SHORT COMMUNICATION

Development of a Digital Sunflower Germplasm Catalog for the Query Based Retrieval of the Trait Specific Germplasm

Mangesh Yuwaraj Dudhe[*], Karusala Alivelu, Mulpuri Sujatha, Hari Prakash Meena, Paladugu Madhuri

Abstract

For the benefit of Indian breeders, a comprehensive germplasm catalog of 3126 accessions based on nine seed yield and yield-related traits of sunflower in collaboration with UAS Bangalore, ORS, Latur, and ICAR-NBPGR, New Delhi, India, has been published in 2018, which includes information, on the availability of sunflower accessions conserved and maintained at ICAR-Indian Institute of Oilseeds Research (ICAR-IIOR), Hyderabad, Telangana, India. Searching for trait-specific germplasm accessions would be difficult in the voluminous germplasm catalog as it contains information on 3126 large numbers of accessions. To ease this difficulty, we have developed a searchable query-based Sunflower Germplasm Catalog to facilitate easy and rapid retrieval of information on germplasm accessions in the form of Digital Sunflower Germplasm Catalog (DSGC), which can be accessed by remote users/researchers who do not have access to the physical form of the catalog. The query-based information system is based on Visual Basic.NET (VB.NET). This system generates the information by giving the required query, including more than one character, *i.e.*, a combination of characters.

Key words: Sunflower, accessions, catalog, query-based, information system

Introduction

Sunflower crop is one of the most important oilseed crops in India. Plant genetic resources (PGR) are important for any crop improvement program. The PGR includes primitive forms of cultivated plant species and landraces, modern cultivars, obsolete cultivars, breeding lines and genetic stocks, weedy types, and related wild species of crop plants. In their quest for certain genetic features, today's scientists and crop breeders need access to information on a broad range of crop varieties, landraces, and associated wild species. The mandate of the ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad, India, includes the cultivation of sunflowers as well as fundamental and strategic research aimed at improving sunflower production in India. To meet this objective, a large number of *Helianthus annuus* L. germplasm collection (3400 accessions) is maintained in medium-term storage at the institute repository (Anonymous, 2020). Through systematic characterization and large-scale evaluation of sunflower accessions, a comprehensive germplasm catalog of 3126 accessions has been developed which includes information on nine seed yield and yield-related traits and 24 DUS descriptors, and on the availability of sunflower accessions conserved and maintained at IIOR (Dudhe *et al.,* 2020a). As large numbers of sunflower accessions are being documented, the catalog is voluminous and requires skills to identify the trait-specific sunflower accession for new users and

breeders who are the ultimate beneficiaries of the catalog. Hence, searching for trait-specific germplasm accessions would be difficult owing to the large number of accessions. To ease this difficulty, we have developed a searchable query-based sunflower germplasm catalog to facilitate easy and rapid retrieval of information on germplasm accessions for remote researchers who do not have a catalog in the physical form. Digital Sunflower Germplasm Catalog (DSGC) is a search catalog providing information about sunflower accession collections maintained at ICAR-IIOR, Hyderabad, Telangana, India. This database provides access to information on all germplasm - augmented, developed, and conserved representing the diversity available within the sunflower collection in IIOR.

The available information on passport data of 3126 sunflower accessions in an Excel sheet was collected on nine yield and yield-related traits, which were utilized to prepare DSGC. The sunflower germplasm data was collected from the IIOR germplasm unit and compiled in Excel (Dudhe *et al.,* 2018; Dudhe *et al.,* 2020a; Dudhe *et al.,* 2020b). For data digitization, a database was developed using MS Access. The MS-Access database was further converted into an SQL server database using Visual Basic.NET. The data is stored in MySQL database as tables and integrated with PHP (Hypertext PreProcessor). The XAMPP Database server was used to run the PHP code and was installed (Figure 1). The main database tables have different columns for different characters. Each record in the database comprises 10 fields (name and 9 traits passport data) which represent individual germplasm accession with different quantitative traits. Based on the characterization and evaluation of all 3126 sunflower accessions, the database was prepared in CD (compact disc format) and distributed to all the AICRP breeders for effective utilization of promising inbreds and germplasm accessions in the breeding program. Various traits included in the DSGC are the name of the accession, seed yield (g/plant), oil content (%), days to 50% flowering, days to maturity, number of leaves, head diameter (cm), plant height (cm), seed weight (g). Since the accession number and name of the accession fields have no duplicate values, those are used for unique identification records. The final value is stored in the form of quantitative values of descriptors.

