

# Advantages for the Use of Standardized Phenotyping in Databases

Gayle M. Volk

National Center for Genetic Resources Preservation, U.S. Department of Agriculture, Fort Collins, CO 80521

*Additional index words.* Darwin Core, database, geospatial, interoperability, ontology, phenotype

**Abstract.** Genomic sequence data are becoming prevalent in databases; however, associated phenotypic data are much more expensive and difficult to obtain. Standardized ontologies for trait classification provide a mechanism by which searches can be efficiently performed. Through standardization efforts, results can be compared across years, researchers, and locations. Examples of databases that have developed standardized ontologies within medical, animal, and plant research programs demonstrate that standardization is both possible and desirable. With standardization, genotypic and phenotypic data can be analyzed to materials for improved yield, quality, and stress tolerance.

Biological collections in herbaria, museums, botanical gardens, genebanks, breeder plots, and research institutions share many common attributes. For documentation of accessions within collections, databases have often been developed independently, or locally, with the intention of improving data inventory, organization, use, and retrieval. The most successful databases often include features that improve operability and interoperability with related databases. The use of standardized data within databases increases the potential for interoperability and increased functionality.

Many biological databases have fields for categories such as sample identification, taxonomy, original source (either georeference or locality), institution, and inventory quantities. Depending on the type of database, additional information often includes ancestry (parentage), phenotype, genotype, image, or other related characteristic data. This review provides some examples of the use of standardized methodologies in databases.

Standard fields use specific ontologies, or vocabularies, in defined categories. Along with reference standards, confidence intervals, and detailed measurement methods, users have information to identify trends across years, programs, locations, species, or families. These broad-scale analyses facilitate the discovery of trends that could be cost-prohibitive if performed as independent experiments. For example, trait plasticity and environmental responses could be evaluated across data sets that are compatible as a result of standardized data collection and reporting methods.

## EXAMPLES OF DATABASES AND THEIR APPLICATIONS

Biological collection curators have developed a set of standards with the goal of facilitating the exchange of information about their collections. The set of Biodiversity Information Standards, as organized by the Taxonomic Data Working Group (TDWG) (Darwin Core Task Group, 2009), includes fields for the taxonomic description (kingdom, phylum, class, order, family, genus, species) as well as the unique institution number and inventory number for the accession (Bisby, 1994). Fields are also available for image and source information. These data include continent, country, state, county and locality, elevation, and collection methods. This group of standards is referred to as the Darwin Core and has been adopted by hundreds of institutions worldwide.

The Geospatial Extensions of the Darwin Core provide standards for georeference data. Among the fields described are decimal latitude, decimal longitude, and a spatial description of the collection source. A critical component of the geospatial extensions is the estimation of uncertainty of a specific referenced point (Chapman and Wieczorek 2006; Wieczorek et al., 2004). The combined use of the Darwin Core and the Geospatial Extensions provides a thorough geographical description of a locality associated with an accession. This locality is most often the source location of the item.

The Global Biodiversity Information Facility (GBIF) uses the TDWG standards for the Darwin Core and Geospatial Extensions to document primary plant and animal biodiversity data and associated metadata (Tuama, 2008). The data are available through a Data Portal and as of Sept. 2009, GBIF includes summary information on 8634 data sets from 299 data providers (Global Biodiversity Information Facility, 2009). The Data Portal serves as a search engine to retrieve data relating to taxon, country, or data set (Flemons et al., 2007). The system is built to integrate with other systems to merge key conservation data sets.

A search of the GBIF data set for *Malus sieversii*, a wild apple native to Central Asia, reveals three data sets with a total of 2310 accessions of this species in the GBIF consortium. After the search is initiated, a map is

displayed that highlights the geographical location of the source of each accession. A table indicates that one accession is maintained at the Vascular Plant Collection in the University of Washington Herbarium, 18 accessions are recorded for the Center for Genetic Resources in The Netherlands, and 2291 accessions are documented within the U.S. National Plant Germplasm System Collection. All the accessions are shown to have originated in Central Asia (despite being currently held in The Netherlands and in the United States), except for the herbarium accession. The herbarium accession is shown on the map as from the northwest corner of Washington State—its current holding location rather than its original source location—thus illustrating the importance that data providers document data in similar manners.

