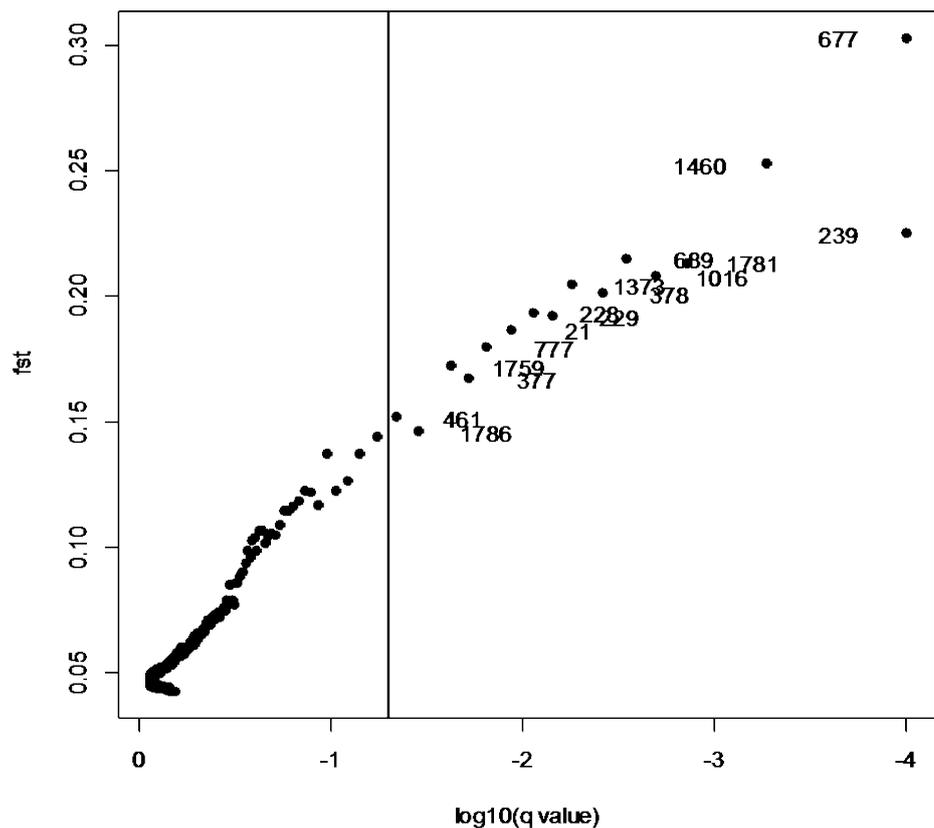
A multivariate Principal Component (PC) analysis based on covariance of genetic distances confirmed that *S. fendleri* populations did not exhibit consistent association patterns based on their regional origins (**Figure 3**). In fact, in contrast, it was observed that populations belonging to different geographic ranges clustered together. To determine genetic relationships among mountain ranges, a distance matrix using Nei's genetic distance coefficients [22] was generated. Nei's coefficients assumed frequency variation of AFLP marker alleles within populations at each mountain range and the variation over all loci. This analysis revealed that, genetically, mountain ranges were very close related. The average genetic distance between two mountain ranges was only 5%, ranging from 2% to 12%. Since PIN had the highest number of unique AFLP alleles, this mountain range was the most differentiated from the rest with an average of 10.5%. Other studies [19] had also reported that PIN populations contained the highest number of unique alleles among all the mountain ranges. On the other hand, RIT showed the least genetic differentiation overall with an average of 3.7%. In general, genetic differentiation levels were quite small which can explain why populations from different mountain ranges failed to resolve a coherent pattern based on genetic and geographic origin. This also corroborates previous results that genetic differentiation in natural potato populations cannot be correlated to geographic separation, or associated with specific ecological or environmental variables at the sites [17] [25].

