BRAIN-be 2.0
BELGIAN RESEARCH ACTION THROUGH INTERDISCIPLINARY NETWORKS - Phase 2

Annual Network Report

To be filled in for the whole network in French, Dutch or English and sent to: BRAIN-be@belspo.be

Contrat nr : B2 / 202 / P2 / 2023

InsectMOoD

Reporting period : 1/1/21 - 31/03/22

The *Annual Network Report* (maximum 15 to 20 pages) is drawn up annually by the coordinator for the entire network and sent to the address BRAIN-be@belspo.be on the dates set in article 7.6 of annex I to the contract. It presents the state of progress and achievements of the research as well as the forecasts for the following year. This information refers explicitly to the tasks and the project schedule defined in articles 2 and 3 of annex I. It also informs of any modification of the data included in the initial reports and gives the list of publications and missions carried out during the past year.

This template can be completed in French, Dutch or English.

NETWORK

COORDINATOR (PARTNER 1)

1. Massimiliano Virgilio, Royal Museum for Central Africa (RMCA)

OTHER PARTNERS

2. Carl Vangestel, Royal Belgian Institute of Natural Sciences (RBINS)

AUTHORS OF THIS REPORT

1. Massimiliano Virgilio (RMCA)
2. Carl Vangestel (RBINS)
3. Lore Esselens (RMCA)

PROJECT WEBSITE, SOCIAL NETWORKS …

https://www.africamuseum.be/en/staff/896/project_detail_view?prjid=7220

https://www.researchgate.net/project/Insect-Museum-Open-omic-Database-InsectMOoD

https://twitter.com/EsselensLore/status/1504795410061303813

TABLE OF CONTENTS

NETWORK 2

COORDINATOR (PARTNER 1) 2

OTHER PARTNERS 2

AUTHORS OF THIS REPORT 2

PROJECT WEBSITE, SOCIAL NETWORKS … 2

1. EXECUTIVE SUMMARY OF THIS REPORT 4

2. ACHIEVED WORK 4

APPENDIX 1: LIST OF PROCESSED SPECIMENS PER SPECIES 10

SELECTED REFERENCES O DIRECT RELEVANCE FOR THIS PROJECT 11

3. INTERMEDIARY RESULTS 13

4. PRELIMINARY CONCLUSIONS AND RECOMMANDATIONS 17

5. FUTURE PROSPECTS AND PLANNING 17

6. FOLLOW-UP COMMITTEE 17

7. VALORISATION ACTIVITIES 18

7.1 PUBLICATIONS 18

7.2 PARTICIPATION/ORGANISATION OF SEMINARS (NATIONAL/INTERNATIONAL) 18

7.3 SUPPORT TO DECISION MAKING (IF APPLICABLE) 19

7.4 OTHER 19

8. ENCOUNTERED PROBLEMS AND SOLUTIONS 19

9. MODIFICATIONS COMPARED TO THE PREVIOUS REPORT 19

9.1 PERSONNEL 19

9.2 COMPOSITION OF THE FOLLOW-UP COMMITTEE 20

10. REMARKS AND SUGGESTIONS 20

1. EXECUTIVE SUMMARY OF THIS REPORT

- The performances of different DNA extraction kits on suboptimal Museum samples have been comparatively tested.
- A cost-effective approach to DNA extraction has been defined.
- A decision map on wet-lab procedures for the whole genome sequencing of vouchers from our Museum insect collection has been drafted.
- The DNA of more than 300 insect vouchers from the RMCA collections of Tephritidae and Syrphidae (Diptera) has been extracted, quality checked and subjected to Whole Genome Sequencing (WGS, genome coverage 10x).
- WGS data and metadata have been archived on the RMCA servers.
- WGS data processing is currently ongoing.
- The setup of an open-access database for voucher archiving (including voucher metadata, genomic data and digital imaging) has been initiated.

2. ACHIEVED WORK

Detailed description of the achieved work and tasks since the previous report

InsectMOoD aims at promoting the genotyping of Museum vouchers as a routine procedure to preserve and valorise the genomic information associated to the collection specimens. Below a list of activities performed in the framework of this project.
At the start of the project a literature survey was carried on and provided an overview of methodologies and approaches to the genomic analyses of Museum specimens (see list of selected references reported below). Following discussion with the follow-up committee, we've set up a test to verify the performances of different DNA extraction kits.

Comparative performances of DNA extraction methods
In this pilot study, different DNA extraction kits were tested on Museum vouchers from the RMCA collections of Tephritidae and Syrphidae (Diptera). We selected 3 to 6 specimens from seven collection series dating from 2008 to 2016. These included three Tephritidae (*Zeugodacus cucurbitae* (Coquillett), *Bactrocera dorsalis* (Hendel), *Dacus bivittatus* (Bigot)) and two Syrphidae (*Eumerus* sp. and *Ischiodon aegyptuis* (Wiedemann)) species, and specimens which were either stored in 100% ethanol at -20°C (Zc, Bd, Db, Eu) or pinned and preserved at room temperature (Ia). Digestions in lysis buffers were implemented on whole bodies for all specimens. For comparative purposes, we also processed forelegs only, rather than whole insects (Fig. 1). The lysates obtained from each specimen were divided in four aliquots and the DNA purified using spin columns from the DNA extraction kits listed in Tab. 1 and following the manufacturer's instructions. This experimental design was based on 30 whole specimens and 18 legs (2 negative controls were also included) which provided aliquoted DNA that was processed through 200 spin columns from four different extraction kits.

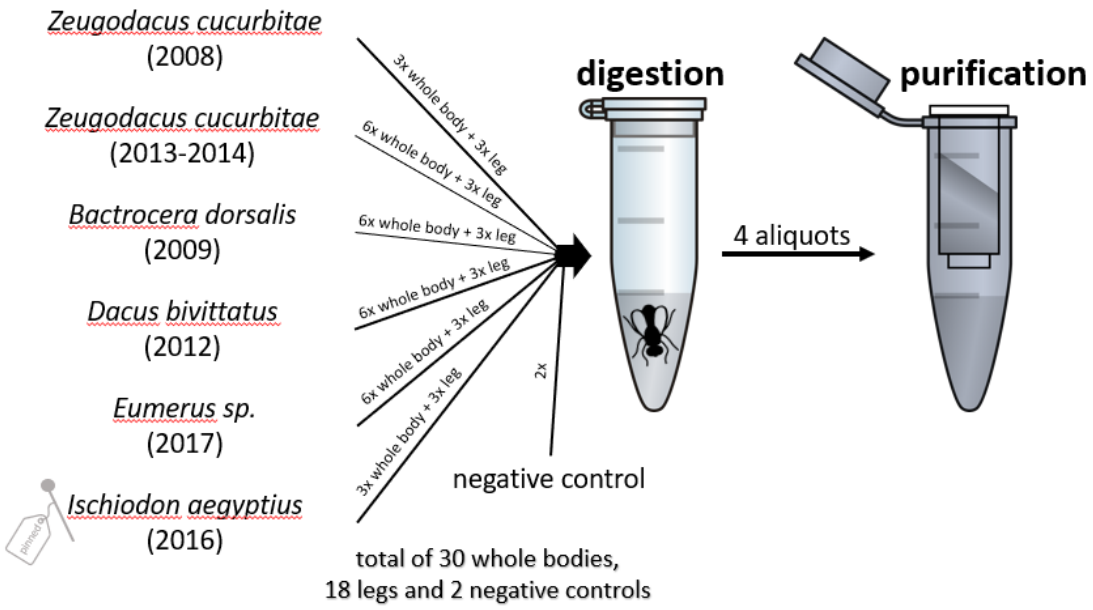

Figure 1: Experimental set-up of samples used for the pilot study of DNA extractions.

Table 1: Qiagen kits with specifications and costs.