The data from each data provider can be viewed in a table reporting the scientific name, data set, institution code, collection code, catalog number, basis or record, data coordinates (latitude and longitude), and country. Additional information such as accession number, detailed locality information, elevation, and inventory data for individual accessions can be viewed on screens that list the records specific to each accession.

The success of the GBIF data portal depends on the use of standardized data fields. Single searches retrieve information available in hundreds of databases worldwide and provide access to information that would not otherwise be available. Currently, the GBIF data portal is specific to data set, taxonomy, and geospatial data. However, adequate information is provided to enable researchers to retrieve information (if available) from institutional databases.

Although the use of standardized data is mandatory in fields such as engineering, the value of standardized phenotyping has more recently been recognized in biological disciplines such as medicine. One prime example of current standardized phenotyping efforts is the standardization and digitization of medical records. By standardizing terminology and reporting methods, uniformity is established among doctors' offices, hospitals, and clinical trials. In fact, agencies such as the National Institutes of Health and the National Center for Biotechnology Information have begun requiring standardized data collection

Received for publication 9 Mar. 2010. Accepted for publication 4 Apr. 2010.

I thank Christopher Richards and Nahla Bassil for helpful discussions and Mark Bohning, Alfredo Alves, and Pankaj Jaiswal for providing manuscript reviews before submission.

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

e-mail Gayle.Volk@ars.usda.gov.

in clinical trials (Garcia et al., 2007; Mailman et al., 2007). Advantages of these efforts include the development of large-scale data sets and the facilitation of meta-analyses of data. With more data, an increased understanding of human pathology can lead to previously undiscovered correlations and improved treatment (Adler, 2007).

Successful standardization and database interoperability is dependent on using controlled vocabularies, nomenclature standards, or ontologies. In these standards, each term is precisely defined and the relationships between and among terms are represented. For example, the human phenotype ontology includes 9500 terms and their relationships (Robinson et al., 2008). This database includes symptoms of hereditary disorders with critical details such as the onset time and speed of progression. Another well-established vocabulary is the Gene Ontology (GO), whereby gene products are assigned ontology terms for their molecular functions, role in biological processes, and localization in cellular components (Gene Ontology Consortium, 2004; The Gene Ontology Consortium, 2008). A web application entitled AmiGO provides access to search and browse the GO and the gene product annotations (Carbon et al., 2009).

Comparative data mining becomes realistic through the use of ontologies. Large whole-genome sequencing efforts in both model organisms such as mice and rats as well as humans have allowed for detailed cross-genome comparisons. Specific relationships between genes, phenotypes, and diseases in model systems can be directly converted from the model systems to results from clinical trials. The European mouse phenotyping resource for standardized screens (EMPreSS) has assisted in these endeavors by standardizing methods for phenotype measurements (Beck et al., 2009; Green et al., 2005). This database provides a portal for searching and retrieving Standard Operating Procedures for mouse phenotyping protocols. One key to its success has been that it was developed through program consortia in multiple European countries (Green et al., 2005).

Genome databases such as the Mouse Genome Database curate genetic, genomic, and phenotypic data. With reference data sets and consensus views, the database links to other bioinformatic resources and promotes use through user-friendly interfaces (Blake et al., 2009; Bult, 2002; Bult et al., 2008).

The use of controlled vocabularies is essential for integrating data resources for analyses, information exchange, and effective browsing. In plants, the Plant Ontology (PO) Consortium has begun to establish ontologies for describing plant structures (anatomical and morphological parts; Ilic et al., 2007), whole plant growth stages (Pujar et al., 2006), and organ development stages in plants (Avraham et al., 2008; Jaiswal et al., 2005). The current ontologies focus primarily on major model plants (Arabidopsis, maize, rice, and Solanaceae), but ongoing efforts will extend this database to include additional agronomic and horticultural species such as

those from Rosaceae. In another effort to standardize the phenotype trait descriptors, Jaiswal et al. (2002) developed the Trait Ontology (TO) for plants. Although its initial development started with populating terms from grasses, recently it has evolved by including Solanaceae and Fabaceae and a mandate from the Consultative Group on International Agricultural Research Generation Challenge Program requires inclusion of cacao and banana as well. The successful integration of the TO in the annotation of phenotypes, GO for gene functions, and PO for annotating the expression of genes and phenotypes (mutants and quantitative trait loci; Ni et al., 2009) in the Comparative Plant Genomics databases such as Gramene (Liang et al., 2007) helped to establish relationships between phenotypic and genotypic data (Ni et al., 2009). The use of these descriptors has facilitated analyses on the variation of populations and how it relates to quantitative and observable traits.