### ***Detection of adaptive genetic markers***

We investigated whether Bayesian methods were useful to identify AFLP marker loci departing from neutral expectations of genetic differentiation in geographically (and environmentally) diverse potato populations. Previous work [26] had indicated that a robust coverage of AFLP markers across genomes would enhance the odds of finding non-neutral markers near to gene regions and/or to segments linked to differentiation in the genome. From that standpoint, 2094 was a significant number of AFLP markers to allow searching for outlier marker loci. With a detection confidence level at 95%, the BayeScan run discovered a total of 16 outlier loci possibly under selection. The parameters used followed threshold recommendations given by the program for a *very strong evidence of selection* ( $FDR < 0.05$ ;  $\log_{10}(PO) = 1.2$ ) (**Figure 4**).



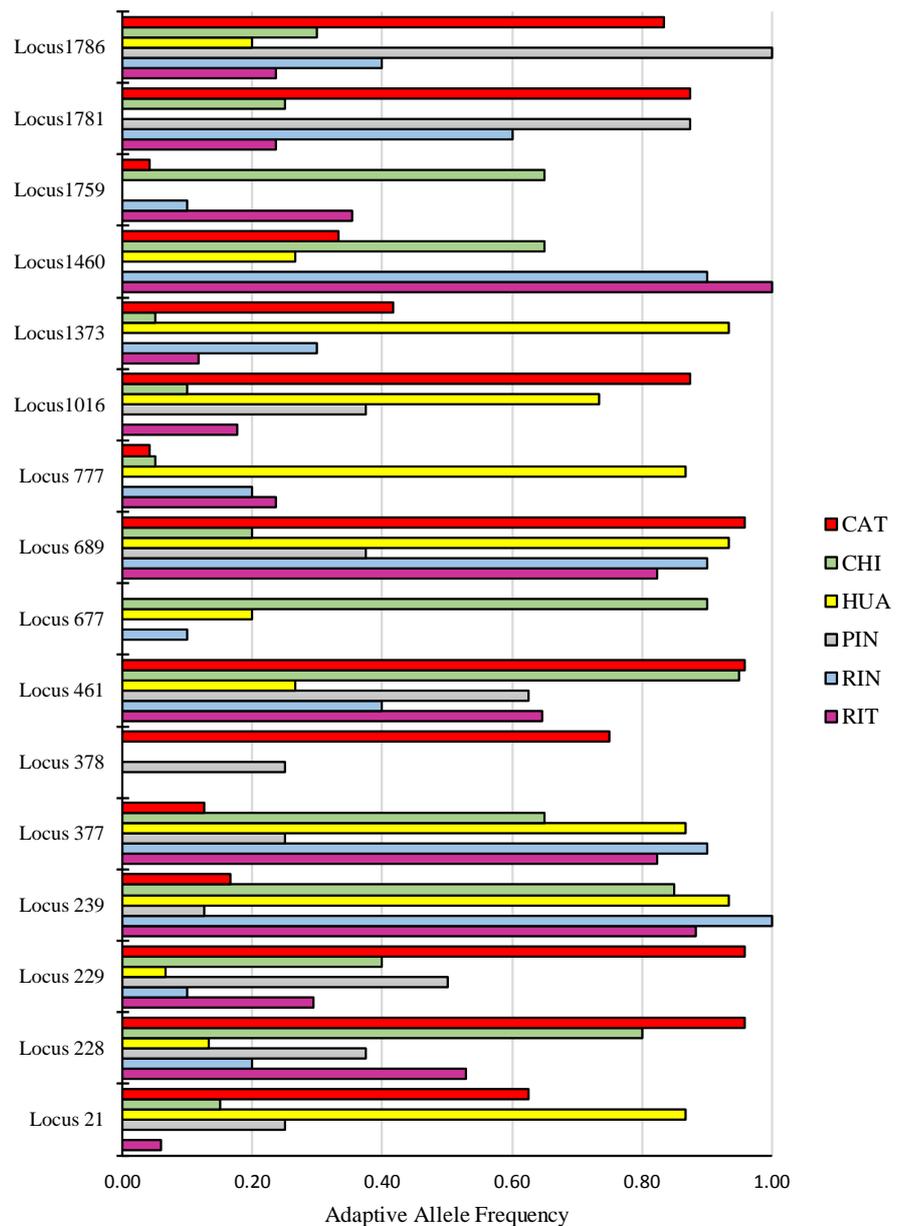
**Figure 3.** Principal component analysis showing genetic associations among the ninety-four *S. fendleri* populations used in the study. This PC graph was created using pairwise genetic distances estimated with 2094 AFLP markers.