Information Retrieval (IR) system finds the relevant documents from a large dataset according to the user query. Queries submitted by users to search engines might be ambiguous, and concise and their meaning may change over time. As a result, understanding the nature of the information that is needed behind the queries has become an important research problem (Khin and Yee, 2018). Hence query-based retrieval system is important to locate the specific object or value or useful information in a short time or with one click rather than spending a long time in a physical search. Here, we have used Visual Basic a

basic program language to create a user-friendly DSGC using graphical interface tools like text box, combo box, list box, command buttons, and embedded pictures.

Originally Visual Basic called Visual Basic.NET (VB.NET), is a multi-paradigm, object-oriented programming language, implemented on .NET, Mono, and the .NET Framework.

To access DSGC, the user has to register and create his/her account on the database (https://icar-iior.org.in/apps/sunflower/) by clicking the above link which will open the home page of the Sunflower germplasm information system. The breeder has different options to choose the mode of data representation depending on their need. The search option is meant for searching the desired accession with the numerical value of its desired characters. It includes a search by country of origin, by GMU number, and by descriptors. All the output can be exported to Excel by selecting that option.

Once registered through confirmation mail, the registration will be authenticated which completes the first step. Using login credentials provided to the user, he/she can enter into the DSGC. At a time, many users can operate the DSGC. A researcher interested in selecting a particular sunflower germplasm accessions may use different options to choose for the retrieval of the data. Complicated queries have been created using the AND or OR operator for getting information on more than one parameter. The user may opt for the output in the form of quantitative data or the form of a graphical representation. The default parameter settings can be performed if a user is interested in retrieving all the
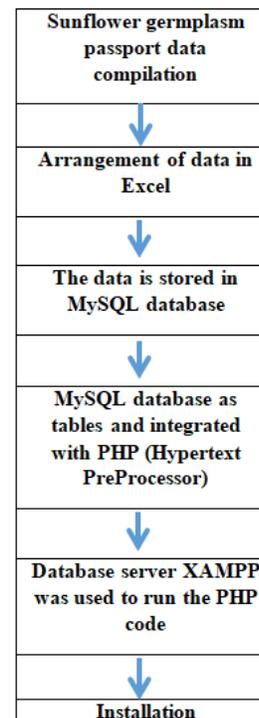

Fig.1: Flow diagram of the development of Digital Sunflower Germplasm Catalog (DSGC)

| 1. cover page of DSGC | 2. login page of DSGC | 3. list of accessions |
| 4. traits listed in DSGC | 5. seed yield trait inDSGC search (>25 g) | 6. seed yield trait in DSGC search and other traits |

Fig. 2: A few screenshots of the DSGC

Table 1: Example query-based results for the search conducted with three parameters

Sunflower high oil content lines (>40%)		Sunflower high seed yield/plant lines (>40 g/plant)		Sunflower-early maturity lines (< 45 days)	
Accession	Oil content	Accession	Seed yield/plant	Accession	Days to 50% flowering
GP6-142	42.0	GMU-363	40.3	GMU-130	43
GP6-176	41.3	GMU-768	44.0	GMU-935	44
GP6-207	40.8	GMU-960	43.2	GMU-1195	45
GP6-282	40.6	GMU-1085	45.5	GP6-891	34
GP6-339	40.1	GP6-126	82.3	EC-601859	31
GP6-451	40.2	GP6-1016	48.3	DRSI-125	44
GP6-564	41.2	GP6-1254	40.2	SCG-36	45
GP6-1035	42.0	GP6-1420	81.3	SCG-77	45
GP6-1076	40.1	ID-12	49.7	EC-625707	43
GP6-1088	40.3	ID-28	44.0	EC-625737	43
GP6-1108	40.4	ID-34	43.2	EC-625785	43
GP6-1233	40.4	SCG-23	43.8	IC-502023	45
GP6-1407	40.2	SCG-39	44.1	IC-502033	42

Table 2: Information on more than one parameter seed yield (>40g/plant) and oil content (>35%)

Accession	Seed yield / plant	Oil content (%)	Seed weight (g)	Head diameter (cm)	Plant height (cm)	No of leaves	Days to maturity	Days to 50% flowering
GMU-363	40.3	36.1	6.2	16.6	203	29	88	58
GMU-960	43.3	36.8	5.5	17.6	176	27	86	56
GMU-1085	45.5	37.1	3.3	15.2	130	22	85	55
GP6-54	45.2	37.5	5.3	14.3	202	21	95	66
GP6-709	48.5	36.8	5.5	13.4	149	26	90	61
GP6-1254	40.2	37.8	4.6	14.1	208	28	96	66

accessions at a time. Similarly, a combination of characters can be used in a query search based on more than two or three features to find important trait-specific sunflower accession according to the needs and breeding goal. Table 1 displays the query results for the search conducted, which included the identification of sunflower accessions with high oil content (>40%), high seed yield/plant (>40 g/plant), and early maturity (< 45 days) lines. In Table 2, information on more than one parameter *i.e.* seed yield (>40 g/plant) and oil content (>35%) was searched and the identified sunflower accessions are listed.

