Breeding the next animal protein source: how genomics will take us to the next step in insect breeding

A Data Management Plan created using DMPonline.be

Creators: Roel Meyermans ⓘ https://orcid.org/0000-0002-5079-5097 , n.n. n.n.

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Roel Meyermans ⓘ https://orcid.org/0000-0002-5079-5097 , n.n. n.n., n.n. n.n.

Data Manager: Roel Meyermans ⓘ https://orcid.org/0000-0002-5079-5097 , n.n. n.n., n.n. n.n.

Project Administrator: Roel Meyermans ⓘ https://orcid.org/0000-0002-5079-5097 , n.n. n.n., n.n. n.n.

Grant number / URL: PDMT2/23/035

ID: 203337

Start date: 01-10-2023

End date: 30-09-2024

Project abstract:

PDM Type 2 grant from KU Leuven IOF.

In preparation of FWO postdoc grant proposal on "Breeding the next animal protein source: how genomics will take us to the next step in insect breeding". In the current PDM year 2023-2024, a pilot will be set up for the project described below (subject of FWO postdoc proposal). A select set of BSF sequences will be generated in preparation of this FWO project, and will allow the setup of this new research line. The current DMP describes the full research line (including the lc-WGS imputation pipeline and the generation of all biological samples, phenotypes, and (imputed) sequences) that will be covered in the FWO project.

Our research group is currently active in genomic research of growth trajectories, resilience, genetic disorders, disease susceptibility and genetic diversity in several livestock and pet species. However, to meet the goals for sustainable food security, there is an increasing need for alternative protein
sources, one of them being insect production. Genetic and genomic research in insects is lagging behind compared to traditional livestock breeds, and the principles and methodology might be tailored to these new animal protein species. Therefore we will initiate a research line on genomic enhanced breeding of insects, exploiting the newest genome sequence pipelines.

Last modified: 13-12-2023

Breeding the next animal protein source: how genomics will take us to the next step in insect breeding

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: ***N**(ew data) or* ***E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
iHerIll2.2.curated.20191125	Reference sequence of BSF - GCA_905115235.1	E	D	Reference Genome	FASTA & online platform	NA	NA
WGS of BSF	Open access WGS of different BSF	E	D	seq data	FASTQ	< 1 Tb	NA
BSF samples	Different biological samples of BSF from different strains	N	P	Frozen samples	NA	NA	Tubes (Eppendorf) N = ± 500
DNA	Extracted DNA of BSF	N	P	Frozen DNA	NA	NA	micro-tubes N = ± 250
New WGS BSF sequences	WGS of BSF (1X and 10X coverage) of different BSF strains	N	D	seq data	FASTQ	< 1 Tb	NA
Imputation pipeline	lc-WGS pipeline for imputation of BSF genotypes	N	D	M	TXT	< 1 Gb	NA
Imputed WGS BSF sequences	Imputed WGS of BSF (10X level) of different BSF strains	N	D	seq data	FASTQ	< 5 Tb	NA
Phenotypes of BSF	Phenotypic measurements of BSF production and reproduction	N	D	T	CSV	< 1 Gb	NA

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

iHerIll2.2.curated.20191125: accession ID GCA_905115235.1 described in 10.1093/g3journal/jkab085
WGS of BSF available via open access: accession IDs PRJNA917807, PRJNA575900, PRJNA547968, PRJNA432297, PRJNA394479

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

Current project is a pilot for the development (pending FWO postdoc proposal).
Imputation pipeline: can be commercialized into spin-off or as tech transfer towards commercial BSF breeding.
WGS of different BSF strains: can contain valuable information for future BSF breeding programs.

KU Leuven (LRD) will, in combination with other partners (Thomas More, Belgium) and the Research Institute of Organic Agriculture, FiBL, in Switzerland, consult on IP for the commercial valorization of the data at the start of the FWO project. During the current 1 year pilot project, there will be no direct commercialization.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

Sequence data will be annotated using metadata standards (https://doi.org/10.25504/FAIRsharing.r2ts5t).
For the imputation protocol, a README.txt will be generated to allow swift transfer of the pipeline for further use, also allowing the transfer to other species.
All biological and DNA samples will be entered in the research group's biobank database, listing origin, date of collection, date of DNA extracting, linking files that contain additional information on the samples, as routinely recorded for all biological samples.
Phenotypes will be described following a standard, annotated (README) csv will containing additional parameters that were used/collected during BSF breeding.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

FASTQ metadata: https://doi.org/10.25504/FAIRsharing.r2ts5t

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Large Volume Storage
- Other (specify below)

Sequence data (including imputed WGS) will be stored on LVS.
Phenotypes and the lc-WGS imputation protocol will be sorted on Shared Network drives (J-Drive).
Biological samples will be stored in the active biobank (freezers -20°C) maintained by the Center of Animal Breeding and Genetics, Department of Biosystems.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

For the current pilot project, no additional storage needs are foreseen.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Access to shared folders on is regulated by ICTS and only members of our research group have access. Access right to the large volume storage will be curated by ICTS to researchers active in this project (incl. partners, see below).
Biobank samples (BSF and DNA) are stored in a lab with restricted badge access. Additional security access measures will be implemented once the biobank is moved to a separate biobank room with badge access (once KA30 building renovations are finished).

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

For LVS, a cost of 104.5€ per TB/year is expected.
For the pilot, no additional storage will be required (low amount of data generated).
For the applied FWO project, an additional benchfee was requested (± 200-300 € for data storage).

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- Shared network drive (J-drive)
- Large Volume Storage (longterm for large volumes)

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

For LVS, a cost of 104.5€ per TB/year is expected.
For the pilot, no additional storage will be required (low amount of data generated) and costs will be covered by the routine credits of the lab.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as open data
- Other (specify below)

If not hampered by the potential setup of commercialization (spin-off) and taking IP between different partners into account, all data will be made available (open data) according to the FAIR principles.
However, test samples from the ongoing pilot project won't be publicly accessible, as they are being prepared for the upcoming FWO postdoc proposal.

If access is restricted, please specify who will be able to access the data and under what conditions.

NA

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- Yes, intellectual property rights

BSF strains owned by Thomas More could be subjected by their IP. KU Leuven LRD will be consulted during the PDM year in preparation of the FWO postdoc grant.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)
- Other data repository (specify below)

Sequence data will be uploaded to www.ncbi.nlm.nih.gov and on KU Leuven's Research Data Repository (RDR) when appropriate and allowed.

When will the data be made available?

- Upon publication of research results

If not jeopardized by any IP restrictions by the partners as discussed on the start of the FWO project (if granted).

Which data usage licenses are you going to provide?

If none, please explain why.

- Data Transfer Agreement (restricted data)
- CC-BY 4.0 (data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

What are the expected costs for data sharing? How will these costs be covered?

NA

Responsibilities

Who will manage data documentation and metadata during the research project?

Dr. Roel Meyermans under supervision of Prof. Dr. Nadine Buys

Who will manage data storage and backup during the research project?

Dr. Roel Meyermans under supervision of Prof. Dr. Nadine Buys

Who will manage data preservation and sharing?

Prof. Dr. Nadine Buys

Who will update and implement this DMP?

Dr. Roel Meyermans under supervision of Prof. Dr. Nadine Buys