Comparison of Reputed Duplicate Populations in the Russian and US Potato Genebanks Using RAPD Markers

J. B. Bamberg[1]*, S. D. Kiru[2], and A. H. del Rio[3]

[1]USDA/Agricultural Research Service, Vegetable Crops Research Unit, Inter-Regional Potato Introduction Station, 4312 Hwy. 42, Sturgeon Bay, WI 54235.
[2]N. I. Vavilov Institute Potato Genebank, 42 Hertzen Street, 190000 St. Petersburg, Russia.
[3]Dept. of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison WI, 53706.
*Corresponding author: Tel: 920-743-5406, Fax 920-743-1080, email: nr6jb@ars-grin.gov
Reported use of brand name products does not imply an endorsement by USDA.

ABSTRACT

The Association of Potato Intergenebank Collaborators (APIC) produced a global inventory of wild potato genetic resources that is available on the Internet (www.potgenebank.org/ipd). This database shows that, in many cases, several genebanks have samples of the progeny from a single original germplasm collection. The assumption has been that these samples are genetically equivalent, so all the characterization and evaluation data gathered on a seedlot from one genebank can be applied to all the other "duplicate" seedlots in other genebanks. This assumption was tested by comparing 25 pairs of reputed duplicates in the VIR (St. Petersburg, Russia) and US (Sturgeon Bay, USA) potato genebanks using RAPDs. In 23 of 25 populations, reputed duplicates among genebanks had significantly less similarity than replicate samples taken from a single population. The average genetic similarity of reputed duplicates was 93%, and the lowest was 81%. Thus, users of germplasm should be aware that reputed duplicate populations from these genebanks may not be genetically identical.

INTRODUCTION

The potato combines a status as a major world crop with high input costs and susceptibility to diseases and pests, high quality demands, and an unusually wide array of closely related wild species that can be crossed with relative ease to the cultivated forms (Hanneman 1989). This makes the use of exotic germplasm for genetic improvement of the crop very attractive. World potato genebanks have the responsibility of collecting, classifying, preserving, evaluating, and distributing these resources. Since 1990, these genebanks have been participating in a formal network to exchange information and techniques, and to work on problems of mutual interest. A comprehensive database of passport and evaluation data has been synthesized for wild potato species. By matching collection numbers, it is evident that in many cases, individual germplasm populations are duplicated in more than one genebank (Huaman et al. 2000). It seems reasonable to assume that evaluation and characterization data collected at one genebank can be attributed to the matching population at another genebank. However, differences in sampling of the original population when it was divided and sent to genebanks and subsequent differences in seed multiplication technique introduce the possibility that reputed duplicates at different genebanks have diverged genetically. Human error in the form of mislabeling, mixing, or mispollinating is also possible. This is exemplified by the study of Steiner et al. (1997) that revealed genetic differences in reputed duplicate oat collections maintained at several sites. The present study was initiated to measure the similarity of some of the presumed duplicate potato populations held both at the Vavilov Institute

Accepted for publication April 19, 2001.

Abbreviations:
VIR: Vavilov Institute Potato Genebank
GS: Genetic similarity

potato collection, VIR, St. Petersburg, Russia, and the US Potato Collection, Sturgeon Bay, WI, USA. To the authors' knowledge, this is the first study comparing the genetic similarity of reputed duplicates in sister potato genebanks.

MATERIALS AND METHODS

Materials

Duplicate populations were identified in the VIR and US collections. Of these, 35 populations representing seven species were selected based on availability of seeds and the fact that they had undergone seed multiplication at least once at each site after being received as samples of the original population (Kiru and Sdvizhkova 1999). Thus, each pair of samples tested was derived from seed increase progeny of different samples from the same original population. The identities of the 25 populations with sufficient germination for RAPD testing are given in Table 1. Most of these are primitive cultivated species (*tuberosum* and *andigena* = 4x, *phureja* and *stenotum* = 2x), the remaining species are wild (*demissum* and *guerreroense* = 6x, *chacoense* = 2x). Complete information on these stocks can be accessed through on-line national and international databases linked to the US Potato Genebank homepage, http://www.ars-grin.gov/nr6.

Handling and Observation of Seedlings

Lots of 100 seeds each were sent from VIR to the US genebank at Sturgeon Bay, Wisconsin. Each of these and the corresponding US seedlot was sown in two replicates of 50 seeds each. All materials were handled identically. Seeds were submersed in 2,000 ppm GA_3 for 24 hr and dispersed over potting medium in 10-cm clay pots, then covered with a thin layer of Vermiculite. Before transplanting, the pairs were visually assessed for differences in germination, size of leaves, height, and presence of albinos.

