

# Seed Database Manager: an open-source tool for germplasm information management

Sam E. Schafer<sup>1\*</sup> and Diane R. Wang<sup>1\*</sup>

<sup>1</sup> Purdue University, United States \* These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [Abhishek Tiwari](#)

## Reviewers:

- [@RobLBaker](#)
- [@jsun](#)
- [@ahmccormick](#)

Submitted: 01 May 2025

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

Record-keeping of plant germplasm in smaller research teams relies on one of several approaches: digital spreadsheets or paper records, commercially available software for database management, or independently developed solutions. The latter two require additional capital or time investment to properly utilize. The Seed Database Manager (SDM) is an open-source Shiny R application designed to provide an affordable, easy-to-use, and accessible method to store, browse, and modify plant germplasm information. The set-up process for the SDM enables deployment either locally or through any hosting platforms that support Shiny R ([Chang et al., 2024](#)) applications and PostgreSQL (PostgreSQL Global Development Group, <https://www.postgresql.org>) databases with minimal modifications made by the user. Upon successful deployment, a simplified user interface allows researchers to browse existing entries, enter in new information, and append new information to entries post-hoc. Each seed packet registered to the database is associated with a unique, customizable tag that is intended to help simplify the relationship between entries while also organizing the database chronologically. Code for the SDM is available on GitHub under the Non-Profit Open Software License 3.0 and can be further modified to fit researcher needs.