Qiagen kit (50 samples)	Column	Range DNA size	Expected DNA yield	Price (BTWI)
DNeasy Blood and Tissue Kit	Dneasy spin column	100bp-50 kb	6-30 µg	€ 260,15
QIAamp Micro Kit	QIAamp MinElute column	<30 kb	<3 µg	€ 349,69
QIAamp Mini Kit	QIAamp Mini spin column	<50 kb	4-30 µg	€ 272,25
DNeasy Blood and Tissue Kit	MinElute column (MinElute PCR Purification Kit)	70bp-4 kb	<5 µg	€ 424,71

The concentration of each DNA extract was measured using a Qubit 3 fluorometer (HS DNA Kit, Thermo Fisher Scientific) and the total amount of DNA was inferred from the final elution volume, which in all cases was 100 µl. An overview of the DNA yields is provided in Fig. 2 and Fig. 3.
The figures show heterogeneous DNA yields across specimens with values ranging from 57,8 to 153 ng for whole bodies and for legs 1,3 to 22 ng (as expected, due to the lower amount of tissue in legs compared to whole bodies).
The comparison of yields obtained from different DNA extraction methods shows that none of the kits tested provides markedly higher yields compared to the widely used Qiagen DNeasy Blood and Tissue Kit (Fig. 4). Therefore, considering const-benefits we decided using the kit with the lowest price among those tested. Therefore, the DNeasy columns will be the columns of choice in the routine processing of vouchers from the target insect collections.

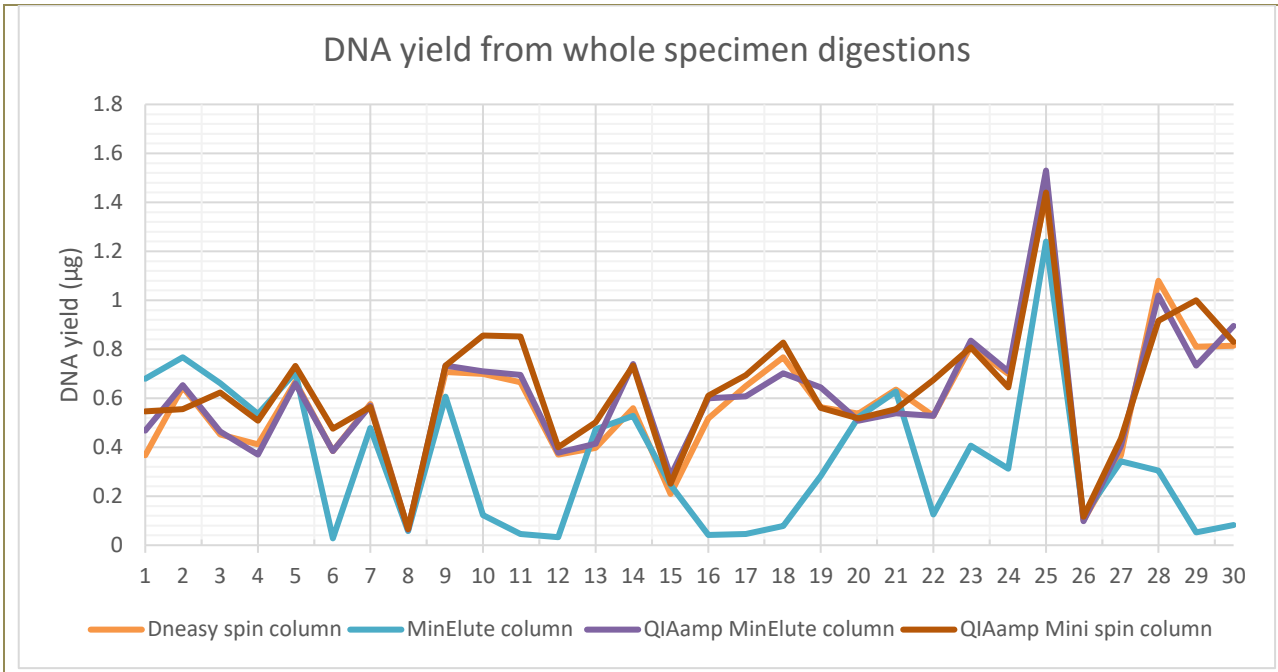

Figure 2: DNA yields from replicated elutions (whole-body digestions). 1-3: Zeugodacus cucurbitae (2016); 4-9: Zeugodacus cucurbitae (2013-2014); 10-15: Bactrocera dorsalis (2009); 16-21: Dacus bivittatus (2012); 22-27: Eumerus sp. (2017); 28-30: Ischiodon aegyptius (2016).

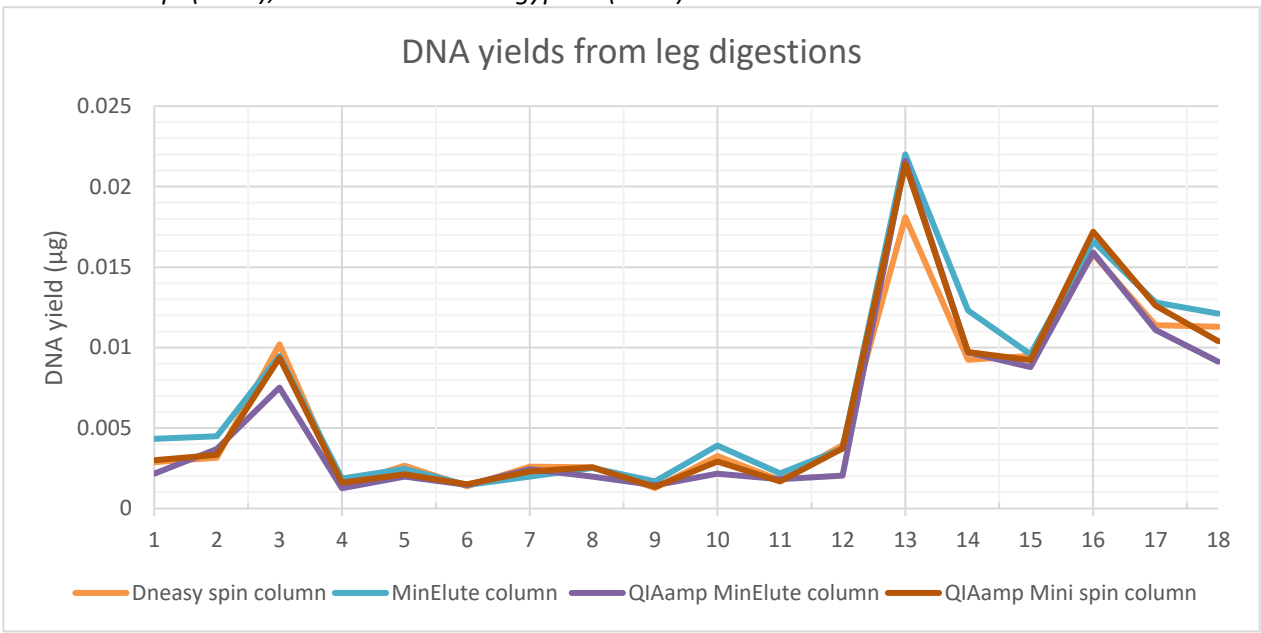

Figure 3: DNA yields from replicated elutions (leg digestions). 1-3: Zeugodacus cucurbitae (2016); 4-6: Zeugodacus cucurbitae (2013-2014); 7-9: Bactrocera dorsalis (2009); 10-12: Dacus bivittatus (2012); 13-15: Eumerus sp. (2017); 16-18: Ischiodon aegyptius (2016).