TABLE 1—*Comparison of reputed duplicate samples at VIR and US potato genebanks using RAPD markers.*

Collector's Number	SPECIES (*Solanum...*)	VIR Code (VIR)	US Code (PI)	Year Split[1]	VIR Seedlot Year	US Seedlot Year	RAPD Bands	GS[2] within US	GS within VIR	GS between genebanks*
FCE 104	*chacoense*	21845	197760	1989	1993	1994	136	0.993	1.000	0.916
OKA 5341	*chacoense*	21323	472819	1987	1992	1996	140	0.993	0.993	0.933
COR 14283	*demissum*	19075	161366	1987	1992	1996	140	1.000	1.000	0.993 ns
COR 14342A	*guerreroense*	21404	161727	1987	1991	1992	117	1.000	1.000	0.940
CCC 122	*phureja*	15246	225674	1977	1991	1996	156	1.000	1.000	0.974
CCC 131	*phureja*	15247	225675	1965	1996	1989	127	1.000	1.000	0.960
CCC 143	*phureja*	8361	225681	1969	1992	1990	160	1.000	1.000	0.969
CCC 256	*phureja*	5949	225689	1965	1984	1966	125	1.000	0.984	0.912
CCC 130	*phureja*	16579	225695	1969	1997	1975	158	1.000	1.000	0.968
GND 63	*stenotomum*	15286	234015	1977	1992	1990	145	1.000	1.000	0.966
CPC 1673x	*andigena*	4712	205623	1962	1973	1994	128	0.992	1.000	0.883
SMI 504	*andigena*	5801	214442	1957	1994	1994	135	0.993	1.000	0.941
CCC 61	*andigena*	5806	225633	1962	1990	1992	142	1.000	0.993	0.884
CPC 1464	*andigena*	4715	230457	1962	1993	1994	136	1.000	1.000	0.904
OCH 1226	*andigena*	5820	230499	1962	1990	1987	123	1.000	1.000	0.935
GND 61	*andigena*	5836	233989	1962	1990	1994	143	1.000	1.000	0.930
GRA 97-2	*andigena*	5847	243343	1962	1984	1991	104	1.000	0.990	0.808
CCC 4	*andigena*	19366	243361	1982	1992	1991	143	1.000	1.000	1.000 ns
CCC 44	*andigena*	18945	243372	1981	1989	1994	144	1.000	1.000	0.972
CCC 114	*andigena*	19367	243384	1962	1990	1994	140	1.000	1.000	0.929
CCC 210	*andigena*	5885	243409	1962	1997	1994	141	1.000	1.000	0.930
CCC 320	*andigena*	17165	243429	1978	1984	1994	128	1.000	0.992	0.922
CCC 425	*andigena*	5912	243438	1962	1997	1986	144	0.993	1.000	0.947
COR C.132	*tuberosum*	10487	245935	1971	1986	1997	120	0.992	1.000	0.858
COR C.133	*tuberosum*	10488	245937	1971	1986	1978	152	1.000	1.000	0.922
						Average:	137	0.998	0.998	0.932

[1]Year samples of the original seedlot were sent to each genebank
[2]GS = genetic similarity
*All <= 0.975 are significant at $p \leq 0.05$. Those not significant are marked "ns".

When seedlings were 3 to 5 cm tall, 27 of each replicate were transplanted to peat pots. Leaf tissue was sampled from each plant and bulked for DNA extraction.

RAPD Marker Technique

DNA was isolated from bulked fresh leaf tissue according to a procedure modified from that described in Williams et al. (1994). PCR amplifications were performed in 15-μL reaction volumes as described in del Rio et al. (1997). Comparisons were based on an average of 137 unique bands generated with 24 primers. All clear bands generated were used to compare replicates and intergenebank samples within a given population. The band or blank status of each DNA bulk was considered to be comparable to the presence or absence of a dominant allele at random loci. The statistic generated was GS, calculated as the percentage of loci with matching band status. For each set of reputed duplicates, replicate GS was calculated between each of the two pairs of replicates and between samples from different genebanks.

Statistical Analysis

The assumption that the distribution of observed replicate GS fit the binomial distribution ($P = 0.998$, n=137) was tested by Chi^2.