**Figure 4.** A BayeScan plot showing all of the 2094 AFLP loci distributed in the ninety-four populations of *S. fendleri* from six different mountain ranges. The vertical line is the threshold ( $\text{Log}(\text{PO}) = 1.2$ ) used for identifying outlier loci. The dots that fall to the right of the threshold line are identified as outlier or adaptive loci. The numbers are the AFLP loci identification code used in the analysis.

These 16 adaptive marker loci exhibited variable frequencies along the mountain ranges (Figure 5). For instance, it was noted that in 14 marker loci the alleles were not found at all in a given mountain range (*i.e.*, allele at locus 677 was not present in CAT, allele at locus 378 was not found in CHI). In the other extreme, three cases showed adaptive marker loci with an allele frequency of 1.00 (see Figure 5). The adaptive AFLP locus most frequently observed was locus 689 with a frequency of 0.698 across mountain ranges, while the less frequent was locus 378 with a frequency of 0.167. In fact, locus 378 was almost unique, it was only detected in two mountain ranges (CAT and PIN) (Figure 5).

A point to remark was that the ability to identify adaptive genetic diversity



**Figure 5.** Complete list of the AFLP marker loci identified with BayeScan as adaptive and their frequencies in populations of *S. fendleri* at each mountain range.

with a Bayesian method required some key factors, in particular, marker coverage and the magnitude of linkage disequilibrium. At the time we wrote this paper, there was no specific data on genome size for *S. fendleri*. but according to the c-values database from the Kew Gardens (<http://data.kew.org/cvalues/>), the largest genome size for potatoes was that of *S. tuberosum*, about 860 Mb. Assuming this value as a possible higher estimate for the genome size of *S. fendleri* and taking into account the total number of AFLP markers used here, a conservative estimate for genome coverage was adequate; about one marker every 750 Kb on average. On the other hand, the level of linkage disequilibrium in *S. fendleri* is unknown, but was probably low because this species has some levels of outcrossing. Therefore, any of the detected outlier marker was probably closely linked to an adaptive gene or to a mutation truly responsible for the atypical pattern of genetic diversity [27].

#### ***Genetic diversity and geographic structure based on adaptive markers***

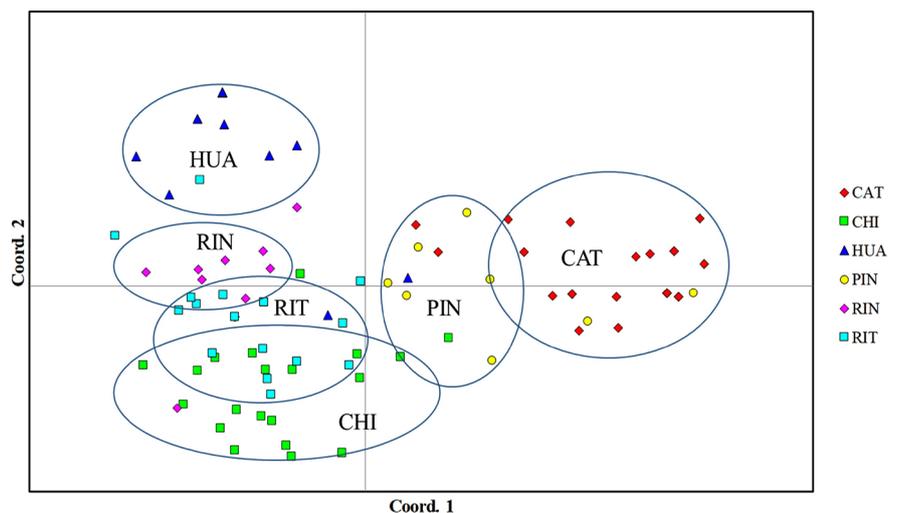
One question was whether adaptive markers were superior than neutral markers on resolving coherent patterns of geographic identity and differentiation. If these markers reflect genome regions under local selection, then populations would reveal some geographic structure. This would not occur if the markers come from neutral genomic regions with no linkage to loci under local selection. To investigate this possibility, we conducted an analysis of genetic diversity but this time including only the 16 AFLP adaptive markers from the BayeScan run. The results showed that, when this new data was compared to the data generated with all of the 2094 AFLPs, the levels of genetic diversity ( $h$ ) within mountain ranges were higher in adaptive markers (~14% more on average). Likewise, the averaged percentage of polymorphic loci increased significantly from 69.5% to 81.2% (Table 3).

