Am. J. Potato Res. (2015) 92:276–283
DOI 10.1007/s12230-014-9428-2

Assessing SNPs Versus RAPDs for Predicting Heterogeneity and Screening Efficiency in Wild Potato (*Solanum*) Species

John Bamberg · Alfonso del Rio · Joseph Coombs · David Douches

Published online: 17 January 2015

Abstract Knowing how genetic diversity is partitioned among and within wild potato species populations is important for efficient sampling for collection, preservation and evaluation. We sought to evaluate the effectiveness of SNPs for assessing germplasm by using the exact set of four model species previously evaluated by RAPDs. To avoid large numbers of SNP samples, population bulks of 25 plants were used, and original RAPD data was adjusted to match SNP genotype data. It was noted that especially for SNPs, it was necessary to examine only loci polymorphic within species to get a realistic view of genetic partitioning within species. This resulted in only a few hundred useful loci for some species. When considering among-population versus within-population partitioning of diversity, both SNPs and RAPDs distinguished the species as expected according to their known breeding system. Primitive wild species were confirmed as very heterogeneous within their populations. Both SNP and RAPD markers can be used to help breeders and genebank managers understand patterns and use in potato germplasm diversity.

Resumen Conocer como la diversidad genética es particionada dentro y entre poblaciones de especies silvestres de papa es importante para lograr muestreos eficientes en colección, preservación y evaluación. Nuestro trabajo fue diseñado para evaluar la efectividad de SNPs en el análisis de germoplasma usando el mismo grupo de especies silvestres usados como modelos para evaluar RAPDs en un trabajo previo. Para evitar el uso de un gran número de muestras individuales en la generación de SNPs, en este trabajo cada muestra representó una composición de 25 plantas y los datos originales previos usando RAPDs tuvieron que ajustarse para coincidir con los genotipos producidos con SNPs. Se observó que en SNPs, especialmente, fue necesario examinar únicamente loci polimórficos dentro de cada especie para tener una visión clara de la partición de diversidad genética dentro de especies. Como resultado solo unos cuantos cientos de loci fueron útiles para analizar algunas especies. Cuando la partición de diversidad genética entre poblaciones fue comparada con la observada dentro de poblaciones, tanto SNPs como RAPDs fueron capaces de distinguir las especies como era esperado de acuerdo a sus sistemas reproductivos conocidos. Las especies silvestres primitivas confirmaron ser muy heterogéneas dentro de sus poblaciones. Ambos marcadores, SNPs y RAPDs, pueden ser usados para ayudar a los mejoradores, y a quienes manejan los bancos de germoplasma, a entender los patrones de diversidad en germoplasma de papa y su uso.

Keywords RAPD · SNP · Genetic diversity · Wild species

Abbreviations

RAPD	Random Amplified Polymorphic DNA
SNP	Single Nucleotide Polymorphisms

Introduction

Impact of Heterogeneity on Germplasm use The partitioning of genetic diversity within exotic potato germplasm in the

J. Bamberg (✉)
USDA/Agricultural Research Service, US Potato Genebank,
4312 Hwy 42, Sturgeon Bay, WI 54235, USA
e-mail: John.Bamberg@ars.usda.gov

A. del Rio
US Potato Genebank, Department of Horticulture,
University of Wisconsin, Madison, WI 53706, USA

J. Coombs · D. Douches
Department of Plant, Soil and Microbial Sciences,
Michigan State University, East Lansing, MI 8824, USA

genebank dictates how one would predict the most efficient approaches to germplasm use. Specifically, if the diversity is primarily among populations, we maximize efficiency of sampling if we collect, preserve and evaluate just a minimum number of individuals from a maximum number of populations. The converse is true if most of the diversity is within populations (Bamberg et al. 2000). If most of the diversity is within populations, "fine screening" among individuals within populations is needed (Douches et al. 2001). With respect to preservation in the genebank, within-population heterogeneity means each population contains more diversity, but that diversity is more susceptible to loss if techniques for sampling the population are not optimized.

Previous RAPD Experiment RAPD data was available to compare with new SNP results. The authors had previously assessed heterogeneity using RAPDs to examine 4 species, 8 populations each, 24 individuals each, with about 50 polymorphic RAPD loci (Bamberg and del Rio 2004). All the RAPD data mentioned in this paper refers to that 2004 study. Two of the species used in the 2004 study have now been combined with related taxa: Thus the official current taxonomic status of *Solanum fendleri* is *S. stoloniferum*, and *S. sucrense* is *S. brevicaule*. However, we refer to the materials used here by their original names *Solanum fendleri*, *jamesii*, *sucrense*, and *verrucosum* for consistency and use the standard species abbreviations fen, jam, scr, and ver (Huaman and Ross 1985). The species were chosen to model unrelated taxa of differing ploidies and breeding systems. However, Cai et al. (2012) provided evidence that fen might be a combination of ver and jam or their progenitors. The species scr has long been recognized as a relative of *tuberosum* (Hawkes 1990; Rodríguez et al. 2010).

