

Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification

David M. Spooner^{*†}, Jorge Núñez[‡], Guillermo Trujillo[‡], María del Rosario Herrera[‡], Frank Guzmán[‡], and Marc Ghislain[†]

^{*}Vegetable Crops Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Department of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706-1590; and [‡]Applied Biotechnology Laboratory, International Potato Center, Apartado 1558, Lima 12, Peru

Communicated by S. J. Peloquin, University of Wisconsin, Madison, WI, October 16, 2007 (received for review August 15, 2007)

Contrasting taxonomic treatments of potato landraces have continued over the last century, with the recognition of anywhere from 1 to 21 distinct Linnean species, or of Cultivar Groups within the single species *Solanum tuberosum*. We provide one of the largest molecular marker studies of any crop landraces to date, to include an extensive study of 742 landraces of all cultivated species (or Cultivar Groups) and 8 closely related wild species progenitors, with 50 nuclear simple sequence repeat (SSR) (also known as microsatellite) primer pairs and a plastid DNA deletion marker that distinguishes most lowland Chilean from upland Andean landraces. Neighbor-joining results highlight a tendency to separate three groups: (i) putative diploids, (ii) putative tetraploids, and (iii) the hybrid cultivated species *S. ajanhuiri* (diploid), *S. juzepczukii* (triploid), and *S. curtilobum* (pentaploid). However, there are many exceptions to grouping by ploidy. Strong statistical support occurs only for *S. ajanhuiri*, *S. juzepczukii*, and *S. curtilobum*. In combination with recent morphological analyses and an examination of the identification history of these collections, we support the reclassification of the cultivated potatoes into four species: (i) *S. tuberosum*, with two Cultivar Groups (Andigenum Group of upland Andean genotypes containing diploids, triploids, and tetraploids, and the Chilotanum Group of lowland tetraploid Chilean landraces); (ii) *S. ajanhuiri* (diploid); (iii) *S. juzepczukii* (triploid); and (iv) *S. curtilobum* (pentaploid). For other classifications, consistent and stable identifications are impossible, and their classification as species is artificial and only maintains the confusion of users of the gene banks and literature.

cultivated | microsatellites | sect. *Petota* | *Solanum tuberosum* | taxonomy

The cultivated potato represents one of the most important food plants worldwide, yet interpretation of its gene pool structure remains controversial. Contrasting taxonomic treatments of the landraces have continued over last century, with the recognition of anywhere from 1 to 21 distinct Linnean species, or of various Cultivar Groups within the single species *S. tuberosum* (1). For consistency in usage in our article and to maintain the names most familiar to scientists, we use the seven species terminology of Hawkes (2). Indigenous cultivated (landrace) potatoes are widely distributed in the Andes from western Venezuela, south to northern Argentina, and with another set of landraces in south-central Chile in Chiloé Island and the adjacent Chonos Archipelago. The Chilean landraces, although once proposed to have arisen independently from central Chile (3), are secondarily derived from the Andean ones (2), likely after hybridization with the Bolivian and Argentinean species *Solanum berthaultii* (4), a species recently combined with the formerly recognized wild species *S. tarijense* (5). Three of the Andean-cultivated species are hypothesized to be of hybrid origins with cultivated potatoes and wild species: *S. ajanhuiri* [*S. stenotomum* cultivated × *S. megistacrolobum* wild (6)], *S. juzepczukii* [*S. stenotomum* × *S. acaule* wild (7, 8)], and *S. curtilobum* [*S. andigenum* cultivated × *S. juzepczukii* (7, 8)]. The latter three “bitter potatoes” are grown in upland habitats and are not

grown nearly as extensively as *S. tuberosum*, as outlined by Huamán and Spooner (1).

The relationships and extent of genetic differentiation between the Andean and Chilean landraces has long been controversial. Based on cytoplasmic sterility factors, geographical isolation, and ecological differences, Grun (9) suggested that Chilean landraces were distinct from Andean landraces. Hawkes (2) distinguished the tetraploid Chilean from Andean landraces by characters of the leaf and flower pedicel. Plastid restriction site data documented five genotypes (A, C, S, T, and W types) in the diploid and tetraploid Andean landraces, and the Chilean landraces had three types, A, T, and W (10, 11). The most frequently observed type in the Chilean landraces (21 of 24 or 87.5% of the accessions examined) is type T, which is characterized by a 241-bp deletion (12). Conversely, 5 of the 113 (4.4%) accessions of *S. tuberosum* subsp. *andigenum* had the T type (10–12).

