Technology and Product Reports

Using Open Source Software in Developing a Web-accessible Database of Sweetpotato Germplasm Collections in Kenya

Arthur Villordon[1,6], Wambui Njuguna[2], Simon Gichuki[3], Philip Ndolo[4], and Don Labonte[5]

ADDITIONAL INDEX WORDS. *Ipomoea batatas*, general public license, database management system

SUMMARY. Web-accessible germplasm databases allow stakeholders to interactively search and locate information in real time. These databases can also be configured to permit designated users to remotely add, delete, or update information. These resources assist in decision-making activities that are related to germplasm documentation, conservation, and management. We report the development of a web-accessible database of Kenyan sweetpotato (*Ipomoea batatas*) varieties using open source software. Kenya is located in eastern Africa, a region that is considered one of the centers of diversity for sweetpotato. We describe the software applications used in developing the germplasm database as well as the web interface for displaying and interactively searching records. This report demonstrates that open source software can be used in developing a web-enabled database with management features similar to those found in proprietary or commercial applications.

C ollaborative research to document and collect Kenyan sweetpotato germplasm has led to the compilation and ongoing development of viazivitamuDB (East African Sweetpotato Project, 2006), a web-based resource of sweetpotato diversity for eastern Africa. The database name is derived from "viazi vitamu," which means sweetpotato in Swahili, the language predominantly spoken in Kenya and in most of eastern Africa. Eastern African is generally considered as a secondary center of diversity for sweetpotato (Austin, 1988; Gichuki et al., 2000; Huamán and Zhang, 1997; Yen, 1982), where the species regularly set seed in the short-daylength tropical conditions. Over time, natural hybridization and selection have resulted in the development of hundreds of native sweetpotato varieties in the region (Gichuki et al., 2003). Most countries in eastern

and central Africa have witnessed regional conflicts and adverse weather in the last two decades that have led to decreased landholdings, which in turn have led to fewer crops with less or little surplus to store (Sperling, 2002), thereby accelerating the risk of germplasm loss. An extreme example is that of Rwanda, where a civil war has disrupted and even devastated agricultural production, including the potential loss of a wealth of local crop varieties (Buruchara et al., 2002).

The majority of the software applications used in the development of VVDB were open source software (OSS). OSS applications are computer programs that are freely available and, depending on the included license, allow users to modify and distribute the application's source code. Dudoit et al. (2003) further defined OSS programs as possessing a clear, well-defined application program interface (API). This allows developers and users to integrate the software with other systems and to add new functionality. This ability to add new utilities and functionality to software assists in developing resources that help advance a scientific field, provide a workbench of tools that allow researchers to explore and expand methods to analyze biological data, and ensure that the international scientific community is the owner of the software tools needed to carry out research (Dudoit et al., 2003). Various OSS solutions are available to the horticultural community (Mann et al., 2005; Rhodus, 2006; Villordon and Franklin, 2005) for image manipulation, web communications, and related applications. The objective of this work was to document the use of OSS applications in developing a web-enabled database that describes the diversity of sweetpotato in Kenya.

Materials and methods

GERMPLASM RECORDS AND DESCRIPTORS. The information in the VVDB database was originally stored in various documents, including paper-based records and spreadsheet-compatible computer files that contained morphological descriptors or passport information. In addition to the variety name and an internal accession number, the database includes data fields (where available) for location (village, town, province, and region), geographic coordinates, and 25

[1]LSU AgCenter, Sweet Potato Research Station, 130 Sweet Potato Road, Chase, LA 71324

[2]Department of Horticulture, Oregon State University, Corvallis, OR 97331

[3]Kenya Agricultural Research Institute (KARI), Biotechnology Centre, P.O. Box 57811, Nairobi, 00200 Kenya

[4]KARI Kakamega Research Station, P.O. Box 169, Kakamega, 50100 Kenya.

[5]Louisiana State University, Department of Horticulture, 137 J.C. Miller Hall, Baton Rouge, LA 70803

[6]Corresponding author. E-mail: avillordon@agctr.lsu.edu.

morphological descriptors previously developed by the International Potato Center (Huamán, 1991). These descriptors include various foliage and vine characteristics (e.g., vine internode length, vine internode diameter, predominant vine color, leaf lobe number, central leaf lobe shape, mature leaf size, abaxial leaf pigmentation, mature leaf color, immature leaf color, and petiole length) and storage root traits (e.g., storage root shape, storage root surface defects, storage root cortex thickness, storage root skin color, intensity of predominant skin color, secondary skin color, and storage root flesh color).

DATABASE DEVELOPMENT. The germplasm accession records and descriptors were compiled into a single comma-separated value (CSV) file and imported into a MySQL (MySQL AB, 2007) database table previously created using phpMyAdmin (php MyAdmin Project, 2007). These applications were pre-installed on a commercial computer server running a modified Linux operating system (Red Hat, Raleigh, NC) and Apache web software (Apache Software Foundation, 2007). Linux (network operating system), MySQL (database), phpMyAdmin (MySQL graphical management tool), and Apache (web server) are open source applications that are available as free software under the GNU General Public License terms (GNU Project, 2007).