In the previous RAPD study, genetic heterogeneity (GH) was calculated, which is the estimated number of heterozygous individuals within a population inferred from the distribution of multiple individual RAPD banded plants in a population. The different species partitioned the observed diversity as expected according to their ploidy, breeding system, and empirical evidence from other studies: outcrossing diploid species jam and tetraploid polysomic species scr had high GH, while self-compatible inbreeding disomic tetraploid species fen and diploid self-compatible species ver had low GH. Furthermore, this RAPD analysis also showed that marker-based estimates of relatedness are strongly biased by within-population heterogeneity if only one or a few plants are used to represent the population.

RAPDs Versus Single Nucleotide Polymorphisms (SNPs) RAPDs are assumed to be random and neutral dominant genetic markers. While inexpensive and technically easy, RAPDs are generally considered an obsolete marker with the advent of SSRs and SNPs. They have been particularly criticized with the claim of poor reproducibility (Spooner et al. 2005). However, other empirical reports show consistent error rates of<½% may be obtained (Bamberg et al. 2001). On the other hand, a SNP locus is codominant, genotyping for the wild allele and, potentially, a second mutant allele. They are mapped mutants in functional genes of *S. tuberosum*, so are not random and may not be neutral. Using SNPs from the Infinium 8303 Potato SNP array to measure diversity will introduce a cultivated potato "ascertainment bias". The concern of this bias is that populations less related to the germplasm from which SNPs were developed necessarily appear less heterogeneous, even potentially reversing the actual order of diversity of two populations (Lachance and Tishkoff 2013). SNPs have the potential for obtaining very abundant data points from the 8303 SNP array in potato that are mapped to a specific physical location in the potato genome. Hirsch et al. (2013) used these SNPs to evaluate 12 wild species and concluded they have far less heterozygosity than cultivated tetraploid potato.

Only a few of the approximately 100 potato species are cultivated, and the remaining wild ones are nearly always collected, preserved, and evaluated as botanical seed populations. SNPs can offer a genome-wide query of nucleotide diversity, but SNPs are too expensive to evaluate populations as many separate individuals as was done in the 2004 experiment. Other researchers have recommended use of a single plant to represent an entire population (Jacobs et al. 2011), but the 2004 RAPD study showed clearly that doing this introduces a great deal of error, and biases resolution strongly in favor of uniform inbreeding species. For the proposed SNP query we elected to use a balanced bulk of DNA from 25 individuals (fully expanded terminal leaflets of seedlings about 6 weeks old) for each population.

Selection of Loci Loci fixed in all populations, whether SNPs or RAPDs, do not provide practical information about within-species heterogeneity. Thus, in the 2004 RAPD study, we recognized that loci of practical interest within a species should be restricted to those that demonstrate the capacity to vary (i.e., are polymorphic).

Special Consideration for Tetraploids Known disomic tetraploids like fen have restricted recombination between their two separate genomes. Thus, a polymorphic SNP genotype does not necessarily mean detectable segregation. Consider, for example, a hypothetical population in which all plants have inbred genotype (AA BB) and a SNP call of AB. For this situation, all plants would appear banded when assessed by dominant RAPDs, so this locus would be eliminated from a RAPD analysis. Thus, when comparing RAPD and SNP data for fen, it was anticipated that SNPs might detect much more within-population variation than could be resolved with RAPDs. Furthermore,

tetrasomic tetraploids like scr could have up to four alleles at each locus, but SNPs and RAPDs are designed to detect only two variants per locus. Thus, in terms of real diversity for breeding, both of these markers might underestimate the value of such tetraploids as compared to diploids.

Materials and Methods

We sought to use a genome-wide set of SNPs to assess partitioning of heterogeneity of model potato species, and to do so particularly by comparing results to a previous RAPD experiment that used identical materials (Bamberg and del Rio 2004). DNA bulks of 25 individuals of the populations listed in Table 1 were examined. The populations randomly selected for replication as a second bulk of different seedlings are underlined, but all calculations of heterogeneity were done only on the first (randomly-chosen) replicate. Full details of origin and evaluation data for these populations are available in GRIN (Germplasm Resources Information Network, http://www.ars-grin.gov/npgs).