Potato landraces have been classified into 21 species (13, 14), 7 species with seven subspecies (2) and 9 species with two subspecies (15, 16), or as the single species *S. tuberosum* with 8 user-defined Cultivar Groups (1). Cultivar Groups are taxonomic categories used by the International Code of Nomenclature of Cultivated Plants to associate cultivated plants with traits that are of use to agriculturists and are not meant to represent natural groups or species in any classification philosophy. Ploidy levels in cultivated potatoes range from diploid ($2n = 2x = 24$), to triploid ($2n = 3x = 36$), to tetraploid ($2n = 4x = 48$), to pentaploid ($2n = 5x = 60$). Huamán and Spooner (1) examined the morphological support for the various classifications of potato landraces using representatives of all seven species from the classification of Hawkes (2). The results showed some morphological support for *S. ajanhuiri*, *S. chaucha*, *S. curtilobum*, and *S. juzepczukii*, lesser support for *S. tuberosum* subsp. *tuberosum*, and no support for *S. phureja* and *S. stenotomum*. Whatever morphological support for these entities was present was only by using a suite of characters, all of which are shared with other taxa (polythetic support). These results, combined with their likely hybrid origins, multiple origins, and evolutionary dynamics of continuing hybridization, led Huamán and Spooner (1) to recognize all landrace populations of cultivated potatoes as a single species, *S. tuberosum*, with the eight Cultivar Groups: Ajanhuiri Group, Andigenum Group, Chaucha Group, Chilotanum Group, Curtilobum Group, Juzepczukii Group, Phureja Group, and Stenotomum Group (the latter containing all landraces of the Goniocalyx Group).

Author contributions: J.N., G.T., M.d.R.H., F.G., and M.G. performed research; and D.M.S. wrote the paper.

The authors declare no conflict of interest.

[†]To whom correspondence should be addressed. E-mail: david.spooner@ars.usda.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0709796104/DC1.

© 2007 by The National Academy of Sciences of the USA

assess diversity in the *S. phureja* collection at the International Potato Center. *Solanum phureja* is widely grown in the Andes from western Venezuela to central Bolivia and has been defined by short-day adaptation, diploid ploidy ($2n = 2x = 24$), and a lack of tuber dormancy. SSR results, in combination with chromosome counts, uncovered fully 31% (32 of 102 accessions examined) triploid and tetraploid accessions from the International Potato Center collection of *S. phureja* that were long assumed to be exclusively diploid.

The purpose of our study is to reexamine the support for classification categories for landrace potatoes, using nuclear SSR markers developed for optimal utility in *S. tuberosum* regarding polymorphism, quality scores, and genomic coverage (20), supplemented with a plastid DNA deletion marker as discussed below. Nuclear SSRs have been shown to be ideal markers for detecting phylogenetically significant diversity within cultivated potatoes (19, 21, 22). In addition, their codominant nature allows them to identify polyploids when three or four bands (alleles) are found, as was shown by Ghislain *et al.* (19). This is particularly important for our study, where few cultivated accessions have been characterized for chromosome number yet ploidy has been so important conceptually in defining the cultivated species. Our study also used the 241-bp plastid deletion marker distinguishing most populations of Chilean from Andean potato landraces (4, 12, 23).

Results and Discussion

SSR Neighbor-Joining (NJ) Tree. NJ results highlight a tendency to separate three broad groups: (i) putative diploids and triploids (all accessions in the diploid cluster Fig. 1), (ii) putative tetraploids and triploids (most accessions in the polyploid cluster of Fig. 1), and (iii) the hybrid cultivated species *S. ajanhuiri* (diploid), *S. juzepczukii* (triploid), and *S. curtilobum* (pentaploid), grouped with the wild species. Bootstrap support above 50% is common in many small groups of species in the terminal branches of the NJ tree (not shown because they greatly complicate the graphic). However, bootstrap support above 50% is present only in the lower nodes supporting *S. ajanhuiri*, *S. juzepczukii*, and *S. curtilobum* and the wild species.

