CONTRAT N° B2/202/P2/InsectMOoD
Insect Museum Open -omic Database
BELSPO programme manager: Véronique Feys

project follow-up Committee
3th meeting, 18th of July 2023

Participants:
Katerina Guschanski – University of Edinburgh, InsectMOoD follow-up committee

Massimiliano Virgilio - project coordinator, Royal Museum for Central Africa (RMCA)
Gontran Sonet - project collaborator, RBINS
Lore Esselens - project research assistant, RMCA

Peter Hollingsworth (Royal Botanic Garden Edinburgh, InsectMOoD follow-up committee), Francesca Scolari (University of Pavia, InsectMOoD follow-up committee) and Carl Vangestel (project partner, Royal Belgian Institute of Natural Sciences (RBINS)) could not attend the online meeting, but kindly provided *a posteriori* feedback after receiving an overview of the topics discussed during the meeting and the slides presented by Lore Esselens during the meeting.

Minute of the meeting

LE presented an overview of the project results (see the presentation attached to this report), showcasing the main achievements of InsectMOoD. Below a general overview of the topics discussed:
Data collection (completed)
- 1406 DNA extracts recovered from collection vouchers and quality checked
- high throughput sequencing (HTS) data represented by Whole Genome Sequencing (WGS) data (10x coverage) obtained from 1305 collection samples

Genomic voucher databasing (ongoing and to be completed before the project end, 31/8/2023)
- assembly of 643 genomic vouchers completed
- uploading of metadata on open-access databases ongoing

Data analysis (ongoing and to be completed before the project end, 31/8/2023)
- relationships between sample age, DNA quality/quantity, HTS performance do not follow obvious trends, with the possible exceptions of
    - absorbance ratios of 260 and 280 nm, which suggest decreasing DNA quality in older samples possibly due to protein contamination,
    - absorbance ratios of 260 and 230 nm, which also seem to decrease in older sample possibly due to contamination of RNA, EtOH, phenol and/or other organic compounds.

Project scientific output
- two abstracts presented at the ISFFEI 2022 - Sydney

- manuscript in prep (submission planned in 2023)

Input from ongoing projects at RMCA-RBINS:

- ISeBAF: use of quality check metrics obtained for Syrphidae and Tephritidae processed within this project.
- SYNTHESYS+: use of quality check metrics obtained for historical pollen samples processed within this project.
- FF-IPM: use of quality check metrics obtained for Tephritidae processed within this project (*Ceratitis capitata*, samples ageing 1-20yrs).
- DISPEST: use of quality check metrics obtained for Tephritidae processed within this project (*Ceratitis rosa, C. quilicii*, samples ageing 1-5yrs).
- DARWIN: use of available databasing protocols and infrastructure available at RMCA-RBINS.

MV mentioned the list of project deliverables which should be met in due time.

KG indicated that RNA is degrading faster which should be visible in the absorbance ratio of 260 over 280 nm. We tried to measure the RNA contamination levels with a Qubit fluorometric quantification which was not possible because of too low concentrations.

GS provided an overview on the efforts made by JEMU-RBINS for standardizing the databasing of sample metadata and discussed possible options to incorporate the genomic vouchers of InsectMOoD in public databases such as DARWIN or GBIF. KG agreed that the standardisation of collection management procedures is a general problem of many research institutions.

MV mentioned a list of potential research partners who showed interest in the genomic vouchers made available in InsectMOoD. Among them Dr. Giuseppe Saccone (University of Naples, Biology Department, IT) interested in WGS data from *Ceratitis capitata* (Tephritidae); Dr. Domenico Rizzo (Phytosanitary Service, Region Tuscany, IT), Dr. Negin Ebrahimi (Flanders Research Institute for Agriculture, Fisheries and Food, BE) and Dr. Jan Mertens, Netherlands Institute for Vectors, Invasive plants and Plant health (NIVIP), National Plant Protection Organization (NPPO) interested in WGS data from the "FARQ" complex (Tephritidae).

The discussion was raised about the publication of the data collected during the InsectMOoD project. The raw genomic data will be gradually released, pending publication of results from the different research projects contributing to InsectMOoD (and including ISeBAF, FF-IPM, etc…). Conversely, all sample metadata will be publicly accessible via open access databases which will be provided on dedicated domains available on the website of the AfricaMuseum.

For checking the contamination of the HTS data, KG suggested to align the WGS data to an outgroup species equally distant from tephritids and flower flies (e.g. *Drosophila*). This should eliminate biases due to distances between the reference genome and the members of the ingroup. LE will initiate a pilot test to explore this option. FS mentioned that additional genomic and transcriptomic data are being produced for tephritid species in the frame of the EU REACT project, and collaborations may be established to perform additional mapping.

Regarding the fragment sizes of the DNA, KG asked if we looked at the median fragment size. We have this data, so this could be examined to see if there are significant trends to be observed over the different variables. Additionally, we only looked at the short fragment sizes, but CV pointed out that he had the best predictive power using large fragment sizes (700-10380bp).

KG asked about the use of the library preparation: is sample fragmentation, enrichment or tagmentation applied? To know if there is a sequencing bias towards longer fragments. We confirmed that the genomic DNA was randomly sheared into short fragments. According to the quality report of the sequencing company, KG believes that DNA tagmentation was performed.

FS and CV indicated that the different specimen preservation methods (within and between projects) could have an effect on the DNA quality/quantity. FS also added that the less optimal absorbance ratios of 260 over 280 nm may be due to the low DNA concentration in older samples, and microbial contamination and environmental DNA may have played a role in (older) samples causing a less optimal absorbance ratios of 260 over 230 nm

Tervuren July 19th, 2023

Lore Esselens, Gontran Sonet, Carl Vangestel, Massimiliano Virgilio
Joint Experimental Molecular unit
of the Royal Museum for Central Africa and the Royal Belgian Institute of Natural Sciences