A multivariate cluster analysis based on a genetic distance matrix built with adaptive markers showed stronger genetic similarities among populations from the same mountain range, and enhanced genetic differentiation among mountain ranges (Figure 6). This improved discriminating power can be explained by the higher genetic distances found among mountain ranges. The averaged Nei's genetic distance between mountain ranges with adaptive markers was 37% (compared to only 5% when all the AFLP markers were included). The genetic distance between HUA and CAT mountain ranges was the highest (63%) while the lowest was found between RIN and RIT (5%). The PC graph clearly showed how adaptive markers had better power to merge populations from the same site, and to set apart the geographic ranges (Figure 6). Other than some exceptions, most of the time populations from the same mountain range clustered together. A statistical comparison to determine how the genetic distances estimated with the 16 adaptive markers and with all the markers confirmed a very low, no significant correlation. The Mantel test for matrix comparisons indicated an  $r$  value of only 9% (data not shown). This analysis corroborated the lack of agreement between information given by these two types of markers.

**Table 3.** Genetic diversity and genetic distance parameters estimated with 16 adaptive AFLP markers identified by the program BayeScan for populations of *S. fendleri* in six different mountain ranges of southern Arizona, USA.

Mountain Range	Genetic Diversity Analysis			Pairwise Population Matrix of Nei's Genetic Distance				
	<i>n</i>	<i>h</i>	% Polymorphic loci	CAT	CHI	HUA	PIN	RIN
CAT	24	0.217	93.75%	0.000				
CHI	20	0.277	93.75%	0.554	0.000			
HUA	15	0.202	81.25%	0.627	0.543	0.000		
PIN	8	0.246	62.50%	0.175	0.376	0.575	0.000	
RIN	10	0.224	75.00%	0.550	0.242	0.271	0.321	0.000
RIT	17	0.269	81.25%	0.470	0.155	0.297	0.349	0.046