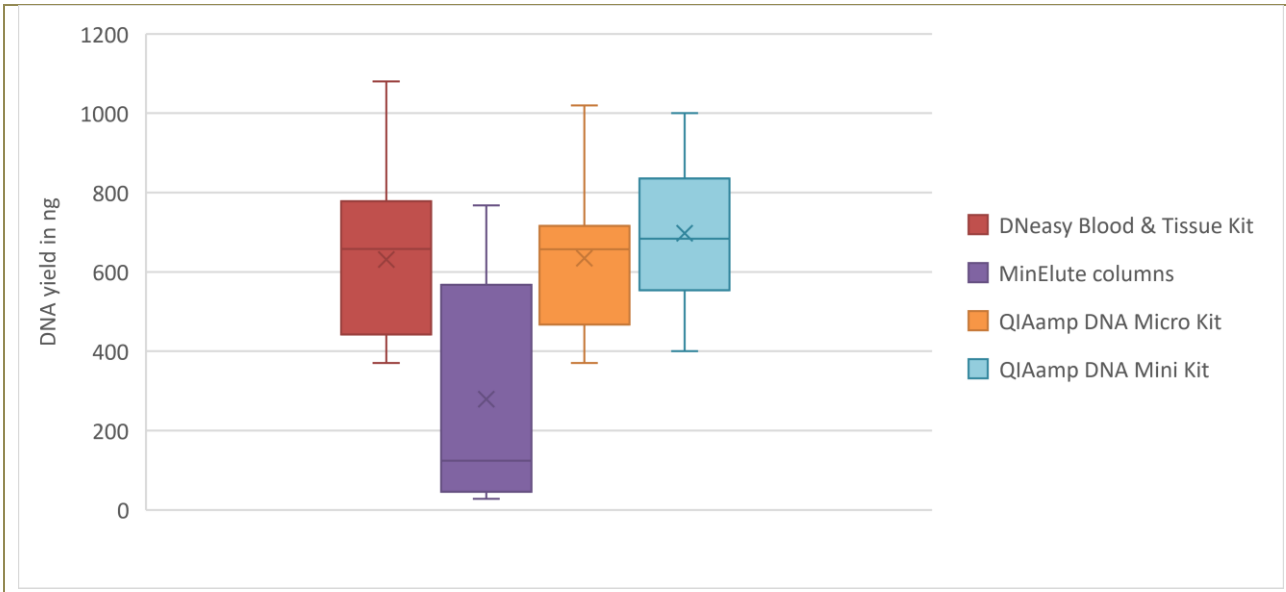

Figure 4: Boxplots of DNA yields from replicated elutions (whole body digestions) of the different extraction kits.

Comparisons between genomic library preparation and whole genome sequencing (WGS) services provided by external companies

After preliminarily considering the cost-benefits of different approaches to the whole genome sequencing (WGS) of the Tephritidae and Syrphidae archived in our Museum collections we considered that outsourcing the preparation of genomic libraries was the most suitable strategy for the routine-processing of vouchers. In fact, external companies, whose costs are steadily decreasing over years, now provide significantly lower prices compared to the in-the-house processing of samples. Therefore we decided that, rather than testing in our labs a range of commercial kits and experimental protocols, we would rather quantitatively compare the performances of services provided by different companies. In this respect, we targeted the performances of two companies providing standard genomic library preparation (Berry Genomics and Novogene) or low-input genomic library preparation (Novogene) (Tab. 2). Sample submitted to these companies are preliminarily subjected to an extended quality check with the objective of describing relationships between levels of DNA concentration, contamination and fragmentation and quality / quantity of reads obtained from the different genomic library preparation protocols implemented (Fig. 5).

All samples are sequenced at 10x coverage on an Illumina NovaSeq platform (150 PE reads, 6Gb raw data output / sample). These comparative analyses are currently ongoing and the collection samples processed so far are listed in Appendix 1.

Table 2: Sequencing companies with library preparation requirements and costs.

Company	Library preparation (Illumina NovaSeq150PE)	Amount of DNA required	Concentration of DNA required	Price/ sample	Extra costs
Berry Genomics	Standard	>2 µg	40 ng/µl	€ 72,6	€ 260 clearance fee/batch
Novogene	Standard	>0,4 µg	10 ng/µl	€ 128	-
	Low input	0,03-0,2 µg	NA	€ 148	-

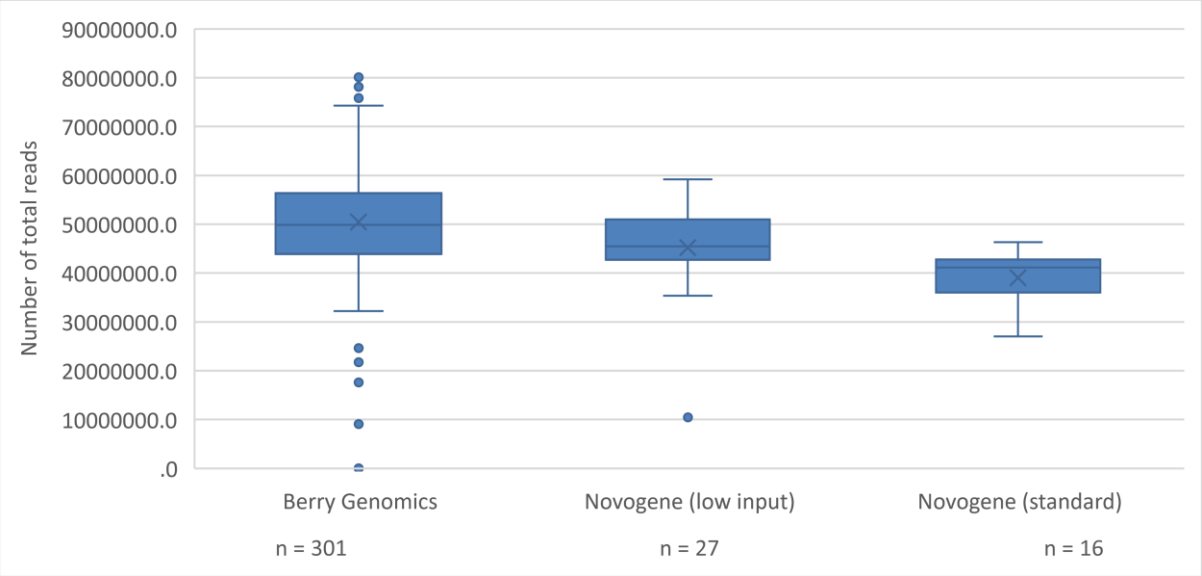

Figure 5: The total number of reads obtained from different sequencing companies (Berry Genomics and Novogene) and different library preparations protocols (low input DNA and standard library preparation).

Relationships between voucher DNA quality and WGS performance

In order to describe relationships between WGS performance and voucher (a) age and preservation and (b) DNA quality and quantity we targeted a total of 344 specimens archived in the collections of RMCA collected from 1999 to 2020 (Fig. 6) and preserved either in ethanol at -80°C (n = 288), pinned at room temperature (n = 14) or dried DNA stored at room temperature (n = 42). All DNA extractions were done using the Qiagen Dneasy blood and tissue kit. We quantified on the one hand the quantity of DNA extracted as measured by a Qubit 3 fluorometer (HS DNA Kit, Thermo Fisher Scientific), on the other hand the quality of DNA as quantified by (a) the DNA fragment size distributions as measured using the DNF-930 dsDNA Reagent Kit (75 bp – 20000 bp) on the fragment analyzer of the Genomics Core (Leuven) and (b) the ratios of absorbance at 260 and 280 nm (A260/280) and at 260 and 230 nm (A260/230). These latter were measured in triplo using the Implen NanoPhotometer N60 Touch. The ratios of absorbance are used to assess the purity of DNA. Values of A260/230 = 1.8 and 2.0 < A260/280 <2.2 are indicative of low DNA contamination.

In our tests, the A260/230 values are clearly related to the amount of DNA and suggest higher contamination for lower DNA concentrations. This trend is not observed for A260/280 (Fig. 7).