DEVELOPMENT OF WEB INTERFACE. The database web interface was configured using Dadabik (Tacchini, 2007), a hypertext preprocessor (PHP) programming language-based OSS application that creates highly customizable web form interfaces for MySQL database tables. The Dadabik installation file was downloaded from the source web site, uploaded to the web server, and installed using a web graphical user interface (GUI) similar to desktop computer application installation software. After all the germplasm records were consolidated into a spreadsheet file, the length of time required to upload the information into the MySQL database table and install software (Dadabik) as well as configure and test the database web interface did not exceed 8 h. The annual cost of the shared server space is $60.

Results and discussion

The VVDB database currently contains 327 unique records of

sweetpotato varieties that have been collected from various Kenyan locations and maintained in two main regional collections. Figure 1 shows the web interface to VVDB. Figure 2 shows the VVDB search interface. This database includes digital images of in situ plant specimens for ≈18%
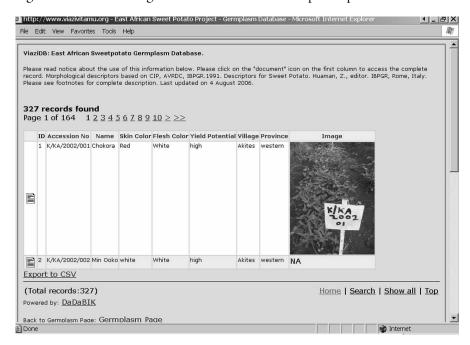

Fig. 1. Screenshot of the main web interface to the Kenyan sweetpotato accession database, viazivitamuDB (VVDB). Dadabik, version 4.0 (Tacchini, 2007), was used to generate the default web interface to the database. The default number of records and fields displayed can be changed from the software's configuration page. The "home" link returns the user to this default page from record searches and retrieval activities. The "search" link opens the records search page. The "show all" link displays all records. The "top" link assists the user to return to the top of the page when browsing several records.

Fig. 2. Screenshot of the default search interface to the Kenyan sweetpotato accession database, viazivitamuDB (VVDB). Dadabik version Dadabik, version 4.0 (Tacchini, 2007), was used to generate the default search interface. In this configuration, the database engine will perform a search and return records even if only one field is specified.

of the records; ≈92% of the accessions have been fully characterized using the morphological descriptors described in the Methods section. At present, we are unaware of any similar published database resource for the eastern Africa region, particularly inclusion of images for certain records. Although comprehensive online crop species germplasm databases are available, the sweetpotato diversity available in eastern Africa is often underrepresented. For example, the SINGER [Consultative Group on International Agricultural Research (CGIAR), 2007] database only lists 27 accessions for sub-Saharan Africa (three from Kenya). The SINGER database represents a gateway to the world's plant diversity as represented by collections of 15 international agricultural research centers supported by the CGIAR.

VVDB enables global access by stakeholders to a shared resource. For example, researchers performing remote database queries can determine if likely duplicates exist in their respective germplasm collections. In addition, the database provides an overview of the status of sweetpotato diversity in eastern Africa and facilitates collaborative partnerships in further conserving sweetpotato germplasm diversity in the region. An offline version of VVDB is also available for local, non-networked computer searches. This offline version is stored on optical disks (CD-ROM format) and is available upon request from the authors. This report demonstrates that enabling remote access to databases need not be cost-prohibitive or tied to a proprietary system, thus helping to promote efficient use of research and development resources. PhpMyAdmin, Dadabik, and similar OSS database management systems (DBMS) offer database management features that are comparable to those found in proprietary web-enabled DBMS like Sybase (Sybase, Dublin, CA) and FileMaker (FileMaker, Santa Clara, CA). Considering that the server versions of these commercial databases cost at least $999 per license, the startup cost associated with developing VVDB was very minimal. The Dadabik installation process detected the MySQL database table and created default display and search web interfaces. MySQL is widely used in the scientific computing community to run databases that contain sequence data (Martin et al., 2002), genome information (Close et al., 2004; Hubbard et al., 2002), and other scientific information. Similar programs that create web interfaces for MySQL databases are commercially available. These web interfaces serve as the GUI for the database tables and allow users without an in-depth knowledge of Structured Query Language (SQL) to perform record searches, insertions, updates, and deletions. SQL is a standard interactive and programming language used by DBMS applications to query, update, and manage data. These tools helped project collaborators to develop web access to the MySQL database table without writing a single line of programming code. Other features included file uploading, export to comma-separated values (CSV), checking for possible duplication during an insert, authentication, and authorization restrictions during edit operations (view, update, and delete). Graphic image formats that are supported include JPEG, GIF, and TIFF. The database administration page can be accessed by authorized users for record addition, deletion, and modification. This enabled remote access and real-time data entry or editing by designated Internet-enabled stakeholders stationed in various locations. As digital images and descriptions of collections are completed, records can be added or updated accordingly. A centralized germplasm database further increases efficient use of resources by reducing potential duplication of collections and promoting resource and information sharing among institutions, organizations, and individual collaborators.