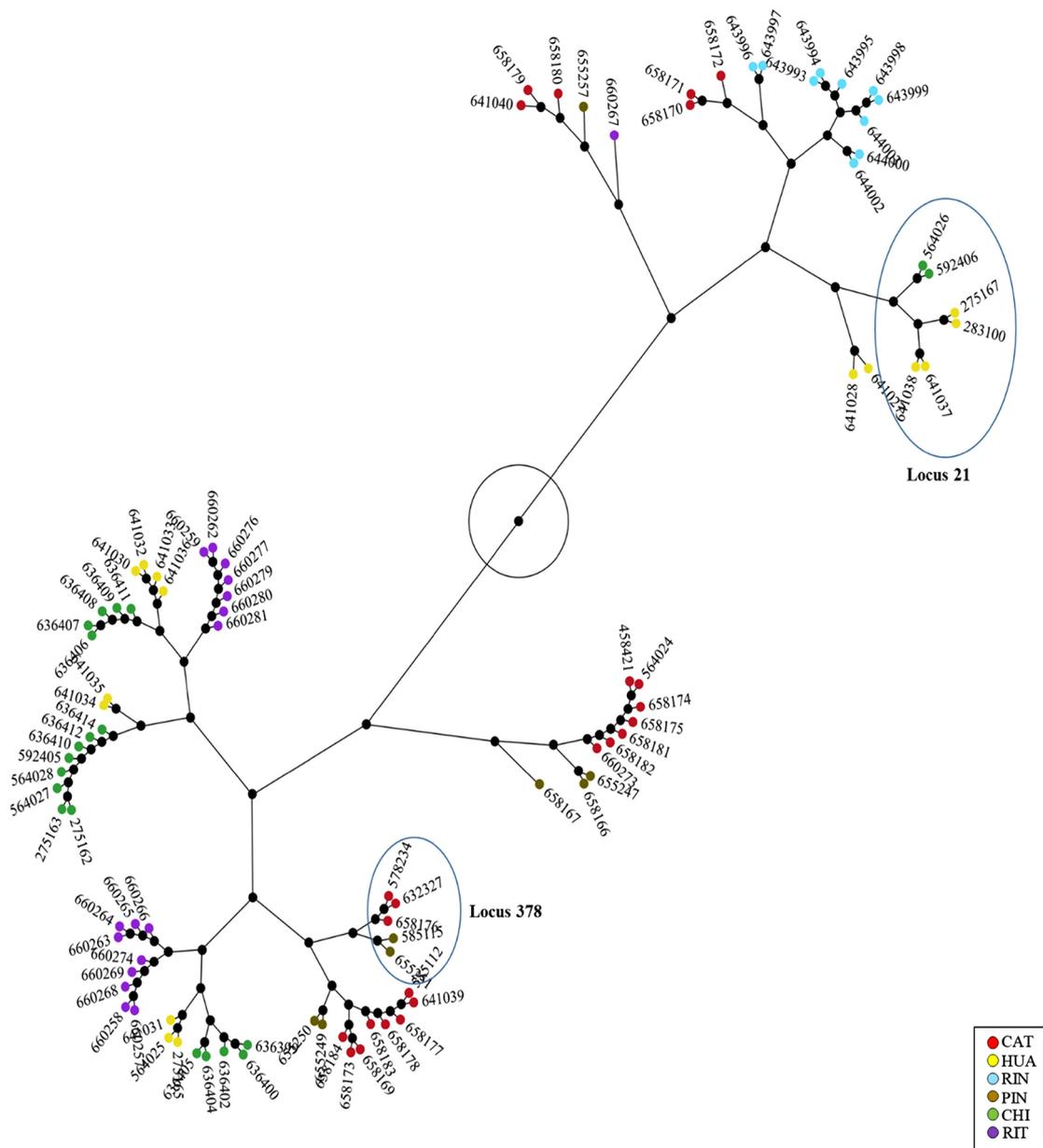
*n* = number of populations. *h* = genetic diversity at the mountain range.



**Figure 6.** Principal component analysis based on 16 adaptive markers showing the genetic diversity and structure of 94 *S. fendleri* populations in six different mountain ranges in Arizona, USA.

#### *Association of adaptive markers with climate variables*

Since adaptive markers showed coherent clustering based on geographic origins, we tested if adaptive loci were associated to specific environmental variables. A total of 55 climate variables were used in a hierarchical cluster approach to organize populations in climate-based relationships (Figure 7). A Constellation plot reflecting these associations revealed that most of the time populations from the same mountain range have similar climate structures. However, in some cases populations from different mountain ranges were placed together because of similar climate patterns (Figure 7). This apparent contradictory sorting is not unexpected. Climate relies on various topographical and ecological factors that influence variability. One known example is altitudinal gradients which can make environmental conditions change quickly over short



**Figure 7.** A hierarchical constellation plot showing the results of the cluster analysis using 55 different climate variables for the ninety-four *S. fendleri* populations distributed in the six mountain ranges. The populations included in the circles represent the clusters of populations from different mountain ranges with similar climate structure associated to a particular adaptive locus indicated on the side.

distances and make habitats very different even when they are located nearby, or very similar when they are far apart [28].