An individual observation must have a frequency of no more than about 0.002 in order to *not* occur at least once in a sample of 25 with $P \leq 0.05$. Thus, we calculated the GS level expected to occur at frequency ≤ 0.002 in the observed replicate GS distribution ($P = 0.998$, n = 137) using the standard binomial formula. This was set as the $p \leq 0.05$ critical (statistically significant) level for any single observation of GS between genebank samples.

RESULTS

Characteristics of Seedlings

Seedlings of VIR origin tended to have larger leaves, be taller, and contain albinos. Eight of the VIR populations did not germinate at all. These differences were not quantified, but because replicates were always very similar, they probably represent real seedlot effects. The overall effect of the source from which seedlots originated (VIR or US) could be measured by Chi^2 tests against an expectation that each genebank's seedlot would be judged superior an equal number of times by chance if no real differences existed. In this way, the superiority of VIR seedlots for leaf size and superiority of US seedlots for >0% germination were significant at $P \leq 0.05$.

Because of poor germination in either of the seedlots, only 25 of the originally planted 35 pairs could be adequately compared using RAPDs.

RAPD Comparisons and Their Statistical Significance

Table 1 shows the GS among replicates and between genebank samples for each population. GS of replicates averaged 99.8%. This similarity of samples from the same seedlot shows that RAPD profiles were very consistent, providing very high resolution of treatments.

It was found that the distribution of GS within replicates was very similar to expectations for a binomial distribution where $P = 0.998$ (Chi^2 *probability* = 92%). This good fit suggests that variation of replicate GS was well explained by random effects, i.e., there is no reason to suspect that detection of certain bands was more or less efficient in different populations.

A GS of 0.975 or less has $P \leq 0.05$ of occurring in a random sample from the observed replicate GS distribution. Thus, any GS between reputed duplicate genebank samples ≤ 0.975 was considered statistically significant. All but two of the 25 comparisons of reputed duplicates from different genebanks had GS this low or lower (Table 1).

DISCUSSION

Visual assessments of seedlings before transplanting suggested differences in the physiological status of these duplicate populations. The observation of albinos in only the VIR sample of population 33 is obviously a genetic difference, but probably not of the type that would be detected by RAPDs in this experiment. It is likely that the US sample also contains the recessive albino allele, but perhaps at a lower frequency, such that none of the observed segregants were nulliplex. Similarly, RAPDs used here on bulks did not detect possible changes in band frequencies, except when a RAPD band was completely lost. Thus, RAPDs detected only extreme changes among the genebanks' samples in the form of alleles lost from one of the paired populations.

The observed distribution of 50 replicate GS fits a binomial distribution for $P = 0.998$, n = 137 quite well. But because binomial distribution variances are not symetrical around this estimated hypothetical population mean, the best estimate of the true population P of replicate GS is slightly lower than 0.998. This consideration slightly lowers the critical limit for signifi-

cance, but not enough to change declarations of significance of any of the GS of pairs of duplicate genebank samples.

These populations are expected to be particularly vulnerable to genetic changes. Most of the populations tested are *Solanum tuberosum* ssp. *andigena*, a taxon whose populations were found by Hosaka and Hanneman (1991) to exhibit particularly high seed protein variability. This implies genetic heterogeneity within populations, the basis of vulnerability to genetic drift.

One objection to using bulk DNA samples is based on the contention that bands present in a small proportion of plants in the bulk will not be detected (Divaret et al. 1999; Gilbert et al. 1999). However, our previous work using very heterogeneous species indicates that even bands present in only one plant in a 24-plant bulk are nearly always detected (del Rio and Bamberg 1998). Others have also reported efficient detection in bulks (Tinker et al. 1994; Williams et al. 1993).

There are several reasons why the ability to detect low frequency bands is not an unqualified advantage. Such bands are very prone to sampling error unless a large total number of plants are sampled. Thus, ironically, more sensitive detection of low-frequency bands may result in an overall loss of resolution. Also, concern for detecting bands at frequency lower than 1/20 seems inconsistent with the fact that no more than 20 plants are used for seed multiplication at these genebanks. Also, if bulking reduced minor band detection, the polymorphic bands analyzed here would tend to be the ones at higher frequency in the populations and *less* susceptible to loss. Thus, the differences observed reflect the detection of more extreme changes than if the use of truly random polymorphic markers had been ensured. Finally, the reader should bear in mind that the differences detected here are with respect to random DNA markers, not traits of practical value. One might argue that differences observed in random polymorphic DNA overestimate the vulnerability of most *useful* traits since such traits tend not to be conferred by alleles at low frequencies. This is a reasonable assumption to the extent that traits conferring natural fitness for the plant also coincide with the desires of human beings with respect to cultivation (which they sometimes do, e.g., disease resistance, fertility).