The objective of this study was to compare Infinium 8303 Potato SNP array measures of heterogeneity to those of the previous RAPD experiment on the same materials. This comparison was made by examining the proportion of SNP loci that are polymorphic in the context of loci that demonstrated the capacity to be polymorphic.

Technical Error We assessed technical error as the percent of loci with mismatched genotypes among replicate bulks (e.g., AA in one rep, AB in the other), with those replicated bulk populations identified with an underlined Plant Inventory (PI) number in Table 1.

Monomorphic Loci Among all Species The proportion of SNP loci that exhibited only one type of allele across all populations of all four species was determined.

Heterogeneity of a Species Then, of those loci polymorphic considering all species, we measured heterogeneity within species as the proportion of loci polymorphic within the species in question. That is, for example, what proportion of loci have (AB) genotype in at least one fen population?

Heterogeneity of a Population Then, of loci polymorphic within species, we assessed the proportion of polymorphic loci within a population. That is, for example, of all the loci potentially polymorphic within fen what proportion have heterogeneous genotype (AB) within the fen population in question? This is "Index of Variation" (IV) shown in Table 1.

Genotype Frequencies of Polymorphic Loci Within a Species Two further measures of heterogeneity within polymorphic loci within species were made. Of loci polymorphic within species, what proportion of those loci have both alleles fixed within the individual populations? That is, for example, how many polymorphic fen loci have at least one popluation (AA) and at least one population (BB)? The only other possible condition for loci polymporphic in a species is that some fen populations are, for example, (AA) and some are (AB). For those loci, we assessed the number of (AB) popluations.

SNP Generation, Filtering, and "Genotype" Evaluation DNA was purified from a composite sample of 25 young seedling leaves from each population using Qiagen DNeasy Plant Mini kits (Qiagen, Germantown, MD) and assayed at Michigan State University on the Infinium 8303 Potato SNP Array with the Illumina iScan Reader (Illumina, San Diego, CA). The array surveys 8303 biallelic SNP loci designed to transcribed sequences (Felcher et al. 2012). SNP genotypes were determined in GenomeStudio (Illumina, San Diego, CA) under a diploid model supporting biallelic calls, such that non-homozygous calls in polyploid accessions were

Table 1 Materials tested and Results: Percent of loci in a population with two alleles, considering those loci that vary in the species=Index of Variation (IV). The populations with underlined PI numbers were replicated

fen	275157	275160	275161	275163	564028	564031	564039	564040	Average
RAPD	33 %	17 %	17 %	22 %	16 %	28 %	10 %	12 %	19 %
SNP	91 %	90 %	89 %	95 %	90 %	91 %	90 %	90 %	91 %
jam	275169	458423	458424	458425	458426	458427	564047		
RAPD	85 %	71 %	65 %	65 %	77 %	73 %	52 %		70 %
SNP	43 %	43 %	46 %	42 %	80 %	49 %	52 %		51 %
scr	473506	473532	498286	498300	498301	498302	498306	566798	
RAPD	63 %	50 %	75 %	69 %	84 %	72 %	69 %	50 %	66 %
SNP	56 %	43 %	61 %	69 %	57 %	56 %	66 %	45 %	57 %
ver	161173	275256	275258	275260	310966	365404	558485	570643	
RAPD	37 %	10 %	27 %	27 %	20 %	13 %	30 %	30 %	24 %
SNP	48 %	48 %	48 %	53 %	49 %	48 %	46 %	59 %	50 %

rated as simple "heterozygote" (AB) calls (Hirsch et al. 2013). The diploid SNP genotypes were initially called using the SolCAP custom three cluster calling file (Felcher et al. 2012) and then manually curated to correctly call heterozygous and no-call SNPs due to no signal. Loci were filtered to remove non-informative SNPs and SNPs that aligned to multiple regions of the International Potato Genome Sequencing Consortium (PGSC) version 4.03 pseudomolecules. Following filtering, 5590 quality SNP loci were used for this analysis.

Meaning of SNP Genotype Notation Used for Bulks For simplicity, we use the letters A and B to represent the genotypes of the two possible SNP alleles per locus, regardless of the actual SNP allele A, C, G, or T. A diploid bulk with a one-allele SNP "genotype" (AA) at a certain locus is assumed to be composed of all AA plants. However, a bulk DNA sample with a two-allele SNP call (AB) indicates the presence of both alleles, although a diploid population, for example, could be composed of various combinations of AA, AB, and BB individuals.

Results and Discussion

Filtering resulted in a useful set of 5590 SNP loci over all four species, or about 67 % of the potential 8303 on the *tuberosum* array.