However, there are many exceptions of clustering by ploidy. Landraces of *S. goniocalyx*, as in the morphological study of Huamán and Spooner (1), were invariably intermixed with those of *S. stenotomum* and are so labeled as this species on Fig. 1. There are 28 putative triploid landraces of *S. chaucha* present on the main polyploid cluster of the tree and 123 on the main diploid cluster. There are 28 putative tetraploids (*S. tuberosum* subsp. *andigenum* and subsp. *tuberosum*) on the diploid cluster and 28 putative diploids (*S. phureja* and *S. stenotomum*) on the tetraploid cluster. In addition, *S. phureja* (the only cultivated species with extensive chromosome counts) shows 18 of the 89 accessions to be triploid or tetraploid. Regarding the *S. phureja*, Fig. 1 shows the position of all accessions formerly identified as this species in the International Potato Center collection (24) but shown to be polyploid in the study of Ghislain *et al.* (19). Most are present in the main “*S. phureja* cluster” (red dots in the diploid cluster of Fig. 1), but this cluster also contains two accessions of *S. tuberosum* subsp. *andigenum*, two of *S. stenotomum*, and one of *S. chaucha*, and with 10 accessions elsewhere on the diploid cluster and 16 on the polyploid cluster. As expected, most (22 of 27) of the *S. tuberosum* subsp. *tuberosum* accessions clustered together (the area in the wedge in the polyploid cluster in Fig. 1). However, this cluster also contained three accessions of *S. tuberosum* subsp. *andigenum*.

241-bp Plastid Deletion. We determined the presence or absence of the 241-bp plastid deletion for all 742 cultivated accessions examined. As expected, most (22 of 23) of the *S. tuberosum* subsp. *tuberosum* accessions in the main cluster of this subspecies

(designated by the wedge in the polyploid cluster of Fig. 1) possessed this deletion. Also in the area of the wedge are three tetraploid accessions from Peru (dark blue); two of these possess the deletion, and one lacks it (the two accessions lacking the deletion are marked with “X”).

All four remaining accessions from Chile falling outside of this cluster (gray dots marked with “X”) lack the 241-bp deletion characteristic of this subspecies, suggesting misidentifications of possible recent introductions of the *S. tuberosum* subsp. *andigenum* into Chile. Thirteen of the 251 *S. tuberosum* subsp. *andigenum* accessions (5.2%; marked with “T” outside of the gray *S. tuberosum* subsp. *tuberosum* cluster) possessed the deletion, similar to the 4.4% reported in prior studies (above). These 13 accessions are widely distributed throughout the Andes in Venezuela (2 accessions), Colombia (1 accession), Ecuador (13 accessions), Peru (4 accessions), Bolivia (2 accessions), and Argentina (1 accession). In addition, one of the *S. stenotomum* accessions (putatively diploid) and two *S. phureja* accessions (known as diploid) possessed this deletion, the first report of diploid potatoes possessing this deletion, because none of the accessions of *S. stenotomum* (215) and *S. chaucha* (150) previously screened was found with this marker (19, 25). Unfortunately, these three accessions do not have reliable collection information.

Reconsideration of the Classification of Cultivated Potato. In combination with a recent morphological study (1), the SSR data support the reclassification of the cultivated potatoes into four species: (i) *S. tuberosum*, (ii) *S. ajanhuiri* (diploid), (iii) *S. juzepczukii* (triploid), and (iv) *S. curtilobum* (pentaploid). We support dividing *S. tuberosum* into two Cultivar Groups (Andigenum Group of upland Andean genotypes containing diploids, triploids, and tetraploids, and the Chilotanum Group of lowland tetraploid Chilean landraces). Because Cultivar Groups are taxonomic categories used to associate cultivated plants with traits that are of use to agriculturists, this classification is convenient to separate these populations that grow in different areas, are adapted to different day-length regimes, and have some degree of unilateral sexual incompatibility to the Andean populations. For the remaining “species” or Cultivar Groups, consistent and stable identifications are impossible, their classification as Linnean species is artificial, and their maintenance as either species or Cultivar Groups only serves to perpetuate confusion by breeders and gene bank managers, and the instability of names in the literature. For example, Ghislain *et al.* (19) showed *S. phureja* to be indefinable as traditionally recognized because prior authors incorrectly assumed that their assumption of diploidy was incorrect for 31% of the accessions, and our results showed many accessions of *S. phureja* to cluster with the polyploids. The recognition of *S. phureja* as either a species or Cultivar Group (Phureja Group), therefore, is no longer tenable because it is no longer diploid, does not exclusively possess low-dormancy tubers, is not short-day adapted, and is not morphologically coherent (1). The other species (or Cultivar Groups) have ploidy as a major identifying criterion. The results from *S. phureja* and this study indicate that chromosome counts from other accessions of cultivated potatoes will uncover a high proportion of counts not matching expectations based on their identifications.