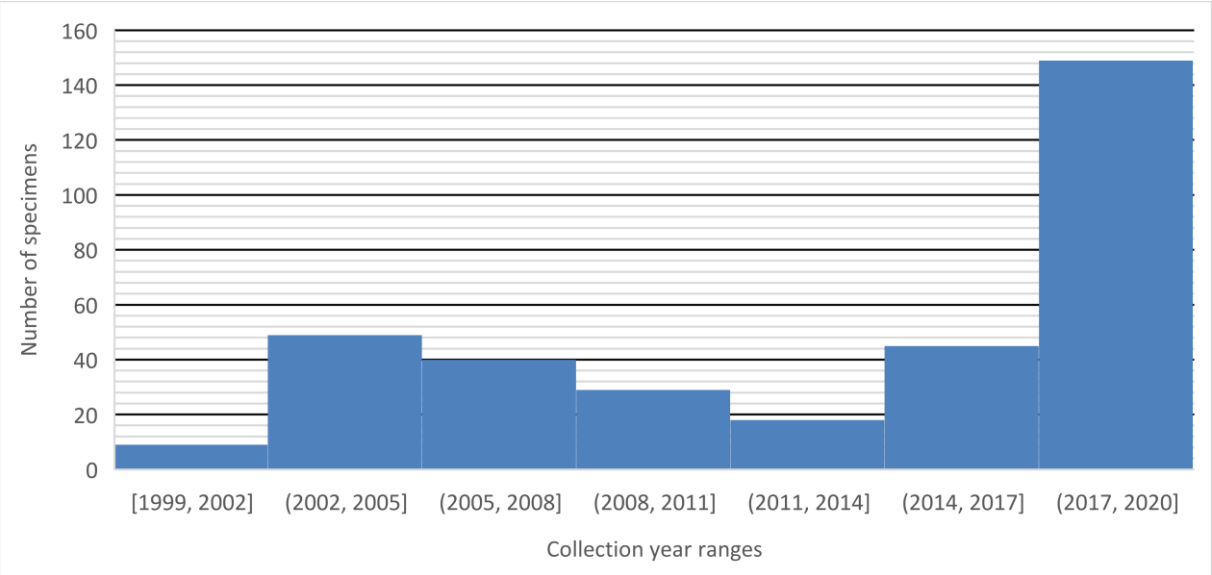

Figure 6: Frequency of processed specimens per age range of 3 years.

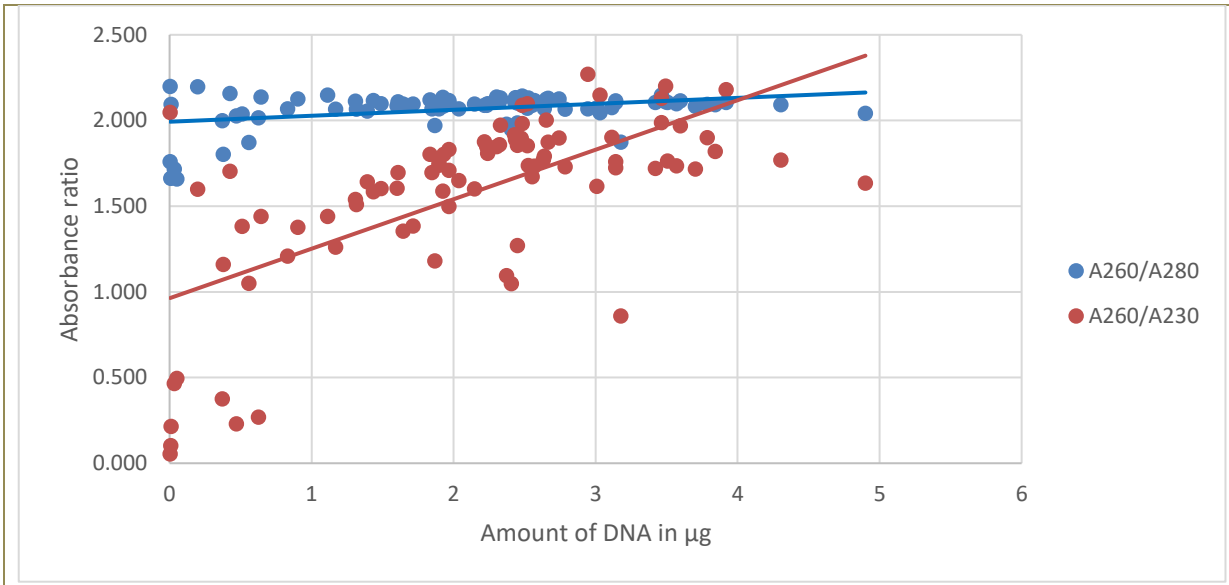

Figure 7: Scatterplot of the absorbance ratios at 260 nm over 280 nm (blue) and 260 nm over 230 nm (red) in relation to the amount of DNA (n = 86).

All DNA extracts with DNA concentrations above 10 ng/µl were subjected to standard genomic library preparation with samples submitted either to Berry Genomics (n = 301) or to Novogene (n = 43). A reduced number of samples with DNA concentration below 10 ng/µl were submitted to Novogene and subjected to low-input DNA library prep (n= 27).

After whole genome sequencing of samples, the raw reads were filtered, trimmed and mapped against a suitable reference genome: reads from *Bactrocera sp*. and *Ceratitis sp.* were mapped against a genome assembly of *Bactrocera dorsalis* (GCA_020283865.1) or *Bactrocera latifrons* (GCA_001853355.1) and reads from *Dacus sp., Eristalinus sp.* and *Melanostoma sp.* were mapped against a genome assembly of *Zeugodacus cucurbitae* (GCA_000806345.1), *Eristalis pertinax* (GCA_907269125.1) and *Melanostoma mellinum* (GCA_914767635.1) respectively. Genomic data and quality parameters of the DNA of 344 specimens originating from more than 60 species have been collected (see Table 3 and Appendix 1) and the preliminary analyses seems to suggest positive correlation between initial DNA concentration and the percentage of reads successfully mapped (Fig. 8). An in-detail characterization of relationships between voucher age and preservation, DNA quality and quantity and WGS performance is currently ongoing.

Table 3: Number of collection vouchers processed from three Tephritidae and two Syrphidae genera.

Collection	**Genus**	**Number of specimens**
Tephritidae	*Bactrocera*	15
Tephritidae	*Dacus*	134
Tephritidae	*Ceratitis*	96
Syrphidae	*Eristalinus*	75
Syrphidae	*Melanostoma*	24

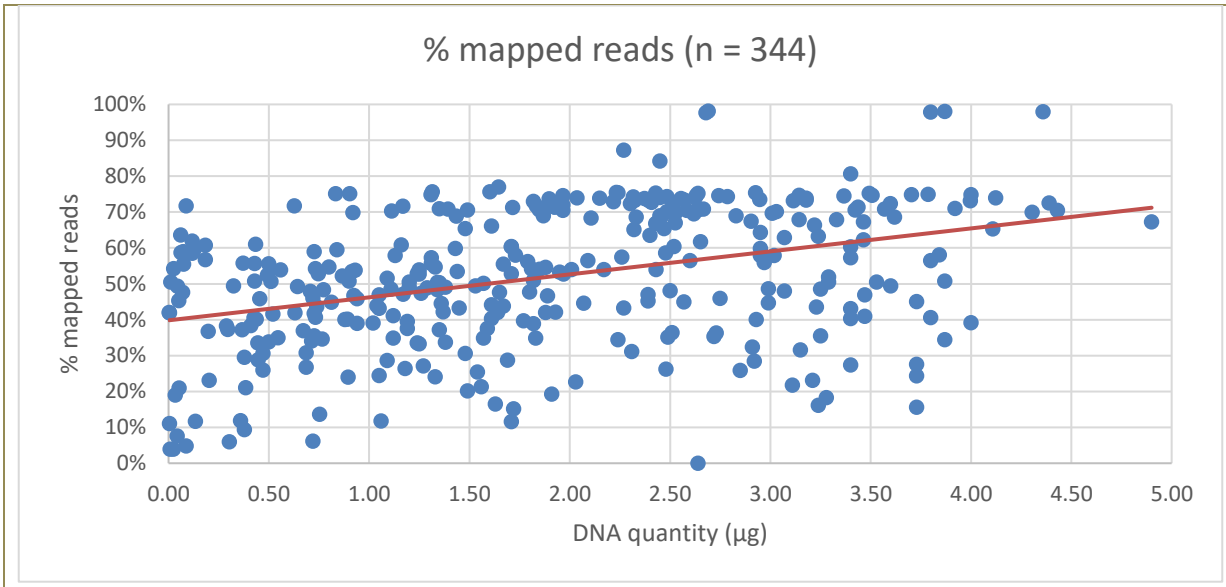

Figure 8: Relationships between quantity of DNA extracted and percentage of reads mapped.