Technical Error Previous tests of RAPD technical error rate using replicated bulks of 25 different seedlings resulted in <½% blank-band mismatches for all species (Bamberg et al. 2001). For SNPs, species jam, fen, and ver all had <1 % error, but scr had 2.3 % and 6.1 % mismatches in population 473506, and 498300, respectively. SNP technical error is usually very low in genotyping identical DNA samples, or when only one allele is present, or when two alleles are present in a ratio not more unbalanced than 1:3 (the limit in a tetraploid). It does not seem plausible that there are so many loci with the very rare alleles needed to cause two 25-seedling bulks of a tetraploid scr to truly have as great as 6 % different SNP genotypes, when this was not observed with RAPDs. More experimentation is needed to explain the unexpected high replicate differences in scr.

Monomorphic Loci Among all Species A total of 2425 loci (43 %) had the same one-allele genotype (AA) over all populations of all species, thus were not informative. For RAPDs, one observes a higher percentage (60–80 %) of completely fixed bands (Bamberg et al. 2000). RAPD markers probably show more fixed bands because band presence is a dominant phenotype on a gel, hiding the presence of the non-band allele in heterozygotes. The 2004 RAPD data avoided this problem by genotyping many individual plants rather than bulking them. Indeed, inspecting that data shows that a high proportion of actual heterogeneous loci would have appeared uniformly banded, thus not informative, had RAPD bulks been used: fen=36 %, jam=65 %, scr=50 %, ver=37 %.

The SNPs used were selected from nucleotide mutants detected among six named cultivars (Hamilton et al. 2011), so the species *Solanum tuberosum* has the maximum potential SNP diversity (potentially having both A and B alleles for 100 % of loci). Thus, the frequency of two-allele loci in wild species can also be used to measure genetic similarity to cultivated tetraploid *S. tuberosum* in terms of common mutants. Pooling all four species tested, only 3165 loci, or about 57 % have both *tuberosum* SNP variants, suggesting a substantial differentiation of wild species from modern cultivars.

Heterogeneity of a Species Once loci monomorphic across all species are eliminated, we may examine the proportion of those loci polymorphic within species. Polymorphic loci show that wild species possesses both cultivar alleles (e.g., A and B), which is a measure of its relatedness to *tuberosum*, perhaps due to ancient descent or more modern intentional introgression through breeding. Two-allele loci were fen=21 %, jam=6 %, scr=55 %, ver=6 %. This is expected for the heterogeneous scr, which is reputed to be a close relative of *tuberosum* (Hawkes 1990). *Solanum stoloniferum*, now encompassing fen in GRIN taxonomy has been used for virus resistance breeding, and the relative *Solanum demissum* has been used extensively in late blight resistance breeding, perhaps explaining the evidence here of common alleles with the cultivar genepool. Inbreeding disomic tetraploid fen would allow fixed heterozygosity across its two genomes, protecting from inbreeding depression, drift, and loss of the mutant allele if it is associated with less fitness.

We expect the primitive species jam (tertiary gene pool), not in the pedigree of any cultivar, to share few of the mutant SNP alleles identified in *tuberosum*, and thus have mostly homozygous wild-type SNP alleles, despite the evidence that jam is, in fact, a quite heterozygous self incompatible outcrossing diploid (Bamberg et al. 2009). The lack of two-allele heterozygous SNP genotypes in ver is also expected according to its breeding system as a self compatible inbreeding diploid which has not appeared much in the pedigrees of cultivars.

A comparison of jam and scr SNP loci that are polymorphic or not polymorphic illustrates why relatedness to *tuberosum* and heterozygosity are confounded in these species and why meaningful results depend on examining only SNPs polymorphic within a species. In the context of only SNP loci polymorphic in scr, the average diversity (AB) appears to be much less for jam (4 %) than for scr (57 %). But in the context of only SNP loci polymorphic in jam, the average diversity (AB) of jam is high (51 %) and about the same as scr (57 %) (see Table 1).

After removing loci with missing data and those lacking variation (only one-allele, e.g., AA), SNPs provide up to 3078

markers for study, far fewer than the 8303 possible, but still many more than RAPDs provided in the 2004 study (Table 2).

Heterogeneity of a Population After eliminating loci with only a one-allele genotype (AA) in all populations within a species, an unbiased estimate of the average heterogeneity within an individual population can be calculated. In the 2004 RAPD study, testing individual plants made it possible to calculate Genetic Heterogeneity (GH), an estimate of actual *heterozygote* frequencies in the population for a given locus. For SNP population bulks, we could only detect the presence or absence of *heterogeneity*, so we standardized both RAPD and SNP data to compare "Index of Variation" (IV): the percentage of two-allele loci in a population for all loci that vary within the species. These comparisons are presented in Table 1.