Figure 2 shows a few screenshots from the Digital Sunflower Germplasm Catalog (DSGC). The systematically categorized germplasm will assist users in identifying the material and choosing new germplasm that will serve a variety of functions, such as serving as promising parents for breeding programs. This information may be useful to the researcher before the actual utilization of the selected accessions through DSGC in their sunflower improvement program. The National Bureau of Plant Genetic Resources (ICAR-NBPGR), New Delhi, has developed the PGR Portal and PGR database management system (http://www.nbpgr.ernet.in/PGR_Databases.aspx; Jacob *et al.*, 2015). Through the PGR Portal, all the germplasm accessions conserved and maintained at ICAR-NBPGR can be searched with different parameters and information on the desirable germplasm can be retrieved. Genesys is an online platform where interested researchers can find information about Plant Genetic Resources for Food and Agriculture (PGRFA) conserved in genebanks worldwide (Guzzon and Ardenghi, 2018). A similar PGR database is maintained at ARS, USDA, USA (USDA, 2022).

In conclusion, we have developed a database called DSGC which is easily accessible to all the researchers who may be interested in sunflower genetic resources that are available in the ICAR-IIOR, India, repository and do not have access to the sunflower germplasm catalog in the physical form. The passport data and the information associated with the particular accessions can be retrieved by using different search parameters and through NBPGR, New Delhi, an indent can be submitted to the corresponding scientist.

References

Anonymous (2000) Annual report sunflower (2019-20), ICAR-Indian Institute of Oilseeds Research, Rajendranangar, Hyderabad-500 030.

Dudhe MY, Sujatha M, Meena HP, Alivelu K, Ghodke MK, Shadakshari YG, Tyagi RK, Radhamani K, Ranganatha ARG, Varaprasad KS, Reddy AV (2018) Germplasm catalog of Sunflower (*Helianthus annuus* L.), Vol 1. ICAR-Indian Institute of Oilseeds Research, Hyderabad.

Dudhe MY, Sujatha M, Meena HP, Ghodke MK, Shadakshari YG, Radhamani J, Alivelu K, Tyagi RK, Ranganatha ARG, Varaprasad KS and Vishnuvardhan Reddy A (2020a) Sunflower germplasm catalog: Necessity and benefit to sunflower researchers. *J. Oilseeds Res.,* 37 (Special issue): 75-76.

Dudhe MY., Mulpuri S, and HP Meena et al. (2020b) Genetic variability, diversity and identification of trait-specific accessions from the conserved sunflower germplasm for exploitation in the breeding programme. *Agric Res* 9, 9-22. https://doi.org/10.1007/s40003-019-00406-w.

Guzzon Fand Ardengh NMG (2018) Could taxonomic misnaming threaten the ex situ conservation and the usage of plant genetic resources?. Biodivers Conserv 27, 1157-1172 . https://doi.org/10.1007/s10531-017-1485-7.

Jacob SR, Singh N, Srinivasan K, Gupta V, Radhamani J, Kak A, Pandey C, Pandey S, Aravind J, Bisht IS and Tyagi RK (2015) Management of Plant Genetic Resources, National Bureau of Plant Genetic Resources, New Delhi, 323 p.

Khin NT W, Yee N N (2018) "Query Classification Based Information Retrieval System,". International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), pp. 151-156,. doi: 10.1109/ICIIBMS.2018.8549988.

USDA Agricultural Research Service (ARS) (2022) Germplasm Resources Information Network (GRIN). USDA Agricultural Research Service.

Visual Basic .NET (VB.NET): https://en.wikipedia.org/wiki/Visual_Basic_.NET.

Yin Z, Lan H, Tan G, Lu M, Vasilakos AV, Liu W (2017) Computing platforms for big biological data analytics: perspectives and challenges, *Computational Struct. Biotech. J.* 15, 403-411.