Implementation of large-scale use of standardized phenotypic characters is expensive and intensive. Efforts to reduce the labor and errors in data collections have included the use of bar coding. Through the use of bar codes, handheld devices can document sample and inventory codes as well as the associated data point for each phenotypic character (Flint-Garcia et al., 2005). This automated data capture saves time and improves accuracy by minimizing the typographical or transposition errors.

The RIKEN Arabidopsis phenome information database was developed to document phenotypic information to aid in identification of transposon-insertional mutants in a large collection (Kuromori et al., 2006). Phenotypic data for eight primary (seedling, leaves, flowering and growth, stems, branching, flowers, fruits, seed yield) and 50 secondary categories were captured for the mutant collection. The database is searchable based on the traits of interest and after querying the database, phenotypic character information and images are presented. These data are intended to be used in combination with the cDNA, genome map, and transposon mutant information provided within the RARGE database (Sakurai et al., 2005; Seki and Shinozaki, 2009).

An example of mass-phenotypic data collection is that provided by a company that specializes in imaging, LemnaTec (LemnaTec GmbH, Wurselen, Germany). The high content screening provides morphological phenotype assessment of seedling plants and has been implemented in quantifying salinity tolerance (Rajendran et al., 2009). Through digital imaging, measurements of compactness, symmetry, surface coverage, leaf-width index, and leaf numbers can be measured using an automated platform.

#### DEVELOPMENT AND USES OF STANDARDIZED PHENOTYPIC TRAITS

Development of standards is not a trivial process. According to the American National

Standards Process, as published by the American National Standards Institute (1997), standards are developed as a result of consensus on proposed standards by a group of representatives. This process depends on broad-based public review and comments on draft standards that are publicized. Feedback is considered and changes are incorporated into the proposed standards. Importantly, there is a right to appeal the process by any participant that alleges that the principles of standard development were not respected. It is clear that a thorough, careful, and inclusive development of standards is needed to ensure that the products are both useful and informative.

Standardized phenotypic traits are best developed through collaborations among interested parties. The European Cooperative Program for Plant Genetic Resources/Bioversity has established ECPGR working groups specific for 18 sets of crops (Maggioni, 2001). These groups aim to develop standardized phenotypic descriptors for crops across the 41 European member countries and report information in unified databases. For some crops, the development of descriptors has been highly successful using this collaborative process. This provides access to most of the *ex situ* plant genetic resource collections in Europe and makes key data such as passport information, characterization, and evaluation data publicly available (Lipman et al., 1997).

Within the United States, the Agricultural Research Service (ARS) National Plant Germplasm System (NPGS) maintains a diverse collection of more than 500,000 accessions. This national resource has become very valuable as genomic technologies progress and it becomes possible to retrieve genetic information from one species and use it in another (Rubenstein et al., 2006; Volk et al., 2008). Representatives from universities, federal research programs, and industry have the opportunity to join Crop Germplasm Committees (CGCs). These working groups of specialists provide recommendations on genetic resources within specific crops or groups of related crops of present or future economic importance. One function of these 42 CGCs is to prioritize traits for evaluation and develop proposals to implement the evaluations. The CGCs provide feedback directly to the USDA-ARS NPGS crop curators and work to ensure that standardized, accurate, and useful information is entered into the GRIN database, which is a national database system documenting plant genetic resources available through the NPGS (Mowder and Stoner, 1988; Perry et al., 1988; USDA, ARS, National Genetic Resources Laboratory, 2009a). Phenotypic descriptors are crop-specific; however, there are efforts to standardize some descriptors across multiple crops. Descriptors are currently classified into categories of chemical, cytologic, disease, morphology, phenology, production, and stress. Data for descriptors can either be quantitative or qualitative, and both types of descriptors are usually coded into categories. As a result, it is possible to retrieve both quantitative and categorical

data types for many descriptors. Definitions of descriptors and examples of germplasm representing the code classifications are available online (USDA, ARS, National Genetic Resources Laboratory, 2009b).