APPENDIX 1: LIST OF PROCESSED SPECIMENS PER SPECIES

Table 4: List of specimens per species.

Collection	Species	Number of vouchers processed
Tephritidae	*Bactrocera dorsalis*	6
Tephritidae	*Bactrocera latifrons*	5
Tephritidae	*Bactrocera mesomelas*	4
Tephritidae	*Ceratitis quilicii*	76
Tephritidae	*Ceratitis rosa*	20
Tephritidae	*Dacus africanus*	1
Tephritidae	*Dacus armatus*	7
Tephritidae	*Dacus bivittatus*	20
Tephritidae	*Dacus brevistriga*	1
Tephritidae	*Dacus ceropegiae*	1
Tephritidae	*Dacus chiwira*	3
Tephritidae	*Dacus ciliatus*	4
Tephritidae	*Dacus demmerezi*	6
Tephritidae	*Dacus diastatus*	3
Tephritidae	*Dacus durbanensis*	2
Tephritidae	*Dacus eclipsis*	2
Tephritidae	*Dacus famona*	4
Tephritidae	*Dacus frontalis*	4
Tephritidae	*Dacus fuscovittatus*	1
Tephritidae	*Dacus fuscovitttatus*	1
Tephritidae	*Dacus humeralis*	16
Tephritidae	*Dacus hyalobasis*	2
Tephritidae	*Dacus kariba*	2
Tephritidae	*Dacus langi*	3
Tephritidae	*Dacus longistylus*	7
Tephritidae	*Dacus masaicus*	4

Tephritidae	*Dacus namibiensis*	2
Tephritidae	*Dacus pallidilatus*	2
Tephritidae	*Dacus phloginus*	1
Tephritidae	*Dacus pulchralis*	1
Tephritidae	*Dacus pullescens*	2
Tephritidae	*Dacus punctatifrons*	4
Tephritidae	*Dacus quilicii*	1
Tephritidae	*Dacus siliqualactis*	4
Tephritidae	*Dacus sphaeristicus*	1
Tephritidae	*Dacus telfaireae*	4
Tephritidae	*Dacus tenebricus*	1
Tephritidae	*Dacus theophrastus*	3
Tephritidae	*Dacus umehi*	1
Tephritidae	*Dacus venetatus*	1
Tephritidae	*Dacus vertebratus*	10
Tephritidae	*Dacus woodi*	1
Tephritidae	*Dacus xanthopterus*	1
Syrphidae	*Eristalinus aeneus*	1
Syrphidae	*Eristalinus albus*	4
Syrphidae	*Eristalinus arvorum*	1
Syrphidae	*Eristalinus cupreus*	1
Syrphidae	*Eristalinus dubiosa*	1
Syrphidae	*Eristalinus eclarus*	1
Syrphidae	*Eristalinus euzonus*	1
Syrphidae	*Eristalinus flaveolus*	1
Syrphidae	*Eristalinus fuscicornis*	4
Syrphidae	*Eristalinus goergeni*	2
Syrphidae	*Eristalinus gymnops*	1
Syrphidae	*Eristalinus lemmyi*	1
Syrphidae	*Eristalinus lineifacies*	1
Syrphidae	*Eristalinus macrops*	25
Syrphidae	*Eristalinus megametapodus*	1
Syrphidae	*Eristalinus modestus*	1
Syrphidae	*Eristalinus myiatropinus*	2
Syrphidae	*Eristalinus sexvittatus*	1
Syrphidae	*Eristalinus smaragdinus*	2
Syrphidae	*Eristalinus sp.*	9
Syrphidae	*Eristalinus tabanoides*	4
Syrphidae	*Eristalinus taeniops*	5
Syrphidae	*Eristalinus vicarians*	5
Syrphidae	*Melanostoma sp.*	24

SELECTED REFERENCES O DIRECT RELEVANCE FOR THIS PROJECT

1. Besnard, G. *et al.* (2016) 'Valuing museum specimens: High-throughput DNA sequencing on historical collections of New Guinea crowned pigeons (Goura)', *Biological Journal of the Linnean Society*, 117(1), pp. 71–82. doi: 10.1111/bij.12494.

2. Bi, K. *et al.* (2013) 'Unlocking the vault: Next-generation museum population genomics', *Molecular Ecology*, 22(24), pp. 6018–6032. doi: 10.1111/mec.12516.
3. Blaimer, B. B. *et al.* (2016) 'Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens', *PLoS ONE*, 11(8), p. 161531. doi: 10.1371/journal.pone.0161531.
4. Buenaventura, E. (2021) 'Museomics and phylogenomics with protein-encoding ultraconserved elements illuminate the evolution of life history and phallic morphology of flesh flies (Diptera: Sarcophagidae)', *BMC Ecology and Evolution*, 21(1). doi: 10.1186/s12862-021-01797-7.
5. Burrell, A. S., Disotell, T. R. and Bergey, C. M. (2015) 'The use of museum specimens with high-throughput DNA sequencers', *Journal of Human Evolution*, 79, pp. 35–44. doi: 10.1016/j.jhevol.2014.10.015.
6. Call, E. *et al.* (2021) 'Museomics: Phylogenomics of the Moth Family Epicopeiidae (Lepidoptera) Using Target Enrichment', *Insect Systematics and Diversity*, 5(2). doi: 10.1093/isd/ixaa021.
7. Cridland, J. M. *et al.* (2018) 'Genome Sequencing of Museum Specimens Reveals Rapid Changes in the Genetic Composition of Honey Bees in California', *Genome Biology and Evolution*, 10(2), pp. 458–472. doi: 10.1093/gbe/evy007.
8. Faircloth, B. C. *et al.* (2015) 'Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among hymenoptera', *Molecular Ecology Resources*, 15(3), pp. 489–501. doi: 10.1111/1755-0998.12328.
9. Gauthier, J. *et al.* (2020) 'Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland', in *Molecular Ecology Resources*. Blackwell Publishing Ltd, pp. 1191–1205. doi: 10.1111/1755-0998.13167.
10. Gillett, C. P. D. T. *et al.* (2014) 'Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea)', *Molecular Biology and Evolution*, 31(8), pp. 2223–2237. doi: 10.1093/molbev/msu154.
11. Guschanski, K. *et al.* (2013) 'Next-generation museomics disentangles one of the largest primate radiations', *Systematic Biology*, 62(4), pp. 539–554. doi: 10.1093/sysbio/syt018.
12. Hung, C. M. *et al.* (2013) 'The De Novo Assembly of Mitochondrial Genomes of the Extinct Passenger Pigeon (Ectopistes migratorius) with Next Generation Sequencing', *PLoS ONE*, 8(2), p. 56301. doi: 10.1371/journal.pone.0056301.
13. Mayer, C. *et al.* (2021) 'Adding leaves to the Lepidoptera tree: capturing hundreds of nuclear genes from old museum specimens', *Systematic Entomology*, 46(3), pp. 649–671. doi: 10.1111/syen.12481.
14. Mikheyev, A. S. *et al.* (2017) 'Museum Genomics Confirms that the Lord Howe Island Stick Insect Survived Extinction', *Current Biology*, 27(20), pp. 3157-3161.e4. doi: 10.1016/j.cub.2017.08.058.
15. Papanicolaou, A. *et al.* (2016) 'The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species', *Genome Biology*, 17(1), pp. 1–31. doi: 10.1186/s13059-016-1049-2.
16. Rohland, N., Siedel, H. and Hofreiter, M. (2010) 'A rapid column-based ancient DNA extraction method for increased sample throughput', *Molecular Ecology Resources*, 10(4), pp. 677–683. doi: 10.1111/j.1755-0998.2009.02824.x.
17. Rowe, K. C. *et al.* (2011) 'Museum genomics: Low-cost and high-accuracy genetic data from historical specimens', *Molecular Ecology Resources*, 11(6), pp. 1082–1092. doi: 10.1111/j.1755-0998.2011.03052.x.
18. Suchan, T. *et al.* (2016) 'Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens', *PLoS ONE*, 11(3), p. e0151651. doi: 10.1371/journal.pone.0151651.
19. Timmermans, M. J. T. N. *et al.* (2016) 'Rapid assembly of taxonomically validated mitochondrial genomes from historical insect collections', *Biological Journal of the Linnean Society*, 117(1), pp. 83–95. doi: 10.1111/bij.12552.
20. Tin, M. M. Y., Economo, E. P. and Mikheyev, A. S. (2014) 'Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics', *PLoS ONE*, 9(5), p. 96793. doi: 10.1371/journal.pone.0096793.