Literature cited

Apache Software Foundation. 2007. Apache Software Foundation. 14 Mar. 2007. <http://www.apache.org>.

Austin, D.F. 1988. The taxonomy, evolution, and genetic diversity of sweetpotatoes and related wild species, 27–59. In: P. Gregory (ed.). Exploration, maintenance, and utilization of sweetpotato genetic resources. Report of the First Planning Conf. 1987. International Potato Center, Lima, Peru.

Buruchara, R.A., L. Sperling, P. Ewell, and R. Kirkby. 2002. The role of research institutions in seed-related disaster relief: Seeds of Hope experiences in Rwanda. Disasters 26:288–301.

Consultative Group on International Agricultural Reseach (CGIAR). 2007. The system-wide information network for genetic resources (SINGER). 14 Mar. 2007. <http://singer.cgiar.org>.

Close, T.J., S.I. Wanamaker, R.A. Caldo, S.M. Turner, D.A. Ashlock, J.A. Dickerson, R.A. Wing, G.J. Muehlbauer, A. Kleinhofs, and R.P. Wise. 2004. A new resource for cereal genomics: 22K barley GeneChip comes of age. Plant Physiol. 134:960–968.

Dudoit, S., R.C. Gentleman, and J. Quackenbush. 2003. Open source software for the analysis of microarray data. Biotechniques 34:S45–S51.

East African Sweetpotato Project. 2006. ViazivitamuDB: Kenya sweetpotato germplasm database. 12 July 2007. <http://www.viazivitamu.org/index.php>.

Gichuki, S.T., M. Berenyi, D. Zhang, M. Hermann, J. Schmidt, J. Glössl, and K. Burg. 2003. Genetic diversity in sweetpotato [*Ipomoea batatas* (L.) Lam.] in relationship to geographic sources as assessed with RAPD markers. Genet. Resources Crop Evol. 50:429–437.

Gichuki, S.T., M. Berenyi, J. Schmidt, J. Glöbl, and K. Burg. 2000. Evaluation of genetic diversity of sweetpotato cultivars with RAPD and AFLP molecular markers. African Potato Assn. Conf. Proc. 5:19–21.

GNU Project. 2007. The GNU project. 14 Mar. 2007. <http://www.gnu.org>.

Huamán, Z. (ed.). 1991. Descriptors for sweet potato. International Board for Plant Genetic Resources, Rome.

Huamán, Z. and D.P. Zhang. 1997. Sweetpotato, p. 29–38. In: D. Fuccillo, L. Sears, and P. Stapleton (eds.). Biodiversity in trust. Cambridge Univ. Press, Cambridge, United Kingdom.

Hubbard, T., D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen, T. Down, R. Durbin, E. Eyras, J. Gilbert, M. Hammond, L. Huminiecki, A. Kasprzyk, H. Lehvaslaiho, P. Lijnzaad, C. Melsopp, E. Mongin, E.R. Pettett, M. Pocock, S. Potter, A. Rust, E. Schmidt, S. Searle, G. Slater, J. Smith, W. Spooner, A. Stabenau, J. Stalker, E. Stupka, A. Ureta-Vidal, I. Vastrik, and M. Clamp. 2002. The Ensembl genome database project. Nucleic Acids Res. 30:38–41.

Mann, H., D. Bedford, J. Luby, Z. Vickers, and C. Tong. 2005. Relationship of instrumental and sensory texture measurements of fresh and stored apples to cell number and size. HortScience 40:1815–1820.

Martin, S.L., B. Blackmon, R. Rajagopalan, T. Houfek, R. Sceeles, S. Denn, T.K. Mitchell, D.E. Brown, R. Wing, and R. Dean. 2002. 2002. MagnaportheDB: a federated solution for integrating physical and genetic map data with BAC end derived sequences for the rice blast fungus *Magnaporthe grisea*. Nucleic Acids Res. 30:121–124.

MySQL AB. 2007. MySQL developer zone. 14 Mar. 2007. <http://www.mysql.org>.

phpMyAdmin Project. 2007. The phpMyAdmin project. 14 Mar. 2007. <http://www.phpmyadmin.net>.

Rhodus, T. 2006. Open source solutions for extension communication. HortScience 41:932 (abstr.).

Sperling, L. 2002. Emergency seed aid in Kenya: some case study insights on lessons learned during the 1990s. Disasters 26:329–342.

Tacchini, E. 2007. Dadabik interfaces creator. 14 Mar. 2007. <http://www.dadabik.org>.

Villordon, A. and J. Franklin. 2005. Public domain computer programs for quantifying size and shape in horticultural research. HortScience 40:1028 (abstr.).

Yen, D.E. 1982. Sweetpotato in historical perspective, p. 17–30. In: R. Villareal and T. Griggs (eds.). Sweet potato: proceedings of the first international symposiumAsian Vegetable Research and Development Center, Tainan, Taiwan.