

3-15-2017

Collaborative Research: The interplay between fluctuating selection on symbionts and life cycle evolution in sap-sucking insects

Carol von Dohlen

Utah State University, carol.vondohlen@usu.edu

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

 Part of the [Biology Commons](#)

Recommended Citation

von Dohlen, Carol, "Collaborative Research: The interplay between fluctuating selection on symbionts and life cycle evolution in sap-sucking insects" (2017). *Funded Research Records*. Paper 13.

https://digitalcommons.usu.edu/funded_research_data/13

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



Data Management Plan

All researchers associated with this project will be given training by the PIs in professionalism and ethics specifically relating to the importance of scientific integrity, honesty, and standards expected regarding experiment planning and execution as well as data collection and storage.

Data and Materials Produced

Nucleic acid samples (Genomic DNA and RNA): Genomic DNA will be extracted for at least 9 new samples. Genomic DNA will be quantified using a fluorometric quantitation and quality examined on an agarose gel. RNA will be extracted for at least 12 new samples. RNA will be quality checked and quantified with an Agilent 220 Tape Station.

Sequence data: The sequence data generated in this project will be Illumina sequence reads, expression data, and sequence assemblies (transcriptome and genome). Raw sequence data will be in FASTQ format as this is the standard output for Illumina sequencing. A minimum of 7 Illumina sequencing runs will be generated (genome data and RNAseq) while completing project activities. Data will be captured at the Georgia Genomics Facility.

Quantitative PCR (qPCR) data & cell counts: Raw qPCR result files will be generated on a qPCR machine at UGA. Cell counts will be done manually under a microscope at UGA.

Curriculum materials: All researchers participating in this project will be required to keep laboratory notebooks recording daily activities, experimental protocols, plans, and results, that are inspected on a regular basis and will be kept in the PI's lab. Data will be generated from broader impacts evaluation (e.g., surveys, etc. from undergraduate and high school laboratory experiences).

Standards, Formats and Metadata

Nucleic acid samples (Genomic DNA and RNA): These samples will be labeled according to voucher IDs and sample concentrations, and stored in the -80° freezer of PI von Dohlen.

Sequence data: Sequence data will be maintained in standard formats that allow for reproducibility and ease of sharing. Raw files can be up to 100Gb in size; the volume will be low enough that data can be stored on desktop computers with good storage capacity. However, data will be backed up on servers at both USU and UGA. Raw sequence read files will be kept in zipped format (.gz) located in folders with standardized directory naming. Expression data will be stored as reads mapped to reference genomes in BAM format (.bam), and also summarized in tables as FPKM values in excel or raw text formats. Transcriptome and genome sequence assemblies will be saved as compressed sequence files (.fasta.gz) along with intermediate files used to generate the assemblies. All insect voucher samples will be COI barcoded, and barcode sequences will be saved as raw (.abi) and consensus sequence files (.fasta).

Quantitative PCR (qPCR) & cell counts: Raw qPCR result files will be saved on the computer that runs the qPCR machine, and analyzed data will be saved in excel files and tables generated by the statistical program JMP. Cell count data will be stored in Excel files with correct sample attributions. The PIs will train staff and students in documentation and data management when they join the project; conventions for directory and file naming will be used for all qPCR and count data files.

Curriculum: Laboratory notebooks will be hard copies. Electronic files will be Word and Excel docs.

Metadata associated with each sample profiled by sequencing or qPCR will be recorded in laboratory notebooks and computer "Readme" files, and associated to samples by sample names and date labels. Minimally, metadata required by the National Center for Biotechnology Information (NCBI) will be recorded, including sample name, species name, date isolated, and isolation source, as well as details of experimental projects such as the sequencing instrument used and read length, etc.

Roles and Responsibilities

PIs von Dohlen and Burke will be responsible for ensuring that sequence and genome data are securely stored on USU and UGA cloud storage space through Google Drive or Box. Both PIs will be responsible

for ensuring that finalized genome and transcriptome data are properly uploaded to the NCBI database. PI von Dohlen and graduate student will also be responsible for ensure genomic DNA is properly stored in the -80°C freezer, and vouchers in 80% ethanol. As a Director of USU Insect Collection, PI von Dohlen will be responsible for any requested loans of genomic DNA and voucher specimens. Any publications produced from this research will include the respective database numbers from NCBI to ensure that the data has been properly archived there. Once original personnel are no longer available the director or curator of the USU Insect Collection will have long-term responsibility for decisions about the data remaining at USU. NCBI will have long-term responsibility for data included in their database.

Dissemination Methods

Sequence data and annotated genomes and transcriptomes uploaded to GenBank will be in the public domain. PI von Dohlen's Insect Systematics and Evolution laboratory is associated with the USU Insect Collection and all DNA samples will be treated similarly to museum specimens. Loans can be requested through the museum curator. Data will be submitted to NCBI before publication, but will be made publicly available when manuscripts describing the data are published to ensure high quality of information. Sequence datasets will be made publicly available by submission to the NCBI according to their instructions. Transcriptome assemblies will be submitted to the NCBI Transcriptome Shotgun Assembly (TSA) database, and genome assemblies will be submitted to the Whole Genome Shotgun (WGS) database. qPCR raw data will be available by request from the PIs. Open-access journals will be targeted for publications resulting from this work to ensure continued and easy access to these data.

Policies for Data Sharing and Public Access

The data produced in this project will not have permission restrictions placed on it. Other researchers working with bacterial genomes or symbionts may be interested in using this data in the future and will be free to do so. To the extent possible, relevant data will be made publicly available in peer-reviewed publications. Open-access journals will be targeted for publications resulting from this work to ensure continued and easy access to these data. Any other nucleic acid samples or raw qPCR data will be made available upon request from the PI. No ethical provisions are necessary for the projects described in this proposal. Any intellectual property provisions that become necessary within the award period will be negotiated with the USU and UGA Research Foundations.

Archiving, Storage and Preservation

Nucleic acid samples (Genomic DNA and RNA): Genomic DNA and RNA will be stored in the -80° C freezer, and vouchers in 80% ethanol in PI von Dohlen's lab. Once original personnel are no longer available, the director or curator of the USU Insect Collection will have long-term responsibility for decisions about the samples remaining at USU.

Sequence data: Minimally, all raw files as described above will be stored on hard drives within the lab, as well as being backed up with automatic backups (at least weekly) on an external hard drive and instantly off-site with the cloud-based CrashPlan program, as well as USU cloud storage space through Google Drive or Box. Raw data will also be backed up on library servers at USU. Data will be stored for a minimum of 7 years after the date of data publication. NCBI will have long-term responsibility for data included in their database.

Quantitative PCR (qPCR) data & cell counts: Data files along with metadata will be uploaded to the USU library server for long-term storage.

Curriculum materials: Laboratory notebooks will be kept in the PI's lab at the conclusion of the project and for at least 7 years thereafter. Any data generated from broader impacts evaluation (e.g., surveys, etc. from undergraduate and high school laboratory experiences) will be stored on both PIs computers and backed up on library servers. The data will be stripped of any personal/confidential identifiers.