21. Waku, D. *et al.* (2016) 'Evaluating the phylogenetic status of the extinct Japanese otter on the basis of mitochondrial genome analysis', *PLoS ONE*, 11(3), p. 149341. doi: 10.1371/journal.pone.0149341.

3. INTERMEDIARY RESULTS

Detailed description of the deliverables completed since the previous report

Deliverables completed:

D 1.1.2: report on suitable JEMU datasets for validation of results (M4)

An additional dataset from the JEMU project KEARAD will contribute to the project analyses. The data obtained from this dataset are currently being processed and will provide additional information on relationships between DNA quantity and WGS performances in cichlid fishes dating from 1984 to 2019 from the collections of RMCA. The graph below suggests correlation between DNA concentrations and voucher age (Fig. 9).

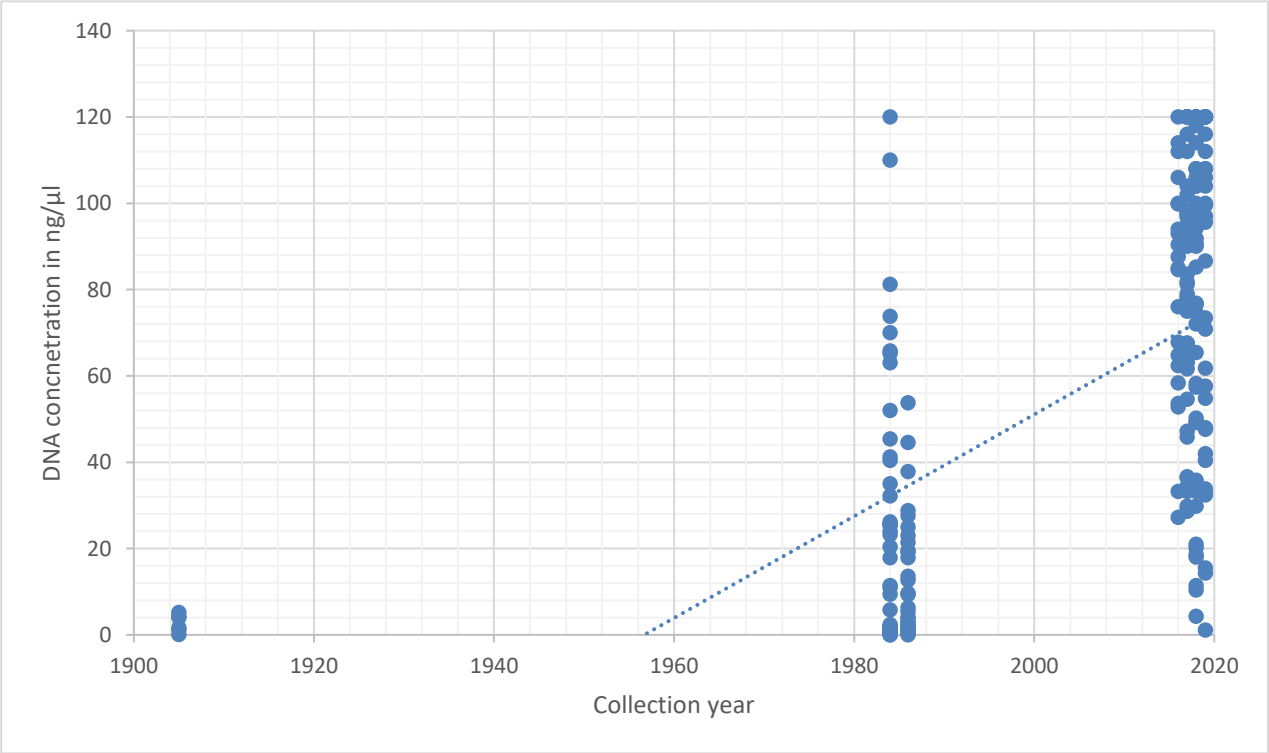

Fig. 9: DNA concentrations in 238 cichlid fishes dating from 1984 to 2019

D 2.1.1: project experimental setup (M6)

The general project experimental setup to characterise relationships between sample age and preservation in Syrphidae and Tephritidae, DNA quality ad quantity and WGS performance has been defined. Vouchers from the Diptera collections of RMCA are currently being selected and processed and metrics form DNA quality check recorded.

D 2.1.2: report on test for target taxa (syrphids, tephritids) (M12)

Lab tests completed. Figure 8 and 9 illustrate DNA concentrations and WGS reads obtained in Museum specimens from the two target taxa: Syrphidae (flower flies) and Tephritidae (fruit flies).

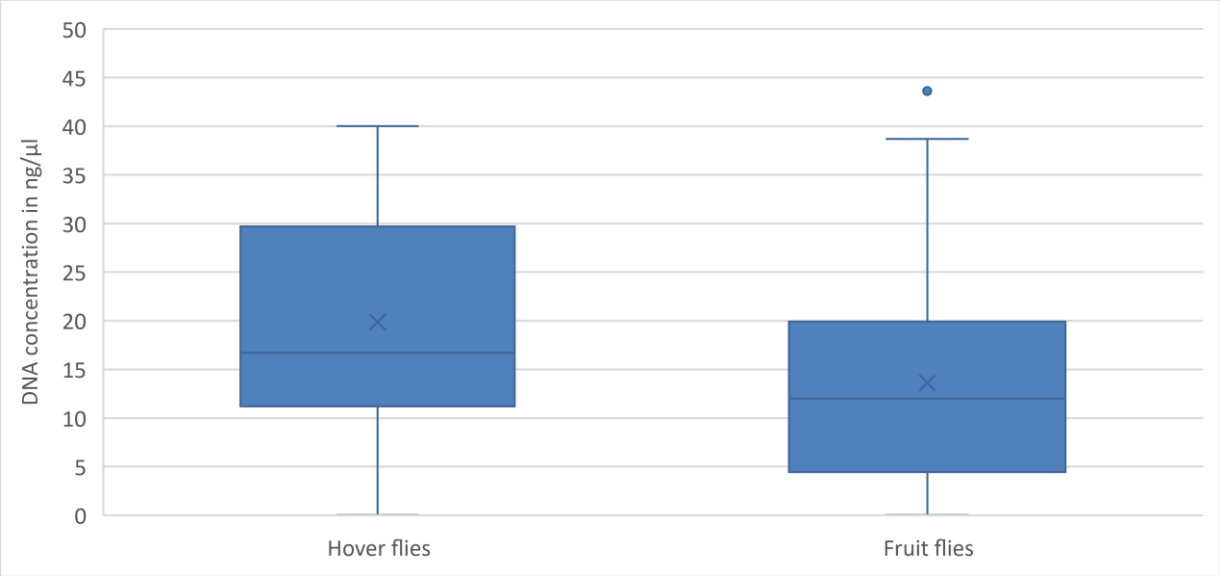

Figure 9: Boxplots of DNA concentration (ng/µl) of hover flies (n = 99) and fruit flies (n = 149).

