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

project follow-up Committee
2nd meeting, 7th of December 2022

Participants:
Francesca Scolari – University of Pavia, InsectMOoD follow-up committee
Peter Hollingsworth – Royal Botanic Garden Edinburgh, InsectMOoD follow-up committee

Massimiliano Virgilio - project coordinator, Royal Museum for Central Africa (RMCA)
Carl Vangestel - project partner, Royal Belgian Institute of Natural Sciences (RBINS)
Gontran Sonet - project collaborator, RBINS
Lore Esselens - project research assistant, RMCA

Report

Katerina Guschanski (Uppsala University, InsectMOoD follow-up committee) 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. The input of KG has been incorporated into this report.

Minute of the online meeting
A brief overview of the project timeline and implementation was provided by LE (see presentation attached as an Annex to this report) with focus on:

- progress in DNA extraction, quality check, high throughput sequencing (HTS, completed or largely finalised)
- sample metadata databasing (largely finalised)
- quality check HTS (ongoing and benefiting of suitable reference genomes recently assembled at RMCA)
- data modelling (initiated)
- scientific output (one InsectMoOD abstract presented at the ISFFEI 2022 - Sydney)
- manuscript writing (to be initiated in Feb 2023)
- synergies with ongoing projects at RMCA-RBINS:
    - ISeBAF: use of quality check metrics obtained for bees, 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 discussed the option of asking for project prolongation (+3 to +6 months) and reviewed the list of project deliverables, which were already met or which should be met in due time.

FS, PH and CV pointed out the higher variability in DNA concentrations quantified in samples from 2020 compared to other years (slide 5 presentation). Discussion followed about the need of further investigating this pattern and to identify the possible drivers of the observed variability.
FS and MV discussed about the suitability of publicly available reference genomes compared to genomes recently developed within the InsectMOoD consortium and soon to be published. FS and MV agreed that the use of the higher number of closely related genomes now available for genera *Ceratitis* (two Omni-C, annotated ref. genomes + one PacBio Hi-fi scaffold assembly), *Zeugodacus* (one Omni-C, annotated ref. genome) and *Bactrocera* (one PacBio Hi-fi scaffold assembly) will allow higher quality metrics for the performance of HTS in Tephritidae.

PH and GS suggested to verify levels of sample contamination by using respectively DNA barcoding and/or abnormally higher frequencies of rare alleles. The project partners agreed that this could be considered as part of the decision map to be developed by this project and that discrepancies in DNA barcoding ID or allele frequencies should definitely be used as criteria to flag problematic genomic vouchers.

KG recommended also considering reads <100 bp when investigating changes in DNA fragmentation with sample age as this would allow observing trends associated to highly fragmented DNA. LE confirmed that reads <100bp were not pre-filtered when analysing fragmentation patterns and that these fragments were routinely included in the analyses.

KG observed that, as verified in her lab with mitochondrial genomes, mapping the same species to different reference genomes could result in heterogeneous proportion of mapped reads and or of genome coverage. KG suggested to verify the extent of potential biases on clades of known topology using multiple reference genomes. The project consortium agreed to proceed accordingly as soon as all the assembled reference genomes will become available.

Tervuren 6 January 2022

Massimiliano Virgilio, Carl Vangested, Lore Esselens, Gontran Sonet
Royal Museum for Central Africa
Royal Belgian Institute od Natural Sciences