Detected heterogeneity within species differs for SNPs versus RAPDs, particularly for fen and ver. For the self compatible disomic tetraploid fen, SNP average IV is so high because 88 % of SNP loci are uniformly (AB) in all populations. This is probably explained by fixed heterozygosity (every plant AB) in the population. Fixed heterozygosity has been observed previously in allopolyploid species using co-dominant isozyme markers (Douches unpublished data). Such loci would not be detected as variable with RAPDs so not considered in the RAPD analysis. The alternate explanation–that so many SNP loci are really segregating in all fen populations, does not seem plausible considering that this situation was not observed for even a single fen locus with RAPDs when individual plants within populations were genotyped.

For diploid self compatible species ver, the SNP average IV is also high due to many loci (39 %) having two alleles (AB) in every population. This is difficult to explain, since that is a higher proportion of SNP loci segregating in all populations than observed for either outcrossing species (diploid jam= 31 % or tetraploid scr=20 %). And such uniform heterozygosity was observed in only one RAPD locus of ver when individual plants within populations were genotyped. Could it be that many of these *tuberosum* SNPs loci are duplicated in the ver genome?

Genotype Frequencies of Polymorphic Loci Within a Species–Both Alleles Fixed After eliminating loci with only a one-allele genotype (AA) in all populations within a species, we would like to see how often a species fixes both alleles in separate populations. We do not know the inheritance of phenotypes linked to SNP diversity. But if any dominance exists, observing the maximum phenotypic variation depends on isolating each allele. Thus, for germplasm screening, if some populations are AA and some BB, for the locus in question, the effect of A and B can be differentiated by bulk population screening. For the interests of preserving diversity in the genebank, both possible alleles will surely be preserved as long as one AA and one BB population are maintained, even if crude population sampling techniques are used. On the other hand, completely losing unique alleles from the genebank may happen with the loss of a single population of this type. Figure 1 displays the proportions of loci within species that fix both alleles versus loci that do not.

Figure 1 illustrates a similar pattern for jam, scr, and ver regardless of whether assessed by SNPs or RAPDs. Few two-allele loci in jam and scr fix both alleles in pure form. Thus, to observe the maximum diversity of phenotypic expression, one may need to screen within populations. On the other hand, a substantial proportion of ver loci (32–47 %) have at least one population pure AA and at least one population pure BB. The tetraploid disomic self compatible species fen looks much different with RAPDs versus SNPs. Figure 1 shows 0 % of loci fixed for both SNP alleles, but this is an approximation of the only 2 loci of 1151 that had at least one of each possible population genotype (AA, AB, and BB). Apparently, double mutants or recombination between the two genomes that would allow pure BB plants or populations in fen are very rare, a conclusion that is in harmony with a previous study of a pigment mutant in fen (Bamberg et al. 2006). If segregation is happening at most within one genome in fen, actual tetraploid genotypes can be [AA AA] or [AA AB] or [AA BB], but never [BB BB]. However, with RAPDs, this still looks like full segregation to fixation of the B allele if B corresponds to the RAPD band, and A corresponds to the RAPD blank. Thus, the marker which best reflects phenotypic segregation in species like fen depends on dominance of the associated trait. If B is linked to a completely dominant trait, the significant phenotypic segregation predicted by RAPDs will be observed. But if B is linked to a completely recessive trait, its segregation in the context of a fixed dominant (A) allele in the other genome will lack the phenotypic evidence predicted by SNPs.

Genotype Frequencies of Polymorphic Loci Within a Species–Heterogeneous Populations When Both Alleles are not Fixed When populations in a species do not have both (AA) and (BB), but only (AA) or (AB) genotypes for a given locus, one must "fine screen" within (AB) populations to observe a pure (BB) individual. That is a more challenging task if (AB) populations are rare. So we are interested in the frequency of (AB) populations for loci for which there