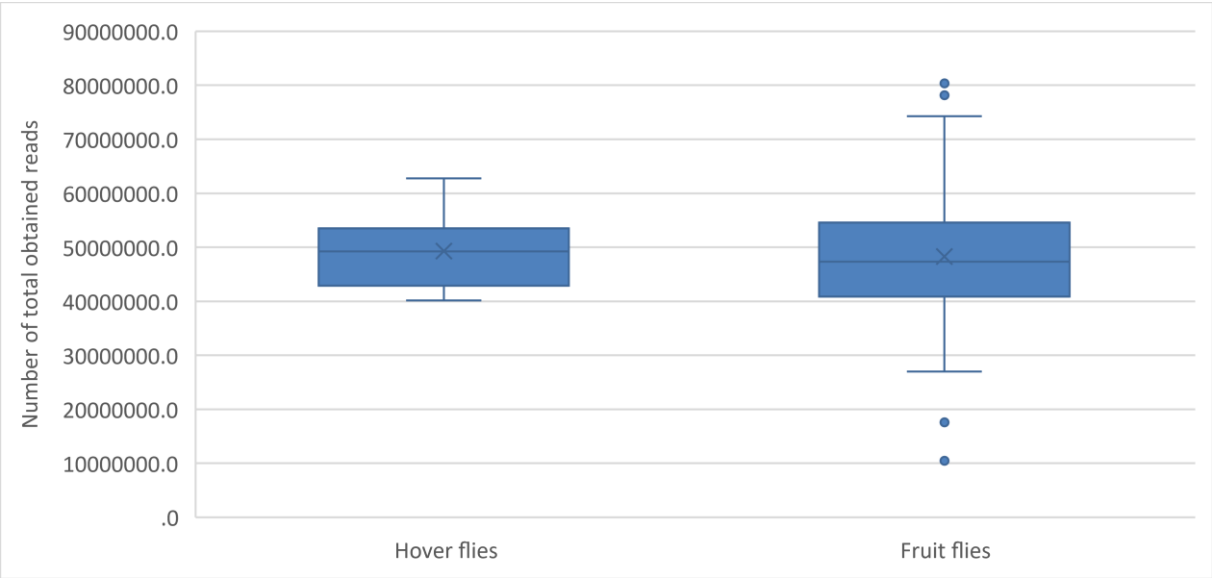

Figure 10: Boxplots of total obtained reads of hover flies (n = 99) and fruit flies (n = 149).

D 2.2.1: cost-benefit analysis (M12)

A cost benefit analysis was implemented on DNA extraction procedures (4 commercial kits compared), the DNeasy kit was deemed to be the most cost-effective for the routine processing of the targeted samples from the collections of Syrphidae and Tephritidae.

D 2.2.2: decision map for the WGS of suboptimal samples (M12)

A decision map for the targeted vouchers form the RMCA collections of Syrphidae and Tephritidae is provided below (Fig. 11)

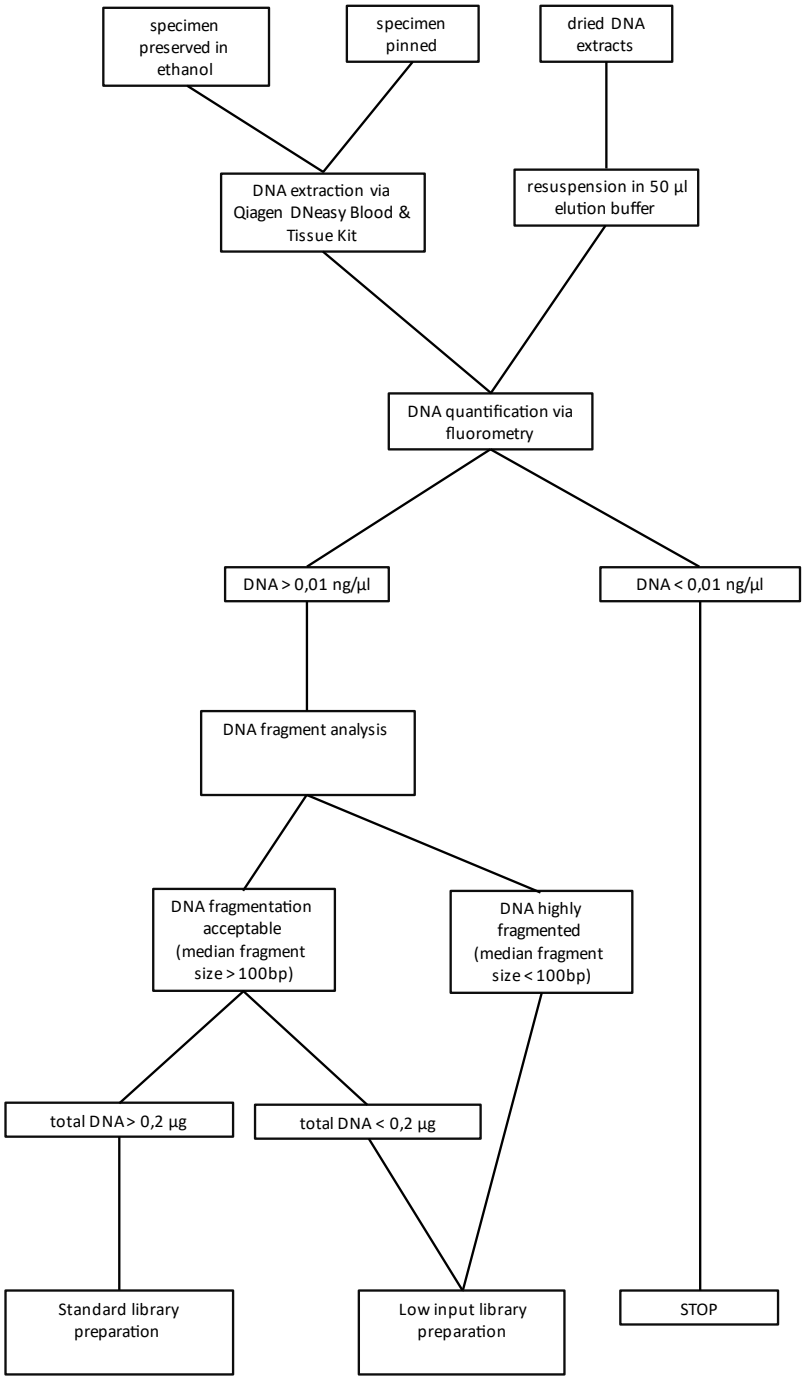

Figure 11: A decision map for the routine genotyping of Diptera from Museum collections.

D 4.1.1: meetings follow-up committee (M1, M12, M24)

The 1st meeting of the project follow-up Committee has been held online on the 1st of July 2021 (see section below).

D 4.1.2: project reports (M3, M12, M24)

The initial project report has been submitted on 15/3/2021

The first year report on 20/6/2021

Deliverables initiated

D 2.2.1: cost-benefit analysis (M12)

The cost-benefit analysis on genomic library preparation protocols as provided by external companies is currently in progress.

D 3.3.2: production of DNA collection vouchers complementing the RMCA collections (M24)

DNA vouchers have been created and linked to the voucher metadata for all the collection vouchers processed so far (>300). The production of additional vouchers for the DNA collection is currently ongoing.

D 5.1.1: backup and long term storage of WGS data and metadata (M24)

The backup and storage of WGS data is currently going on with WGS data stored in the NAS systems of RMCA for all the collection vouchers processed so far (>300)

D 5.1.2: backup and long term storage of digital images and metadata (M24)

The backup and storage of voucher metadata is currently going on with metadata stored in the NAS systems of RMCA for all the collection vouchers processed so far (>300).

D 6.1.1: participation to international congresses (M10, M19)

Due travel restrictions related to the 2021 COVID-19 emergency, we have not taken part in international congresses.

D 6.1.2: post / interviews on RMCA social media accounts (M6, M12, M18, M24).

A post on Twitter on March 18th 2022 was published to inform general public about the progresses in InsectMOoD: https://twitter.com/EsselensLore

Deliverables to be completed in year two:

D 3.1.1: report on voucher WGS and assemblage (M18, M24).