Records were not available as to the number of serial increases separating the tested lots among genebanks, so a possible relationship between this and degree of differentiation could not be tested. No tests of differentiation between generations *within* a genebank were made in this experiment for comparison. However, previous work has shown that similarity between seed increase generations averages about 96% when only polymorphic bands are considered (del Rio et al. 1997), and that about two thirds of total bands in these types of materials can be expected to be monomorphic (Bamberg et al. 1999). So the average GS of generations within the same genebank would be estimated at nearly 99% (not significant) compared to the average GS of 93% detected here for populations in different genebanks.

Although the GS of duplicates had a relatively high average of >93%, most of the comparisons of reputed duplicate samples held in the VIR and US potato genebanks exhibited a statistically significant degree of genetic differentiation (Table 1). The cause and specific practical impact of this is beyond the scope of this experiment. However, these results serve to apprise breeders, curators, and other potato germplasm researchers of the fact that samples of reputed duplicate populations from these genebanks may not be genetically identical.

ACKNOWLEDGMENTS

This project was made possible, in part, by the Cornell-Eastern Europe-Mexico (CEEM) International Collaborative Project in Potato Late Blight Control at Cornell University, and the Wallace Genetic Foundation. The authors also express their thanks to the University of Wisconsin Peninsular Agricultural Research Station program and staff for their cooperation.

LITERATURE CITED

Bamberg, J., C. Singsit, A.H. del Rio, and E.B. Radcliffe. 1999. RAPD analysis of genetic diversity in *Solanum* populations to predict the need for fine screening. Am J Potato Res 77:275-278.

del Rio, A.H., and J.B. Bamberg. 1998. Effects of sampling size and RAPD marker heterogeneity on the estimation of genetic relationships. Am J Potato Res 75:275.

del Rio A.H., J.B. Bamberg, and Z. Huaman. 1997. Assessing changes in the genetic diversity of potato genebanks. 1. Effects of seed increase. Theor Appl Genet 95:191-198.

Divaret, I., E. Margale, and G. Thomas. 1999. RAPD markers on seed bulks efficiently assess the genetic diversity of a *Brassica oleraceae* L. collection. Theor Appl Genet 98:1029-1035.

Gilbert, J.E., R.V. Lewis, M.J. Wilkinson, and P.D.S. Caligary. 1999. Developing an appropriate strategy to assess genetic variability in plants germplasm collections. Theor Appl Genet 98:1125-1131.

Hanneman Jr. , R.E. 1989. The potato germplasm resource. Am Potato J 66:665-667.

Hosaka, K., and R.E. Hanneman Jr. 1991. Seed protein variation within populations of wild and cultivated potato species and inbred *Solanum chacoense*. Potato Res 34:419-428.

Huaman, Z., R. Hoekstra, and J. B. Bamberg. 2000. The Inter-genebank Potato Database and the dimensions of available wild potato germplasm. Am J Potato Res 77:353-362.

Kiru, S.D., and V.P. Sdvizhkova. 1999. Katalog mirovoi collectzii VIR N 707 - Kartofel - Culturny vid *Solanum andigenum* Juz. et Buk. Vavilov Institute (VIR, St. Petersburg, Russia) publication No. 707. 22 pp.

Steiner, A.M., P. Ruckenbauer, and E. Goecke. 1997. Maintenance in genebanks, a case study: Contaminations observed in the Nürnberg oats of 1831. Genet Resources Crop Evol 44:533-538.

Tinker N.A., D.E. Mather, and M.G. Fortin. 1994. Pooled DNA for linkage analysis: practical and statistical considerations. Genome 37:999-1004.

Williams, C.E., and P.C. Ronald. 1994. PCR template-DNA isolated quickly from monocot and dicot leaves without tissue homogenization. Nucleic Acids Res 22:1917-1918.

Williams, G.K., R.S. Reiter, R.M. Young, and P.A. Scolnick. 1993. Genetic mapping of mutations using phenotypic tools and mapped RAPD markers. Nucleic Acids Res 21:2697-2702.