Table 2 Number of useful SNP and RAPD loci (polymorphic among populations within species)

	fen	jam	scr	ver
RAPD	58	48	35	30
SNP	1151	342	3078	323

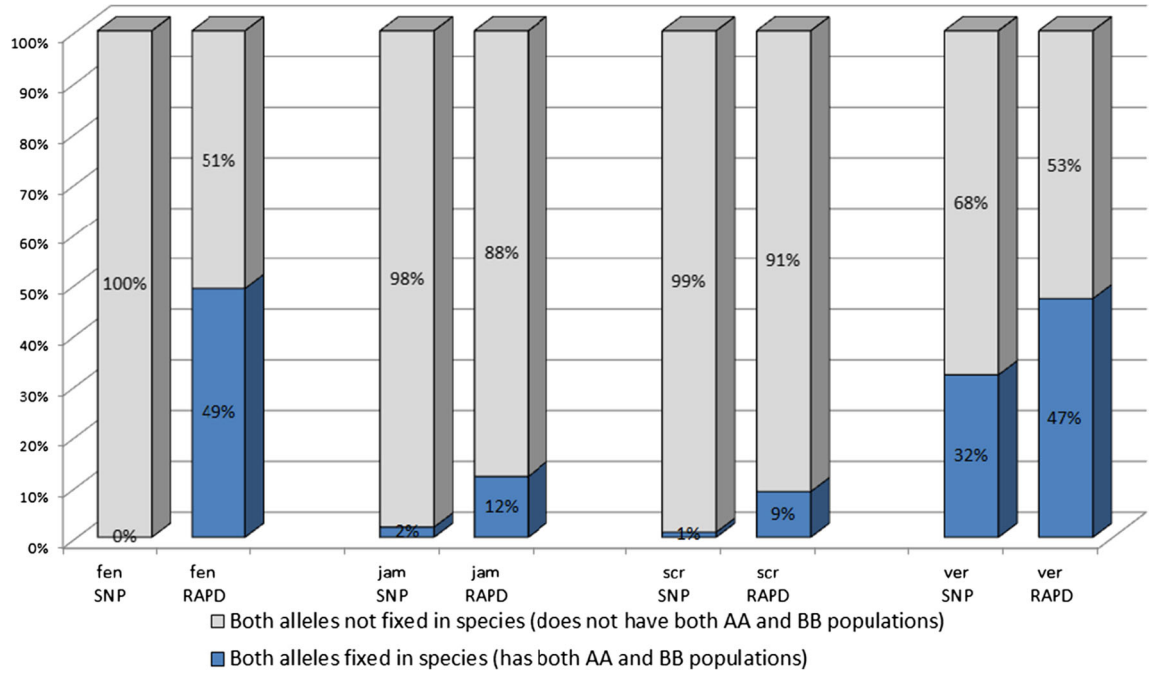

Fig. 1 Proportion of two-allele loci for which both alleles are fixed in a species: At least one population is AA and at least one population is BB

are no (BB) populations. Figures 2a-d show the frequency of populations with such unfixed minor (B) alleles within species for SNP and RAPD data.

For RAPDs, fen (Fig. 2a) and ver (Fig. 2d) show a similar and expected pattern for inbreeders. Unfixed alleles are relatively rare, usually occurring in only one or a few of the