The Genome Database for Rosaceae is in the process of including data for standardized phenotypic descriptors for key economic traits of Rosaceae genera (Jung et al., 2004). Among others, these traits include disease resistance, fruit quality, and yield descriptors. Merging key breeder, genomics, and germplasm collection data sets will facilitate data mining efforts in which novel germplasm with desirable traits linked to key markers can be identified. In the future, interoperability between key databases containing genetic, phenotypic, and passport data such as GDR and GRIN will be critical to improve access to information.

## CONCLUSION

The use of standardized phenotypes is beneficial in that it allows for the development of large-scale data analyses and interpretations that might not otherwise be possible. Some of these issues include trait evolution, domestication, and the identification of underlying architecture across diverse genera such as that performed using comparative genomics. Furthermore, researchers can better apply the known gene/function information from model plant systems to less studied horticultural crops. In addition, researchers can target and identify novel germplasm accessions that may harbor key traits of interest and breeders can identify the extent of trait plasticity through the merging of well-documented data sets. Finally, the use of standards in databases facilitates the interoperability among databases to improve information flow and use.

## Literature Cited

- Adler, D. 2007. Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med. Care* 45:S1–S2.
- American National Standards Institute. 1997. Procedures for the development and coordination of American national standards. American National Standards Institute, New York, NY.
- Avraham, S., C.-W. Tung, K. Ilic, P. Jaiswal, E.A. Kellogg, S. McCouch, A. Pujar, L. Reiser, S.Y. Rhee, M.M. Sachs, M. Schaeffer, L. Stein, P. Stevens, L. Vincent, F. Zapata, and D. Ware. 2008. The Plant Ontology Database: A community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res.* 36:D449–D454.
- Beck, T., H. Morgan, A. Blake, S. Wells, J.M. Hancock, and A.M. Mallon. 2009. Practical application of ontologies to annotate and analyze large scale raw mouse phenotype data. *BMC Bioinformatics* 10(suppl 5):S2.
- Bisby, F.A. 1994. Plant names in botanical databases. *Plant Taxonomic Database Standards* No. 3. Version 1.00. Hunt Institute for Botanical Documentation. Carnegie Mellon University, Pittsburgh, PA.
- Blake, J.A., C.J. Bult, J.T. Eppig, J.A. Kadin, and J.E. Richardson, and the Mouse Genome Database Group. 2009. The Mouse Genome Database genotypes::phenotypes. *Nucleic Acids Res.* 37:D712–D719.
- Bult, C.J. 2002. Data integration standards in model organisms: From genotype to phenotype in the laboratory mouse. *Targets* 1:163–168.
- Bult, C.J., J.T. Eppig, J.A. Kadin, J.E. Richardson, and J.A. Blake, and the Mouse Genome Database Group. 2008. The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.* 36:D724–D728.
- Carbon, S., A. Ireland, C.J. Mungall, S.Q. Shu, B. Marshall, and S. Lewis, and the AmiGO Hub and the Web Presence Working Group. 2009. AmiGO: Online access to ontology and annotation data. *Bioinformatics* 25:288–289. 29 Sept. 2009. <<http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>>.
- Chapman, A.D. and J. Wieczorek (eds.). 2006. Guide to best practices for georeferencing. Global Biodiversity Information Facility. Copenhagen, Denmark.
- Darwin Core Task Group. 2009. Darwin core. 29 Sept. 2009. <<http://rs.tdwg.org/dwc/2009-07-06/index.htm>>.
- Flemons, P., R. Guralnick, J. Krieger, A. Ranipeta, and D. Neufeld. 2007. A web-based GIS tool for exploring the world's biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application. *Ecological Informatics* 2:49–60.
- Flint-Garcia, S.A., A.-C. Thuillet, J. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2005. Maize association population: A high resolution platform for quantitative trait locus dissection. *Plant J.* 44:1054–1064.
- Garcia, S.F., D. Cella, S.B. Clauser, K.E. Flynn, T. Lad, J.S. Lai, B. Reeve, A.W. Smith, A.A. Stone, and K.P. Weinfurt. 2007. Standardizing patient-reported outcomes assessment in cancer clinical trials: A PROMIS initiative. *J. Clin. Oncol.* 25:5106–5112.
- Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32:D258–D261.
- Global Biodiversity Information Facility. Version 1.2.5. Published 6 Aug. 2009. 15 Sept. 2009. <<http://www.gbif.org>>.