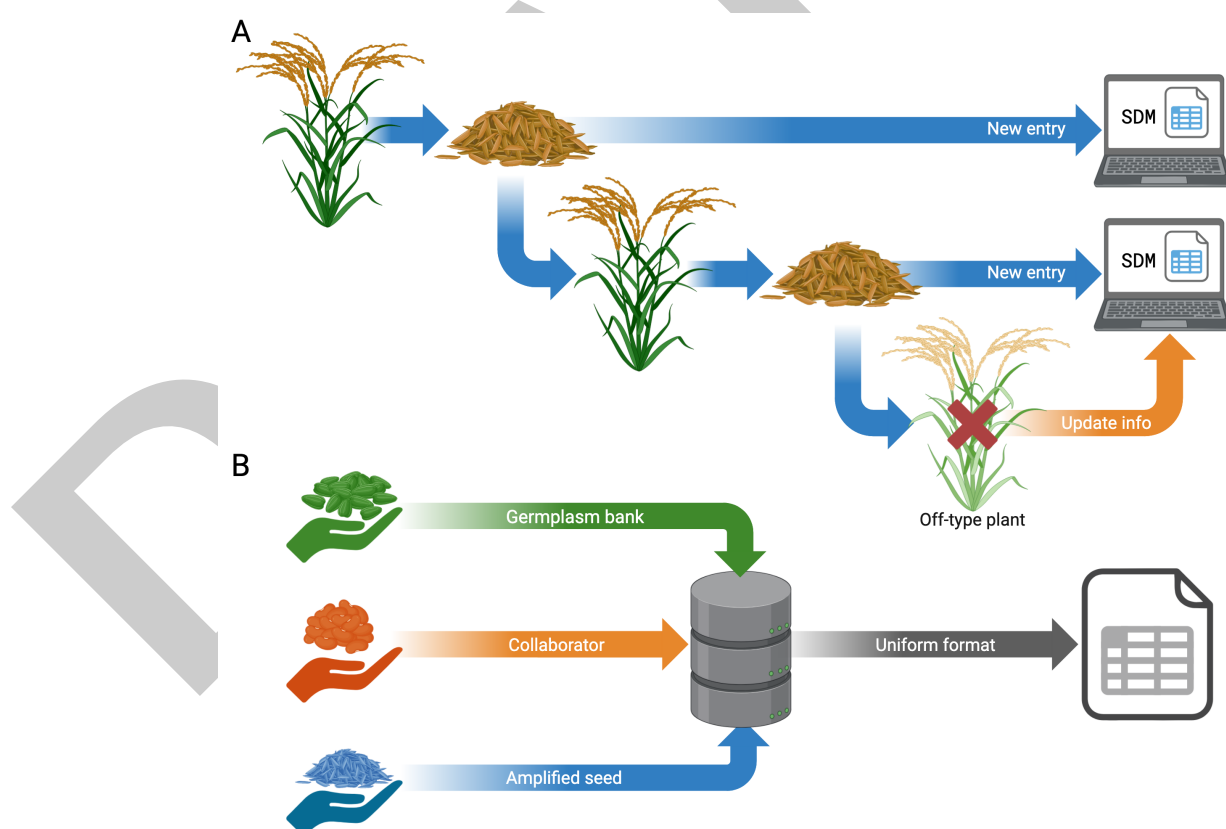
## Statement of need

Research conducted within plant sciences often involve exchange and amplification of seed. A single research group may handle several plant species in addition to numerous distinct varieties within each species. It is critical that regular records of germplasm be made to keep track of germplasm relationships, prevent information gaps associated with staff turnover, and to monitor the progress of the program ([Tiwari et al., 2023](#)). Larger organizations that maintain germplasm collections have their own strategies and specialized collection management systems that are often not applicable to smaller-scale research teams, such as those found at research universities. Existing tools, such as commercially available software, specially designed data management systems, or usage of digital or paper spreadsheets, are not always be easily accessible, require additional capital investment, or are better suited for larger-scale breeding programs. Open-source systems like the Breeding Information Management System ([Jung et al., 2021](#)), Breedbase ([Morales et al., 2022](#)), Germinate ([Shaw et al., 2017](#)), and the discontinued International Crop Information System (ICIS) ([Portugal et al., 2007](#)) require significant technical knowledge to initiate or were intended for larger plant breeding operations and are not ideal for simple germplasm record-keeping. Additionally, acquisition of germplasm from separate sources (e.g., retrieval from external sources such as collaborating groups and nationally maintained gene banks, or generation internally through propagation of seeds), utilize different naming conventions and introduces an extra layer of complication for archiving information as a result. Lastly, due to individual needs among different research groups, it is important that existing tools be easily tailored to better serve the needs of each group. The

development of an accessible and adaptable open-source seed management tool would greatly enhance the capacity of plant science research groups to track their seed collections.

The Seed Database Manager (SDM) is a database-navigating application developed in Shiny R that provides an easy-to-use, centralized, and accessible platform to store all relevant germplasm information. The SDM aims to resolve the germplasm organization problem by adapting each entry into a uniform format that archives information on species, common name and any associated IDs, proximal source of the seed, harvest date, relevant researcher, and other descriptive information. Following submission into the database, each entry receives its own unique, chronological, and customizable identifying tag. This format allows the database to be as flexible as possible to account for different types of information given within each set of germplasm, as well as simplifies pedigree tracking down to the tracking of entry tags. The SDM is designed to connect to a PostgreSQL database through the utilization of the Rpostgres (Wickham et al., 2024) package, which can be easily configured to run via local machines or through deployment onto an external hosting platform, greatly enhancing the accessibility of the stored data. Additionally, all database-modifying operations are restricted to authorized users determined at the set-up process, allowing for a more secure platform. All functionality of this open-source software is achieved using R programming, which is accessible to biologists within and outside of the realm of plant sciences.

## Figures



**Figure 1:** A. Example pipeline for information tracking of amplified seed using the SDM, wherein the discovery of new information, such as the observation of off-type plants, is used to update previous entries. B. Information of seed packets received from various sources are conformed to a uniform format ahead of entry into the database.

## Acknowledgements

We thank all the members of the Wang lab who helped test this application from its inception to release and Jason King for contributing code related to database connections. Development of this application was supported by grants to D.R.W. (USDA AFRI #2022-67013-36205 and NSF-PGRP #2102120).

## References

- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). *Shiny: Web application framework for R*. <https://doi.org/10.32614/CRAN.package.shiny>
- Jung, S., Lee, T., Gasic, K., Campbell, B. T., Yu, J., Humann, J., Ru, S., Edge-Garza, D., Hough, H., & Main, D. (2021). The breeding information management system (BIMS): An online resource for crop breeding. *Database*, 2021, baab054. <https://doi.org/10.1093/database/baab054>
- Morales, N., Ogbonna, A. C., Ellerbrock, B. J., Bauchet, G. J., Tantikanjana, T., Tecle, I. Y., Powell, A. F., Lyon, D., Menda, N., Simoes, C. C., & others. (2022). Breedbase: A digital ecosystem for modern plant breeding. *G3*, 12(7), jkac078. <https://doi.org/10.1093/g3journal/jkac078>
- Portugal, A., Balachandra, R., Metz, T., Bruskiewich, R., & McLaren, G. (2007). International crop information system for germplasm data management. *Plant Bioinformatics: Methods and Protocols*, 459–471. [https://doi.org/10.1007/978-1-59745-535-0\\_22](https://doi.org/10.1007/978-1-59745-535-0_22)
- Shaw, P. D., Raubach, S., Hearne, S. J., Dreher, K., Bryan, G., McKenzie, G., Milne, I., Stephen, G., & Marshall, D. F. (2017). Germinate 3: Development of a common platform to support the distribution of experimental data on crop wild relatives. *Crop Science*, 57(3), 1259–1273. <https://doi.org/10.2135/cropsci2016.09.0814>
- Tiwari, A., Tikoo, S. K., Angadi, S. P., Kadaru, S. B., Ajanahalli, S. R., & Vasudeva Rao, M. (2023). Germplasm management in commercial plant breeding programs. In *Market-driven plant breeding for practicing breeders* (pp. 33–68). Springer. [https://doi.org/10.1007/978-981-19-5434-4\\_2](https://doi.org/10.1007/978-981-19-5434-4_2)
- Wickham, H., Ooms, J., & Müller, K. (2024). *RPostgres: C++ interface to PostgreSQL*. <https://doi.org/10.32614/CRAN.package.RPostgres>