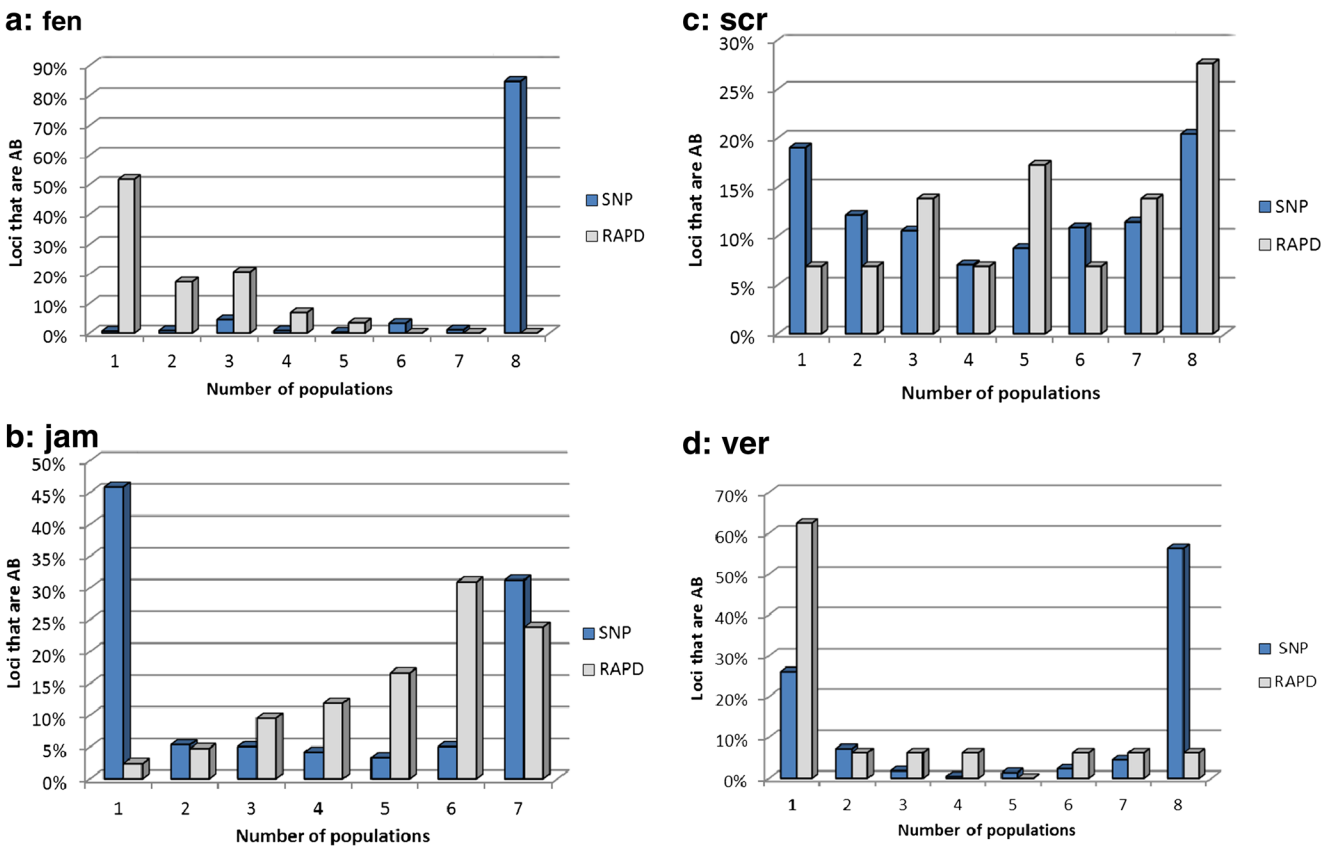

Fig. 2 a-d). Distribution of (AB) loci among populations within species

populations. On the other hand, outcrossing jam (Fig. 2b) and scr (Fig. 2c) often have their unfixed alleles in many or all of the populations. Thus, unfixed alleles are common, but easy to isolate by fine screening within most populations of jam and scr. In contrast, unfixed alleles are rare and difficult to isolate by fine screening within selfing species fen and ver.

The pattern of unfixed SNP alleles is most similar to that of RAPDs for scr (Fig. 2c).

SNP and RAPD distributions of unfixed alleles are different for fen (Fig. 2a), particularly as unfixed SNP alleles being present in all populations for a very high proportion (88 %) of loci, something almost never seen with RAPDs. The likely reason for this has already been explained: These SNP alleles actually are fixed as [AA BB] on one genome of this inbreeding disomic tetraploid. RAPD loci of this type are not considered because they are monomorphic, appearing as a band present in all individuals of all populations.

For both SNPs and RAPDs of jam (Fig. 2b), many unfixed minor alleles are present in all populations (31 and 24 % respectively). On the other hand, while unfixed RAPD minor alleles rarely occur in only one population (2 %), such SNP alleles commonly occur in only one population (46 %). Further investigation showed that for about 75 % of the time, this was attributable to PI 458426 (as suggested by its unusually high SNP IV in Table 2). So, except for this particular single population, SNPs as well as RAPDs predict that minor alleles would likely be isolated as (BB) with ease by fine screening within any random jam population. We have no explanation for why SNPs appear to distinguish PI 458426. It was not one of the replicated jam populations, so we cannot eliminate the possibility of a technical mistake like a mixing of tissue samples. This jam 458426 is known to have the same site of natural origin as fen. But natural introgression between the species is not known, and PI 458426 has typical jam phenotype.

For both SNPs and RAPDs of ver (Fig. 2d), many unfixed minor (B) alleles are present in only one populations (26 and 63 % respectively). On the other hand, while only 6 % of unfixed RAPD minor alleles occur in all populations, with SNPs, many (56 %) of such SNP alleles occur in all populations. It is not apparent why this should be observed in an inbreeding self compatible diploid species that has fixed about 1/3 of its other polymorphic SNP loci (Fig. 1).

In summary, distribution of unfixed minor alleles among populations exhibited unexpected SNP patterns only for jam and ver. However, we already knew that these two species provided by far the smallest numbers of usable SNP loci (Table 1), so they might be expected to be most susceptible to a bias due to the particular small sample of SNP loci examined.