Two clusters containing populations from different mountain ranges revealed associations with specific adaptive loci. One cluster including two populations of PIN, 585115 and 655251, and three CAT populations (578234, 632327 and 658176) (Figure 7), shared the allele from locus 378. Moreover, this allele was exclusive for the two PIN populations. A detailed inspection of all of the climate variables found that all five populations had the same value for *Precipitation of*

*Wettest Quarter*. A second example was found in a cluster of two populations of CHI (564026 and 592406) and four from HUA (275167, 283100, 641028 and 641037). All of them shared the allele from adaptive locus 21. This allele was exclusive for the two populations of CHI (Figure 7). In this case, this cluster had the climate variable *Precipitation of Driest Month* similar for all the populations.

Additional climate-based clusters associating populations from different mountain ranges were examined but no more significant associations with markers were detected. Though this analysis was very preliminary, the results present a glimpse of how adaptive markers can be applied for germplasm screening and trait discovery. For example, using the examples above, the association of adaptive allele at locus 21 with a variable related to low rainfall could be the starting point to investigate if that germplasm offers genetic sources for adaptation to drought. In recent years, tolerance to abiotic stresses have become key genetic traits to identify as they offer chances for developing resilient varieties to climate change [29].

#### ***Core subset for adaptive markers***

An analysis of the presence of adaptive markers along individual populations at all mountain ranges showed that only two populations were needed to capture all of the 16 adaptive markers. Populations 658179 (from CAT) and 641032 (from HUA), when combined, included all the adaptive markers. However, if mountain range representation is sought as a parameter to include in a core subset, to capture the maximum number of adaptive markers at each of the mountain ranges, then a total number of populations needed is 18 (see Table 4).

## **4. Conclusions**

The prospect of using adaptive markers in germplasm is that it could add another way to value, screen and characterize germplasm for conservation and use. Because of their broad distribution, potato wild relatives and native landraces are good materials for mining useful genetic variants associated to environmental adaptation, particularly, as a response to abiotic stresses, or to diseases and pests.

**Table 4.** List of *S. fendleri* populations to be included in a core subset to capture the maximum number of adaptive markers at each mountain range.

<b>Population included in the core.</b> Number in parenthesis is number of adaptive loci in the population	<b>Mountain Range</b>	<b>Maximum number of adaptive loci found in the mountain range</b>
658179* (12), 641040 (11), 658173 (11)	CAT	14
592406 (10), 275162 (8), 564027 (8)	CHI	10
641032* (10), 641034 (10), 275167 (5)	HUA	10
655251 (8), 655257 (5), 658166 (4)	PIN	8
643995 (8), 643994 (7), 643999 (7)	RIN	8
660265 (9), 660259 (8), 660257 (7)	RIT	9

\*Populations that combined capture all of the 16 adaptive markers.

In recent years the expansion of high-throughput sequencing is facilitating genome-wide assessments in plant species. Methods like genotyping by sequencing (GBS) can potentially accelerate processes of identifying adaptive genetic diversity in potato species.

The results from this study proved that adaptive genetic diversity has enhanced capacity for detecting patterns of geographic structure in *S. fendleri*. This has the potential of supporting planning for future germplasm exploration and collecting. Another important finding was detecting marker association with specific environmental variables at the sites so that genebanks have options of identifying *evolutionary significant units*, in other words, chances of detecting germplasm units or groups with unique genetic distinctiveness due to environmental adaptation. Ultimately, the understanding of the mechanisms that shape genetic variation could help to predict responses of populations to future environmental conditions.