Ploidy has been of great importance in the classification of cultivated potatoes, but our results show so many exceptions that it is a poor character to define gene pools. Cultivated potato fields contain mixtures of different ploidy levels (6, 26–32). Bukasov (33) was the first to count chromosomes of the cultivated potatoes and used ploidy variation to speculate on hybrid origins. The strong reliance on ploidy levels was clearly stated by Hawkes and Hjerting (34): “The chromosome number of $2n =$

36 largely helps to identify *S. chaucha*, but morphological characters can also be used.”

Morphology is a poor character to define most species or Cultivar Groups except for the bitter potato species *S. ajanhuiri*, *S. curtii*, *S. juzepczukii*. As shown by Huamán and Spooner (1), most traditionally recognized cultivated potato species have little morphological support, and then only by using a suite of characters, all of which are shared with other taxa (polythetic support).

The International Potato Center has collected cultivated potatoes for 30 years and has invested tremendous effort in their identification. An examination of identification records at the International Potato Center shows many changes over the years, further showing the lack of stability of any character set to reliably define most cultivated species.

Potato gene banks are in great need of an integrated and comprehensive program of ploidy determinations; controlled and replicated studies of tuber dormancy (which we suspect will highlight grades of dormancy, not the present/absent determinations that exist today); photographically documented determinations of tuber and flesh colors and tuber shapes; and determinations of tuber pigments, glycoalkaloid contents, carbohydrates, proteins, amino acids, minerals, and secondary metabolites, using functional genomics approaches, with all data publicly integrated into a readily searchable web-based bioinformatics database. Such a multicomponent system will serve the breeding community much better than the outdated, unstable, and phylogenetically indefensible traditional classifications that exist today.

Materials and Methods

Plant Materials. A total of 742 potato landraces of all cultivated potato species were examined: *S. tuberosum* subsp. *andigenum*, putatively tetraploid (251 accessions); *S. ajanhuiri*, diploid (22); *S. chaucha*, triploid (151 accessions); *S. tuberosum* subsp. *tuberosum*, tetraploid (27 accessions); *S. curtii*, pentaploid (21 accessions); *S. juzepczukii*, triploid (35 accessions); *S. phureja*, diploid (104 accessions); *S. stenotomum*, diploid (131 accessions); 7 diploid wild species accessions in the northern *S. brevicaulle* complex *S. ambosinum* Ochoa (1 accession), *S. bukasovii* Juz. (4 accessions), and *S. multiinterruptum* Bitter (2 accessions); and the wild tetraploid species *S. acaule* Bitter (1 accession) (750 accessions in total with the 8 wild species). Selection of these wild species is based on recent amplified fragment length polymorphism (AFLP) studies that documented the northern *S. brevicaulle* complex wild species to be the progenitors of the cultivated potatoes and *S. acaule* believed to be a wild species parent in the hybrid species *S. juzepczukii* and *S. curtii*. We qualify landrace collection ploidy as “putative” because only *S. phureja* has been counted in detail (19), that showed extensive examples of incorrect assumptions of ploidy as discussed above. Data of these accessions that includes International Potato Center accession number, taxonomic identification, ploidy when known, locality of collection, and average number of SSR alleles per accession are available as a [supporting information \(SI\) Dataset](#).

DNA Extraction, SSR Primers, PCR Conditions, and Electrophoresis. Genomic DNA was obtained by using standard protocols at the International Potato Center (35). DNA concentration was calculated by using PicoGreen dsDNA quantitation reagent (Molecular Probes) and a TBS-380 Fluorometer (Turner BioSystems). DNA dilutions were performed to achieve a final concentration of 3 ng/ μ l, using 96-well plates. We used 50 nuclear SSRs (see [SI Dataset](#)) screened from 88 that included the 22 from the Potato Genetic Identity (PGI) kit (20), 13 from ESTs developed at the Scottish Crop Research Institute (36), 30 identified by using the