Conclusions SNPs derived from *tuberosum* can be used to assess diversity in wild potato species and the SNP diversity was compared to previous RAPD data for species differing in breeding systems. At least 10 times as many useful SNP markers are easily obtained, but far fewer than the total in the 8303 array, since the relatedness of a species to *tuberosum* is inversely related to the number of useful polymorphic SNPs it will contain. SNPs agreed with RAPDs in concluding that outcrossers jam and scr both have high levels of heterozygosity partitioned mostly within populations. SNPs were also similar to RAPDs in revealing that diploid selfing ver is substantially inbred, having fixed both alleles within some population for many loci, partitioning diversity among populations. On the other hand, SNP results implied high levels of heterozygosity in the disomic polyploid fen which likely do not reflect phenotypic variation that can be selected in breeding or is vulnerable to loss in the genebank.

Acknowledgments The authors thank Mrs. Adele Douglass for technical assistance, Nick Garrity and Matt Zuehlke for SNP technical assistance and the Peninsular Agricultural Research Station program and staff for their support.

References

Bamberg, J.B., and A.H. del Rio. 2004. Genetic heterogeneity estimated by RAPD polymorphism of four tuber-bearing potato species differing by breeding system. *American Journal of Potato Research* 81: 377–383.

Bamberg, J., C. Singsit, A.H. del Rio, and E.B. Radcliffe. 2000. RAPD analysis of genetic diversity in *Solanum* populations to predict the need for fine screening. *American Journal of Potato Research* 77: 275–278.

Bamberg, J.B., S.D. Kiru, and A.H. del Rio. 2001. Comparison of reputed duplicate populations in the Russian and US potato genebanks using RAPD markers. *American Journal of Potato Research* 78: 365–369.

Bamberg, J.B., C. Fernandez, and A. del Rio. 2006. A new wild potato mutant in *Solanum stoloniferum* Schltdl. lacking purple pigment. *American Journal of Potato Research* 83: 437–445.

Bamberg, J.B., A.H. del Rio, and R. Moreyra. 2009. Genetic consequences of clonal versus seed sampling in model populations of two wild potato species indigenous to the USA. *American Journal of Potato Research* 86: 367–372.

Cai, D., F. Rodríguez, Y. Teng, C. Ané, M. Bonierbale, L.A. Mueller, and D.M. Spooner. 2012a. Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section Petota). *BMC Evolutionary Biololgy* 12: 70.

Douches, D.S., J.B. Bamberg, W. Kirk, K. Jastrzebski, B.A. Niemira, J. Coombs, D.A. Biognin, and K.J. Fletcher. 2001. Evaluation of wild *Solanum* species for resistance to the US-8 genotype of Phytophthora infestans utilizing a fine-screening technique. *American Journal of Potato Research* 78: 159–165.

Felcher, K.J., J.J. Coombs, A.N. Massa, C.N. Hansey, J.P. Hamilton, et al. 2012. Integration of Two diploid potato linkage maps with the potato genome sequence. *PloS One* 7(4): e36347. doi:10.1371/journal.pone.0036347.

Hamilton, J.P., C.N. Hansey, B.R. Whitty, K. Stoffel, A.N. Massa, and A. Van Deynze. 2011. Single nucleotide polymorphism

discovery in elite north american potato germplasm. *BMC Genomics* 12: 302.

Hawkes, J.G.H. 1990. *The potato: evolution, biodiversity and genetic resources*, 171. London: Belhaven Press.

Hirsch, C.N., C.D. Hirsch, K. Felcher, J. Coombs, D. Zarka, A. Van Deynze, W. De Jong, R.E. Veilleux, S. Jansky, P. Bethke, D.S. Douches, and C.R. Buell. 2013. Retrospective view of north american potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. *G3 (Bethesda)* 3: 1003–1013.

Huaman, Z., and R.W. Ross. 1985. Updated listing of potato species names, abbreviations and taxonomic status. *American Potato Journal* 629–641.

Jacobs, M.M.J., M.J.M. Smulders, R.G. van den Berg, and B. Vosman. 2011. What's in a name; genetic structure in *Solanum* section *Petota* studied using population genetic tools. *BMC Evolutionary Biology* 11: 42.

Lachance, J., and S.A. Tishkoff. 2013. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays* 35: 780–786.

Rodríguez, F., M. Ghislain, A.M. Clausen, S.H. Jansky, and D.M. Spooner. 2010. Hybrid origins of cultivated potatoes. *Theoretical and Applied Genetics* 121: 1187–1198.

Spooner, D.M., R. van Treuren, and M.C. de Vicente. 2005. Molecular markers for genebank management. *IPGRI Technical Bulletin No.* 10.