### Acknowledgements

The authors express their thanks to the US Potato Genebank and staff for their cooperation. We also thank the technical assistance given by the staff at the University of Wisconsin-Madison Biotech Center, and Ms. Natalia del Rio for her constructive comments and assistance in appropriate editing of the manuscript.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Fowler, C. and Hodgkin, T. (2004) Plant Genetic Resources for Food and Agriculture: Assessing Global Availability. *Annual Review of Environment and Resources*, **29**, 143-179. <https://doi.org/10.1146/annurev.energy.29.062403.102203>
- [2] Gepts, P. (2006) Plant Genetic Resources Conservation and Utilization. *Crop Science*, **46**, 2278-2292. <https://doi.org/10.2135/cropsci2006.03.0169gas>
- [3] Byrne, P.F., Volk, G.M., Gardner, C., Gore, M.A., Simon, P.W. and Smith, S. (2018) Sustaining the Future of Plant Breeding: The Critical Role of the USDA-ARS National Plant Germplasm System. *Crop Science*, **58**, 451-468. <https://doi.org/10.2135/cropsci2017.05.0303>
- [4] Bamberg, J.B., del Rio, A.H., Jansky, S. and Ellis, D. (2018) Ensuring the Genetic Diversity of Potatoes. In: Wang-Pruski, G., Ed., *Achieving Sustainable Cultivation of Potatoes Volume 1: Breeding, Nutritional and Sensory Quality*, Burleigh Dodds Science Publishing, Cambridge, 320 p. <https://doi.org/10.19103/AS.2017.0016.02>
- [5] Van Hintum, T.J. and Van Treuren, R. (2002) Molecular Markers: Tools to Improve Genebank Efficiency. *Cellular and Molecular Biology Letters*, **7**, 737-744.
- [6] Luikart, G., England, P.R., Tallmon, D., Jordan, S. and Taberlet, P. (2003) The Power and Promise of Population Genomics: From Genotyping to Genome Typing. *Nature Reviews Genetics*, **4**, 981-994. <https://doi.org/10.1038/nrg1226>

- [7] Bonin, A., Nicole, F., Pompanon, F., Miaud, C. and Taberlet, P. (2007) Population Adaptive Index: A New Method to Help Measure Intraspecific Genetic Diversity and Prioritize Populations for Conservation. *Biological Conservation*, **21**, 697-708. <https://doi.org/10.1111/j.1523-1739.2007.00685.x>
- [8] Petit, R.J., El Mousadik, A. and Pons, O. (1998) Identifying Populations for Conservation on the Basis of Genetic Markers. *Biological Conservation*, **12**, 844-855. <https://doi.org/10.1046/j.1523-1739.1998.96489.x>
- [9] Frankham, R., Ballou, J.D. and Briscoe, D.A. (2002) Introduction to Conservation Genetics. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511808999>
- [10] Bonin, A., Taberlet, P., Miaud, C. and Pompanon, F. (2006) Explorative Genome Scan to Detect Candidate Loci for Adaptation along a Gradient of Altitude in the Common Frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773-783. <https://doi.org/10.1093/molbev/msj087>
- [11] Moritz, C. (2002) Strategies to Protect Biological Diversity and the Evolutionary Processes That Sustain It. *Systematic Biology*, **51**, 238-254. <https://doi.org/10.1080/10635150252899752>
- [12] Nunes, V.L., Beaumont, M.A., Butlin, R.K. and Paulo, O.S. (2011) Multiple Approaches to Detect Outliers in a Genome Scan for Selection in Ocellated Lizards (*Lacerta lepida*) along an Environmental Gradient. *Molecular Ecology*, **20**, 193-205. <https://doi.org/10.1111/j.1365-294X.2010.04936.x>
- [13] Bamberg, J.B. and del Rio, A.H. (2005) Conservation of Potato Genetic Resources. In: Razdan, M.K. and Mattoo, A.K., Eds., *Genetic Improvement of Solanaceous Crops, Vol. 2: Potato*, Science Publishers, Enfield, NH, 451 p.
- [14] Bamberg, J.B., del Rio, A.H. and Penafiel, J. (2011) Successful Prediction of Genetic Richness at Wild Potato Collection Sites in Southeastern Arizona. *American Journal of Potato Research*, **88**, 398-402. <https://doi.org/10.1007/s12230-011-9205-4>
- [15] del Rio, A.H., Bamberg, J.B. and Huaman, Z. (1997) Assessing Changes in the Genetic Diversity of Potato Gene Banks. 1. Effects of Seed Increase. *Theoretical and Applied Genetics*, **95**, 191-198. <https://doi.org/10.1007/s001220050547>
- [16] Bamberg, J.B. and del Rio, A.H. (2014) Selection and Validation of an AFLP Marker Core Collection for the Wild Potato *Solanum microdontum*. *American Journal of Potato Research*, **91**, 368-375. <https://doi.org/10.1007/s12230-013-9357-5>
- [17] del Rio, A.H., Bamberg, J.B., Huaman, Z., Salas, A. and Vega, S.E. (2001) Association of Eco-Geographical Variables and RAPD Marker Variation in Wild Potato Populations of the USA. *Crop Science*, **41**, 870-878. <https://doi.org/10.2135/cropsci2001.413870x>
- [18] Jansky, S.H., Simon, R. and Spooner, D.M. (2008) A Test of Taxonomic Predictivity: Resistance to Early Blight in Wild Relatives of Cultivated Potato. *Phytopathology*, **98**, 680-687. <https://doi.org/10.1094/PHYTO-98-6-0680>
- [19] Bamberg, J.B. and del Rio, A.H. (2011) Diversity Relationships among Wild Potato Collections from Seven "Sky Island" Mountain Ranges in the Southwest USA. *American Journal of Potato Research*, **88**, 493-499. <https://doi.org/10.1007/s12230-011-9216-1>
- [20] Bamberg, J.B., del Rio, A.H., Huaman, Z., Vega, S.E., Martin, M., Salas, A., Pavék, J., Kiru, S., Fernandez, C. and Spooner, D.M. (2003) A Decade of Collecting and Research on Wild Potatoes of the Southwest USA. *American Journal of Potato Research*, **80**, 159-172. <https://doi.org/10.1007/BF02855688>
- [21] Bamberg, J.B. and del Rio, A.H. (2004) Genetic Heterogeneity Estimated by RAPD