- Green, E.C.J., G.V. Gkoutos, H.V. Lad, A. Blake, J. Weekes, and J.M. Hancock. 2005. EM-PRess: European Mouse Phenotyping Resource for Standardized Screens. *Bioinformatics* 21: 2930–2931.
- Ilic, K., E.A. Kellogg, P. Jaiswal, F. Zapata, P.F. Stevens, L.P. Vincent, S. Avraham, L. Reiser, A. Pujar, M.M. Sachs, N.T. Whitman, S.R. McCouch, M.L. Schaeffer, D.H. Ware, L.D. Stein, and S.Y. Rhee. 2007. The plant structure ontology, a unified vocabulary of anatomy and morphology of a flowering plant. *Plant Physiol.* 143:587–599.
- Jaiswal, P., S. Avraham, K. Ilic, E.A. Kellogg, S. McCouch, A. Pujar, L. Reiser, S.Y. Rhee, M.M. Sachs, M. Schaeffer, L. Stein, P. Stevens, L. Vincent, D. Ware, and F. Zapata. 2005. Plant Ontology (PO): A controlled vocabulary of plant structures and growth stages. *Comp. Funct. Genomics* 6:388–397.
- Jaiswal, P., D. Ware, J. Ni, K. Chang, W. Zhao, S. Schmidt, X. Pan, K. Clark, L. Teytelman, S. Cartinhour, L. Stein, and S. McCouch. 2002. Gramene: Development and integration of trait and gene ontologies for rice. *Comp. Funct. Genom.* 3:132–136.
- Jung, S., C. Jesudurai, M. Staton, Z. Du, S. Ficklin, I. Cho, A. Abbott, J. Tomkins, and D. Main. 2004. GDR (Genome Database for Rosaceae): Integrated web resources for Rosaceae genomics and genetics research. *BMC Bioinformatics* 5:130.
- Kurumori, T., T. Wada, A. Kamiya, M. Yuguchi, T. Yokouchi, Y. Imura, H. Takabe, T. Sakurai, K. Akiyama, T. Hirayama, K. Okada, and K. Shinozaki. 2006. A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis. *Plant J.* 47:640–651.
- Liang, C., P. Jaiswal, C. Hebbard, S. Avraham, E.S. Buckler, T. Casstevens, B. Hurwitz, S. McCouch, J. Ni, A. Pujar, D. Ravenscroft, L. Ren, W. Spooner, I. Teclé, J. Thomason, C. Tung, X. Wei, I. Yap, K. Youens-Clark, D. Ware, and L. Stein. 2007. Gramene: A growing plant comparative genomics resource. *Nucleic Acids Res.* 36:D947–D953.
- Lipman, E., M.W.M. Jongen, Th.J.L. van Hintum, T. Gass, and L. Maggioni. 1997. Central Crop Databases: Tools for plant genetic resources management. International Plant Genetic Resources Institute, Rome, Italy/CGN, Wageningen, The Netherlands.
- Maggioni, L. 2001. Achievements and perspectives of the ECP/GR networking activity for the conservation and use of crop genetic resources, p. 133–142. In: Wicicki, W., B. Naganowska, and B. Wolko (eds.). Broad variation and precise characterization—limitation for the future. Proc. of the XVth EUCARPIA Section Genetic Resources Workshop, 16–20 May 2001, Poznań, Poland. Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland.
- Mailman, M.D., M. Feolo, Y. Jin, M. Kimura, K. Tryka, R. Bagoutdiov, L. Hao, A. Kiang, J. Paschall, L. Phan, N. Popova, S. Pretel, L. Ziyabari, M. Lee, Y. Shao, Z.Y. Wang, K. Sirotkin, M. Ward, M. Kholodov, K. Zbicz, J. Beck, M. Kimelman, S. Shevelev, D. Preuss, E. Yaschenko, A. Graeff, J. Ostell, and S.T. Sherry. 2007. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* 39:1181–1186.
- Mowder, J.D. and A.K. Stoner. 1988. Information systems. *Plant Breed. Rev.* 7:57–65.
- Ni, J., A. Pujar, K. Youens-Clark, I. Yap, P. Jaiswal, I. Teclé, C. Tung, L. Ren, W. Spooner, X. Wei, S. Avraham, D. Ware, L. Stein, and S. McCouch. 2009. Gramene QTL database: Development, content and applications. Database: Bap005. doi: 10.1093/database/bap005.
- Perry, M., A.K. Stoner, and J.D. Mowder. 1988. Plant germplasm information management system: Germplasm Resources Information Network. *HortScience* 23:57–60.
- Pujar, A., P. Jaiswal, E.A. Kellogg, K. Ilic, L. Vincent, S. Avraham, P. Stevens, F. Zapata, L. Reiser, S.Y. Rhee, M.M. Sachs, M. Schaeffer, L. Stein, D. Ware, and S. McCouch. 2006. Whole-plant growth stage ontology for angiosperms and its application in plant biology. *Plant Physiol.* 142:414–428.
- Rajendran, K., M. Tester, and S.J. Roy. 2009. Quantifying the three main components of salinity tolerance in cereals. *Plant Cell Environment* 32:237–249.
- Robinson, P.N., S. Köhler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos. 2008. The Human Phenotype Ontology: A tool for annotating and analyzing human hereditary disease. *Amer. J. Human Genet.* 83:610–615.
- Rubenstein, K.D., M. Smale, and M.P. Widrechner. 2006. Demand for genetic resources and the U.S. National Plant Germplasm System. *Crop Sci.* 46:1021–1031.
- Sakurai, T., M. Satou, K. Akiyama, K. Iida, M. Seki, T. Kurumori, T. Ito, A. Konagaya, T. Toyoda, and K. Shinozaki. 2005. RARGE: A large-scale database of RIKEN Arabidopsis resources ranging from transcriptome to