D 3.2.1: report on tests for non-target taxa (bees, fishes) (M18, M24).

D 3.3.1: production of digital vouchers complementing the RMCA digital collections (M24)

D 3.3.2: production of DNA collection vouchers complementing the RMCA collections (M24)

D 3.3.3: establishing a collection of RMCA genomic vouchers M24).

D 3.3.4: online, open access database of genomic vouchers (M24).

D 4.1.1: meetings follow-up committee (M1, M12, M24)

D 4.1.2: project reports (M3, M12, M24)

D 5.1.1: backup and long term storage of WGS data and metadata (M24)

D 5.1.2: backup and long term storage of digital images and metadata (M24)

D 6.1.1: participation to international congresses (M10, M19)

D 6.1.2: post / interviews on RMCA social media accounts (M6, M12, M18, M24).

4. PRELIMINARY CONCLUSIONS AND RECOMMANDATIONS

The use of standard methods for the WGS of Museum specimens in suboptimal conditions from the RMCA collections of flower flies (Syrphidae) and fruit flies (Tephritidae) aging 1-20 years is providing satisfactory results by yielding suitable HTS read output. Further analyses are currently scrutinizing the quality of the output provided, by proceeding with SNP filtering and SNP calling on different datasets. The preliminary results obtained suggest that, even if DNA quality and quantity is often below the minimal standards required by companies for genomic library preparation and sequencing, it might still be cost-effective to proceed processing these suboptimal samples, as in most cases, these will provide a reasonable amount of data to be directly used or phylogenomic reconstructions or population genomics. Further analyses will better define the limit of this pragmatical effective approach.

5. FUTURE PROSPECTS AND PLANNING

Overview of the foreseen activities and planning for the next reporting period, taking into account the current state of the work and the intermediary results

- Meetings with the ICT service of RMCA have been organized to define a suitable pipeline for the online archiving of genomic data and metadata (and including publicly available data and data to be disclosed after publication). Two webpages (one for Tephritidae, one for Syrphidae) to be hosted on the RMCA servers are currently being organized. These webpages will include links to the genomic data and metadata.
- The bioinformatic pipelines for the analysis of the genomic data collected during year 1 are currently being optimized with the main objective of more rapidly processing the whole genome sequencing data obtained.
- An in-depth analysis of relationships between quality and quantity of high throughput sequencing reads obtained from Museum samples of different ages is currently ongoing. These analyses will follow the preliminary analyses described above and aim at describing relationships between the number of single nucleotide polymorphisms (SNPs) recovered form each Museum specimens and the quality and quantity of its DNA.
- The routine processing of the target Museum specimens (Syrphidae and Tephritidae) is ongoing.

6. FOLLOW-UP COMMITTEE

Date(s) of the meeting(s) and overview of the concrete contributions of the follow-up committee

The 1st meeting of the project follow-up Committee was held on 1/7/2021 (online). The participants were
Francesca Scolari – University of Pavia, InsectMOoD follow-up committee
Katerina Guschanski – Uppsala University, 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
Jessica Ody – JEMU research assistant, RMCA
During the meeting, a brief overview of the project timeline and implementation was provided by LE with

focus on:

- project structure and timeline
- sample age distributions in the tephritid and syrphid collections of RMCA
- methodological approach to wet lab pipelines
- methodological approach to the analysis of High Throughput Sequencing (HTS) performances
- expected outcome and valorisation of results

KG, PH and FS initiated discussion and provided input on

- destructive DNA extraction procedures and effects of sample age on proportions of endogenous vs contaminant DNA in Museum vouchers
- best practices for voucher cleaning and sterilization -> include PBS washing step before DNA extraction protocol to decrease contaminant DNA (if samples were collected > 10 years ago)
- fragmentation of degraded vouchers: retain smaller endogenous DNA fragments to distinguish from larger contaminant DNA fragments
- negative controls for DNA extraction and library preparation
- replicates vs pooling samples + use distinct species per batch
- include molecular identification and double indexing library preparation
- Use of Roche columns (High Pure Viral Nucleic Acid Large Volume Kit) as a possible alternative to the DNeasy Blood and Tissue Kit (Qiagen) columns to retain DNA fragments < 100bp
- Use of Qiagen kit for DNA extractions of sperm cells -> QIAamp DNA Mini or Micro kit as a suitable strategy to deal with low DNA concentrations
- MinElute columns (MinElute PCR Purification Kit from Qiagen) retain DNA fragments > 70bp -> in principle they could replace the DNeasy columns in the DNeasy Blood and Tissue Kit (Qiagen) protocol -> check DNA fragmentation as short fragments should be mostly endogenous and contamination is mostly about long fragments
- Library preparation: TRUSEQ Nano DNA Low Throughput Library Prep kit (Illumina) recommended for small yields of (degraded) DNA
- Control procedures in HTS and need for negative + positive controls (including library prep control + DNA extraction control)
- Suggestion for pilot on library preparation focusing on species with published/available reference genome

As a result of this discussion and following the recommendations from the follow-up committee, a pilot experiment was organized and will be carried out to test the possible effects of the use of different extraction protocols/columns on the final amount of DNA yielded from specimens of various populations.

7. VALORISATION ACTIVITIES

7.1 PUBLICATIONS

N/A

7.2 PARTICIPATION/ORGANISATION OF SEMINARS (NATIONAL/INTERNATIONAL)

Oral presentations, posters… and/or organisation of workshops, conferences, etc.

N/A

7.3 SUPPORT TO DECISION MAKING (IF APPLICABLE)

N/A

7.4 OTHER

With reference to article 5.2 of the contract of InsectMOod, the project starts on 15/12/2020 and ends on 15/03/2023. Due to delays in hiring personnel, the project research assistant (12+12 PM) was hired only starting form the 5th of May 2021. Therefore, after communication with BELSPO, the project will be extended for 2 additional months without additional financing. A request for project extension will be submitted in due time no later than 30 days before the end of the contract.

8. ENCOUNTERED PROBLEMS AND SOLUTIONS

Encountered problems/obstacles, implemented and/or considered solutions, if any.

A possible bottleneck when routinely processing relatively high number of specimens from the collections is represented by the bioinformatic pipelines for the preliminary quality check of the obtained WGS reads. In the framework of this project we are currently working to obtain time-effective bioinformatic pipelines to rapidly quality check data from multiple specimens.
Due travel restrictions related to the 2021 COVID-19 emergency, we have not taken part in international congresses.
Similarly, due to repeated lab lock downs in 2021 we experienced some delay in lab activities, so that the cost-benefit analysis on genomic library preparation protocols could not be finalised (analysis is currently in progress, to be embedded in the 2022 report).

9. MODIFICATIONS COMPARED TO THE PREVIOUS REPORT

9.1 PERSONNEL

In case modifications have occurred since the previous report regarding personnel in charge and at the disposal of the project, please list these in the following table conform the instructions given in the *Initial Report*. Send a copy of the employment contracts of the personnel in charge to brain-be@belspo.be.

N/A

A research assistant, Lore Esselens, was selected and hired starting from the 5th of May 2021. See section above for request of project extension.

9.2 COMPOSITION OF THE FOLLOW-UP COMMITTEE

Dr. Francesca Scolari,
University of Pavia (Italy),
Department of Biology and Biotechnology
Via Ferrata 9
I-27100 Pavia, Italy
tel: +39 (0382) 986023
e-mail: francesca.scolari@unipv.it

Prog. Katerina Guschanski,
University of Uppsala (Sweden)
Department of Ecology and Genetics, Animal Ecology
Norbyvägen 18 D
752 36 UPPSALA
tel: +46 18471 2673
e-mail: katerina.guschanski@ebc.uu.se

Prof. Peter Hollingsworth,
Director of Science & Deputy Keeper
Royal Botanic Garden, Edinburgh, UK
Inverleith row 20a, EH3 5LR, Edinburgh

10. REMARKS AND SUGGESTIONS

Concerning for example: the coordination, the use or valorisation of the results, personnel change

NA