potato EST database at TIGR, and 23 from the University of Idaho (37). PCR reactions were performed in a 10- μ l volume containing 100 mM Tris-HCl (Sigma), 20 mM $(\text{NH}_4)_2\text{SO}_4$ (Merck), 2.5 mM MgCl_2 (Merck), 0.2 mM each dNTP (Amersham Biosciences), 0.3 μ M labeled M13 forward primer (LI-COR IRDye 700 or 800), 0.3 μ M M13-tailed SSR forward primer (Invitrogen), 0.2 μ M SSR reverse primer (Invitrogen), 1 unit of Taq polymerase (GIBCO/BRL), and 15 ng of genomic DNA. PCR was carried out in a PTC-200 thermocycler (MJ Research). The program used was the following: 4 min at 94°C, followed by 33 cycles of 50 sec at 94°C, 50 sec at annealing temperature (T^a), and 1 min at 72°C, then 4 min at 72°C as a final extension step. PCR products were separated by electrophoresis on a LI-COR 4300 DNA analyzer system. The molecular weight ladder was the LI-COR IRDye 50–350 bp size standard and was loaded into gel each eight samples.

SSR Allele Scoring. SSR alleles were detected and scored by using SAGA Generation 2 software (LI-COR). Size calibration and an SSR “smiling line” were performed by using the molecular weight ladder (LI-COR IRDye 50–350). The SSR alleles were determined for size in bp of the upper band of the allele and scored as present (1) or absent (0). Missing data were scored as “9.”

Data Analysis. Genetic analysis was performed by using the program DARwin (38). A dissimilarity matrix was calculated by using Jaccard’s coefficient, 60% of minimal proportion of valid data required for each unit pair, and 500 replicate bootstrapping. The dendrogram was built by using the NJ method, using the seven wild species accessions in the northern *S. brevicaulle* complex as outgroup. The NJ method developed by Saitou and Nei (39) estimates phylogenetic trees. Although based on the idea of parsimony (it does yield relatively short estimated evolutionary trees), the NJ method does not attempt to obtain the shortest possible tree for a set of data. Rather, it attempts to find a tree that is usually close to the true phylogenetic tree (40). This method allows the rooting of trees on outgroups (in this case, the seven accessions of the *S. brevicaulle* complex). The polymorphic information content (PIC) was calculated as $\text{PIC} = 1 - \sum(p_i^2)$, where p_i is the frequency of the i th allele detected in all accessions (41). Data of somatic chromosome counts for accessions of *S. phureja* were obtained from Ghislain *et al.* (19).

Plastid DNA Polymorphism Detection. The 241-bp deletion was analyzed for all 742 cultivated accessions by using the primers from ref. 23. PCR amplification was performed in a volume of 10 μ l consisting of 18 ng of genomic DNA, 0.4 μ M each of primers (Invitrogen), 1 \times PCR buffer (PerkinElmer), 2.5 mM MgCl_2 (PerkinElmer), 200 μ M each dNTP (Amersham Biosciences), and 0.25 unit of Taq DNA polymerase (GIBCO/BRL). Thermal cycling was carried out in a PTC-200 thermocycler (MJ Research) (one cycle of 4 min at 94°C, followed by 40 cycles of 45 sec at 94°C, 45 sec at 59°C, and 45 sec at 72°C, then terminated with one cycle of 4 min at 72°C). PCR products were separated by electrophoresis in a 1% agarose gel, and lambda phage digested by PstI was used as a molecular weight marker. The 241-bp plastid polymorphism was determined for size in bp and scored as “T” (\approx 200 bp for deleted type) and “X” (\approx 440 bp for undeleted type).

We thank David Douches and Lynn Bohs for comments on an earlier draft of the manuscript. This work was supported by the International Potato Center, U.S. Department of Agriculture, Generation Challenge Program Grant SP1C2-2004-5, and National Science Foundation Grant DEB 0316614 entitled “A world-wide treatment of *Solanum*” (to D.M.S.).

1. Huamán Z, Spooner DM (2002) *Am J Bot* 89:947–965.
2. Hawkes JG (1990) *The Potato: Evolution, Biodiversity, and Genetic Resources* (Belhaven, London).

3. Ugent D, Dillehay T, Ramirez C (1987) *Econ Bot* 41:17–27.
4. Hosaka K (2003) *Am J Potato Res* 80:21–32.
5. Spooner DM, Fajardo D, Bryan GJ (2007) *Taxon* 56:987–999.