- phenome. *Nucleic Acids Res.* 33:D647–D650. 29 Sept. 2009. <<http://rarge.gsc.riken.jp/>>.
- Seki, M. and K. Shinozaki. 2009. Functional genomics using RIKEN Arabidopsis thaliana full-length cDNAs. *J. Plant Res.* 122:355–366.
- The Gene Ontology Consortium. The Gene Ontology project in 2008. *Nucleic Acids Res.* 36:D440–D444. 29 Sept. 2009. <<http://www.geneontology.org/>>.
- Tuama, E.O. 2008. Metadata requirements for datasets delivered via the Global Biodiversity Information Facility (GBIF) Network. 16 Aug. 2010. <[http://www2.gbif.org/GBIF-metadata-strategy\\_v.06.pdf](http://www2.gbif.org/GBIF-metadata-strategy_v.06.pdf)>.
- USDA, ARS, National Genetic Resources Laboratory. 2009a. Germplasm Resources Information Network (GRIN) [online database]. National Germplasm Resources Laboratory, Beltsville, MD, 29 Sept. 2009. <<http://www.ars-grin.gov/npgs/>>.
- USDA, ARS, National Genetic Resources Laboratory. 2009b. Germplasm Resources Information Network (GRIN) [online database]. National Germplasm Resources Laboratory, Beltsville, MD. 29 Sept. 2009. <<http://www.ars-grin.gov/npgs/html/croplist.pl>>.
- Volk, G.M. and C.M. Richards. 2008. Availability of genotypic data for USDA-ARS National Plant Germplasm System accessions using the Genetic Resources Information Network (GRIN) database. *HortScience* 43:1365–1366.
- Wieczorek, J., Q. Guo, and R.J. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *Intl. J. Geogr. Inf. Sci.* 18:745–767.