

Utah State University

DigitalCommons@USU

Funded Research Records

Data Services

5-23-2022

Social and Ecological Drivers of Life History Evolution in Wild Bees

Karen Kapheim

Utah State University, karen.kapheim@usu.edu

Follow this and additional works at: https://digitalcommons.usu.edu/funded_research_data



Part of the [Biology Commons](#)

Recommended Citation

Kapheim, Karen, "Social and Ecological Drivers of Life History Evolution in Wild Bees" (2022). *Funded Research Records*. Paper 227.

https://digitalcommons.usu.edu/funded_research_data/227

This Grant Record is brought to you for free and open access by the Data Services at DigitalCommons@USU. It has been accepted for inclusion in Funded Research Records by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



DATA MANAGEMENT PLAN

Proper management of data will be key to the success of this project. The PI's lab has regular discussions of the primary literature in data science and data management. In addition to the descriptions below, an outline of the proposed management of data acquired as part of this project is included in Table 1. The PI will coordinate with team members to ensure this data management plan is carried out in full.

1. Data Types

Raw data

Specimen data. All bee data will be captured in hard copy paper worksheets designed to be filled-in during the field collections and dissections. Each bee will be given a unique identifying code, and this code will accompany all records, including sample tubes stored in the -80°C freezer. All worksheets will be scanned and archived in USU Box after the data has been transcribed into spreadsheets, and checked for accuracy. USU Box is backed up to the cloud instantly, and provides version control.

Video data. Videos of bee behavior will be recorded directly to video cameras with built-in hard drives. These will be saved as MTS files on the lab-owned storage space at the University of Utah High Performance Computing Center (UCHPC). I currently have 20 TB of unused storage space. The data storage nodes require authentication for access, and data transfers are encrypted. All team members will be able to access these data upon request. These files will be backed up to hard drive.

Sequence data. DNA and RNA sequences will be received as digital files in the FASTA or FASTQ format via Secure File Transfer Protocols (SFTP), and kept on the lab-owned storage space at the UCHPC, following the same protocol as described for digital data.

Curriculum materials

Lesson plans. The PI will develop new curriculum to incorporate into ongoing courses. This curriculum will focus on bioinformatics analysis, and will be created in Markdown (MD) format on GitHub. GitHub provides version control and is backed up daily to the cloud.

Intern posters. Students that participate in the NASMP 10-week fellowship will produce posters to present at the annual SACNAS conference. The PDF of these posters will be stored in USU Box.

Other materials

Analytical workflows. This research will generate a public repository of data analytical workflows made publically and permanently available in GitHub and the Harvard Dataverse.

2. Data and Metadata Standards

The PI will follow the NSF Public Access Plan recommendations and “FAIR (Findability, Accessibility, Interoperability, and Reusability) guiding principles for scientific data management and stewardship” as described by Wilkinson et al. 2016 *Scientific Data*. The PI will use a dedicated project within the Open Science Framework (OSF) as the central hub for all data, metadata, workflows, and communication about all aspects of the project. The PI will link this OSF space directly to storage and archival locations.

Findability. All data and metadata generated from this project will persist with unique identifiers in the Harvard Dataverse, NCBI Sequence Read Archive (SRA), or USU Digital Commons. Data can be easily searched, accessed, and cited from each of these repositories.

Accessibility. Data and metadata will be openly accessible at the Harvard Dataverse, GitHub repository, NCBI SRA, or USU Digital Commons. Data may be embargoed until publication or 3 years after the end of the funding period, whichever comes first.

Interoperability. The project OSF will include instructions on standards for file naming, spreadsheet organization, file management, and coding, and these will be based on recommendations from the USU Research Data Management Systems. When archiving data, the PI will conform to metadata standards of each repository. Guidance on appropriate vocabulary will be sought from Biosharing.org.

Reusability. Each of the archival repositories we have selected perform validity checks and version control, thus enhancing long-term reusability. Metadata will accompany all data, and workflows written in markdown syntax will accompany all descriptions of data analysis.

3. Access and Sharing Policies

All data, metadata, and data analytical workflows generated from the research will be publically and permanently accessible in the Harvard Dataverse (all data), a GitHub repository designated for this project (analytical workflows and lesson plans), NCBI SRA (sequence data), or USU Digital Commons (student presentations). These data are not subject to protection of privacy, confidentiality, security, or intellectual property rights.

4. Re-use and Re-distribution Policies

Data will be made freely available for replication under the CC-attribution license.

5. Data Archiving Plans

Specimen data. All specimens used for DNA, RNA, or physiological sampling and the materials generated from these samples will be stored indefinitely in a -80°C freezer in my laboratory, equipped with a low/high temperature alarm and a backup CO₂ system that becomes activated in the event of high temperatures. Voucher specimens of wild caught bees will be deposited at the University of Panama or the USDA-PIRU bee collection when appropriate. Specimen data will be archived in the Harvard Dataverse, which is contained in a physically secured, continually-monitored facility, and system backups are made on a daily basis.

Video data. MTS files will be archived at the Harvard Dataverse.

Sequence data. FASTA and FASTQ files will be archived at NCBI.

Lesson plans. Lesson plans will be archived from GitHub via Zenodo.

Student posters. Student presentations from SACNAS will be archived at USU Digital Commons.

Analytical workflows. MD files will be archived at the Harvard Dataverse and Zenodo.

Table 1 Summary of data management plan.

Data type	Format	Storage location	Backup schedule	Maintained	Archive location
Bees	n/a	-80°C freezer	CO ₂ backup	Indefinitely	Voucher specimens
Data collection sheets	PDF	USU Box	Daily	Minimum 5 years past the end of the project	n/a
Specimen data	spreadsheets				Dataverse
Video data	MTS	UCHPC	Monthly		Dataverse
Sequence data	FASTQ				NCBI SRA
Analytical workflows	MD	GitHub	Daily		Zenodo, Dataverse
Lesson plans	GitHub			Zenodo	
Student posters	PDF	USU Box		1 year after presentation	USU Digital Commons