- Polymorphism of Four Tuber-Bearing Potato Species Differing by Breeding Structure. *American Journal of Potato Research*, **81**, 377-383. <https://doi.org/10.1007/BF02870198>
- [22] Peakall, R. and Smouse, P.E. (2012) GenALEX 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—An Update. *Bioinformatics*, **28**, 2537-2539. <https://doi.org/10.1093/bioinformatics/bts460>
- [23] JMP® Pro, Version 15.0.0 (2019) SAS Institute Inc., Cary, NC, 1989-2019.
- [24] Foll, M. and Gaggiotti, O. (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, **180**, 977-993. <https://doi.org/10.1534/genetics.108.092221>
- [25] del Rio, A.H. and Bamberg, J.B. (2002) Lack of Association between Genetic and Geographic Origin Characteristics for the Wild Potato *Solanum sucrense* Hawkes. *American Journal of Potato Research*, **79**, 335-338. <https://doi.org/10.1007/BF02870170>
- [26] Caballero, A., García-Pereira, M.J. and Quesada, H. (2013) Genomic Distribution of AFLP Markers Relative to Gene Locations for Different Eukaryotic Species. *BMC Genomics*, **14**, 528. <https://doi.org/10.1186/1471-2164-14-528>
- [27] Storz, J.F. (2005) Using Genome Scans of DNA Polymorphism to Infer Adaptive Population Divergence. *Molecular Ecology*, **14**, 671-688. <https://doi.org/10.1111/j.1365-294X.2005.02437.x>
- [28] Körner, C. (2007) The Use of “Altitude” in Ecological Research. *Trends in Ecology & Evolution*, **22**, 569-574. <https://doi.org/10.1016/j.tree.2007.09.006>
- [29] Bradshaw, W.E. and Holzapfel, C.M. (2006) Evolutionary Response to Rapid Climate Change. *Science*, **312**, 1477-1478. <https://doi.org/10.1126/science.1127000>

## Abbreviations

AFLP = Amplified Fragment Length Polymorphism;

CAT = Santa Catalina;

RIT = Santa Rita;

HUA = Huachuca;

CHI = Chiricahua;

PIN = Pinaleño;

RIN = Rincon;

USPG = US Potato Genebank.