Plan Overview

A Data Management Plan created using DMPTuuli

Title: Origin and evolution of the hexaploid African nightshade genome

Creator: Péter Poczai

Principal Investigator: Péter Poczai

Data Manager: Péter Poczai

Affiliation: University of Helsinki

Funder: The Research Council of Finland (former The Academy of Finland)

Template: Academy of Finland data management plan guidelines

ORCID iD: 0000-0002-0107-1068

Project abstract:

Orphan crops are plant species used around the world for food and medicine contributing to the livelihoods or producers, but which have not been included in mainstream agricultural research and development agendas. Rapid advancement in genomics provide an unprecedented opportunity to accelerate breeding of such orphan crops, which hold the potential to improve food and nutritional security in the world. African nightshade (Solanum scabrum) is a hexaploid species of the mega-diverse Solanum genus, which has received less attention than other major solanaceous crops originating from the New World, e.g., potato and tomato. Its leaves and berries are the source of coloring plant extracts, inks and dyes, and they are rich in proteins, fibers, iron, vitamins and amino acids. It is also a valuable genetic resource for plant breeding because of its resistance to biotic (e.g. late blight) and abiotic stresses (e.g. drought). However, basic knowledge in this species about its origins is still lacking as well as necessary genomic resources, which are required for its successful breeding. This project will deliver the reference genome sequence and transcriptome atlas of S. scabrum, which will be pivotal to set the baseline for African nightshade breeding by providing opportunities for increasingly detailed analysis of genetic diversity and accelerating trait development in breeding programs. Using genomics data, we will aim to achieve a better understanding regarding the population structure and domestication history of African nightshade germplasm. Such information will be highly beneficial for the development of commercial value chains improving food security, nutrition, livelihood and contribute to income generation in Europe and marginal areas of Africa. Our project will also fill fundamental knowledge gaps by providing the first comprehensive phylogenomic framework about the evolution of African nightshade and elucidate parental origins of its polyploid genome.

ID: 12282

Last modified: 28-04-2021

Copyright information:

Origin and evolution of the hexaploid African nightshade genome

1. General description of data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Also give a rough estimate of the size of the data produced or collected?

Nucleotid and amino acid sequences and alignments, phylograms, herbarium specimen and plant metadata, and species descriptions including drawings and photos. DNA-sequences are stored in standard fasta format (.fasta or .fastq), morphometrics tab-delinated text (.txt), alignments in nexus-format (.nxs) and trees in newick-format (.nwk) . Chromatograms are in ab1 format. Specimen metadata will be stored in university administered public databases with backups in personal MS Access databases. Data produced by this project is estimated to be ~1TB in size.

1.2 How will the consistency and quality of data be controlled?

Electronic lab journal (eLabFTW) and chromatogram files will be maintained throughout the project documenting every step of DNA sequence production. Everyone handling the data has been introduced to the best practices (IRIDA platform for sequencing data managment).

2. Ethical and legal compliance

2.1 What ethical issues are related to your data management, for example, in handling sensitive data, protecting the identity of participants, or gaining consent for data sharing?

No ethical problems are associated with the non-material data collected during the project. As regard to material samples, the project studies plant samples that have been already submitted to public herbaria or genbank collections with high standards and ethical guidelines based on the Nagoya protocol. Mandatory Standard Material Transfer Agreements (SMTA) are available for all accession used in the project. Herbaria and genebank collections have well-established codes of conduct and standards for exchanging specimens, and this network will be relied on for providing the research material.

2.2 How will data ownership, copyright and IPR issues be managed? Are there any copyrights, licences or other restrictions that prevent you from using or sharing the data?

The project does not deal with copyrighted or patented data. Open Access journals are preferred when publishing the results and whenever not possible hybrid OA publishing.

3. Documentation and metadata

3.1 How will you document your data to make them findable, accessible, interoperable and reusable for you and others? What kinds of metadata standards, README files or other documentation will you use to help others understand and use your data?

Each of the data repositories (NCBI-GenBank and Sequence Read Archive) with the exception of Dryad have their own strict formats for the data and associated information, and those will be naturally followed during the submission. When submitting files to Dryad, standard phylogenetic formats will be used: nexus for alignments, newick for phylograms.

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will they be backed up?

Backups of the data will be created each day in the university server to prevent data loss when the data is modified. CSC longterm storage facilities are utilized if server space will be an issue.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

No sensitive data is associated with this project - normal IT safety measures to keep data from being stolen before publication are followed by all team and project members.

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or their metadata, be made available?

Herbarium data and species information will be made publicly available:
1. Finnish Museum of Natural History public database (herbarium specimens and collecting data - submitted upon accession)
http://luomus.fi/en/botanical-and-mycological-collections
Raw sequence data will be published and uploaded to:
2. One of the databases of the Interionational Nucleotide Sequence Database Collaboration (DNA, RNA and protein sequence data - upon article publication)
http://www.insdc.org/
Genome assemblies will be publicly available:
4. NCBI Genome Database (annotated genomes - upon article pubication)
https://www.ncbi.nlm.nih.gov/genome
5. Sol Genomics Network https://solgenomics.net/
Phylogenetic data will be deposited and open to public:
6. TreeBASE and Dryad Digital Repository (phylograms, DNA sequence alignments - upon article publication)
https://treebase.org/, http://datadryad.org/
All data will be made available by the end of the project and when reaching major deliverables of the work packages detailed in the time managment plan of the project.

5.2 Where will data with long-term value be archived, and for how long?

Large-scale public databases have been developed and maintained for long-term data storage. All of the above-mentioned data have long-term value and will be made public upon accession to collections (herbarium and collection metadata) or article publication (other types of data).

6. Data management responsibilities and resources

6.1. Who will be responsible for specific tasks of data management during the research project life cycle? Estimate also the resources (e.g. financial, time and effort) required for data management.

Data managment will be carried out by an expert team of LUOMUS and lead by the PI. Resources and the time needed is considerable for the project but also standard procedure in specimen-based biology. My institution is well prepared for these tasks and their managment with strict internal organization and planing. As a curator of plant collections with 8 years of experties and widely published taxonomist with 17 years of experties I am well aware of the efforts of dealing with metadata related to the project.