

InsectMOoD

Insect Museum Open -omic Database

Rationale

Museum specimens are of particular importance for research on taxonomy, systematics and biogeography and provide tools to tackle a wide range of scientific questions in disciplines such as ecology, evolution and conservation. The rapid advances of massively parallel sequencing technologies have led to a substantial reduction in sequencing costs, so that the routine whole genome sequencing (WGS) of museum specimens represents a cutting-edge intervention to valorise the collection's genetic resources. In this context, the Joint Experimental Molecular Unit (JEMU) of RMCA-RBINS is defining the best practices for the genotyping of **suboptimal vouchers** of insects (i.e., aging 20-25 years, Fig. 1). This approach relies on digital archiving of genomic data as a complement to the physical and digital archiving of museum vouchers.

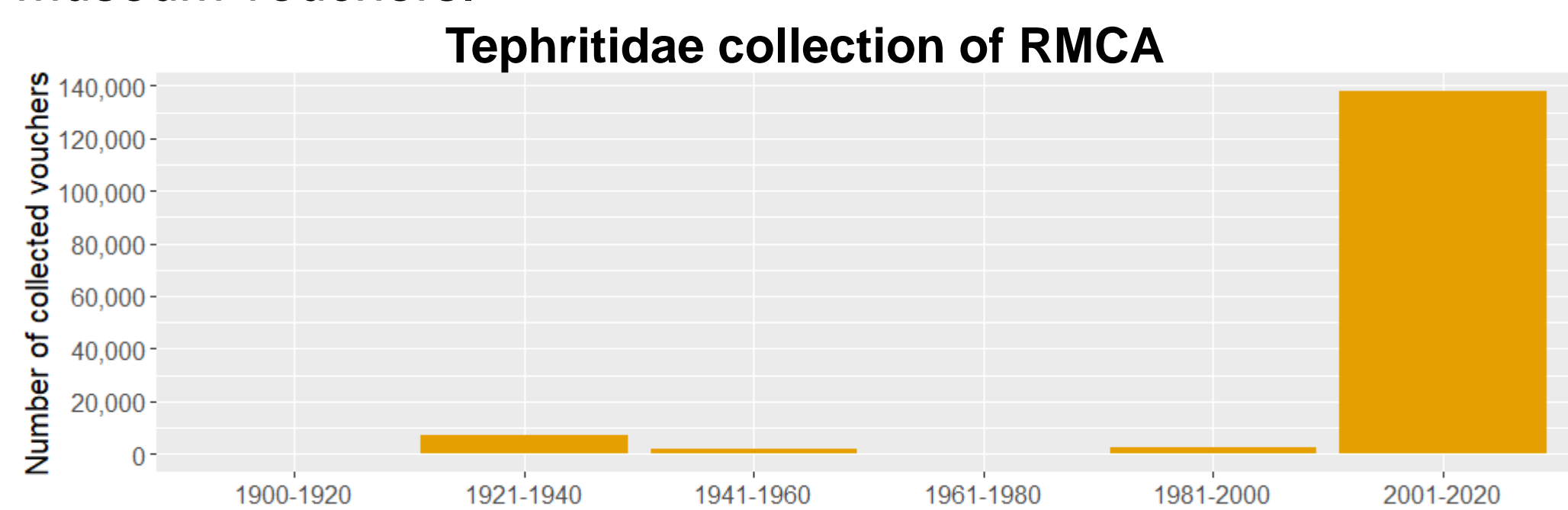


Fig. 1: Sample age distribution of Tephritid collection at RMCA

Methodological approach

The performances of DNA extraction and library preparation were quantitatively compared on approximately **300 tephritid vouchers** of different ages sampled from the collections of RMCA. After DNA extraction and quality check, samples were subjected to WGS at 10x coverage on an Illumina NovaSeq platform (150 PE reads, 6Gb raw data output / sample) (Fig. 2). Relationships between quantity/quality of extracted insect DNA (as estimated by fluorometric methods and by levels of DNA fragmentation and contamination, see Fig. 3, 4) and quantity/quality of WGS output (including total amount of filtered reads, proportion of high-quality reads, etc., see Fig. 5) were investigated using a correlative approach.

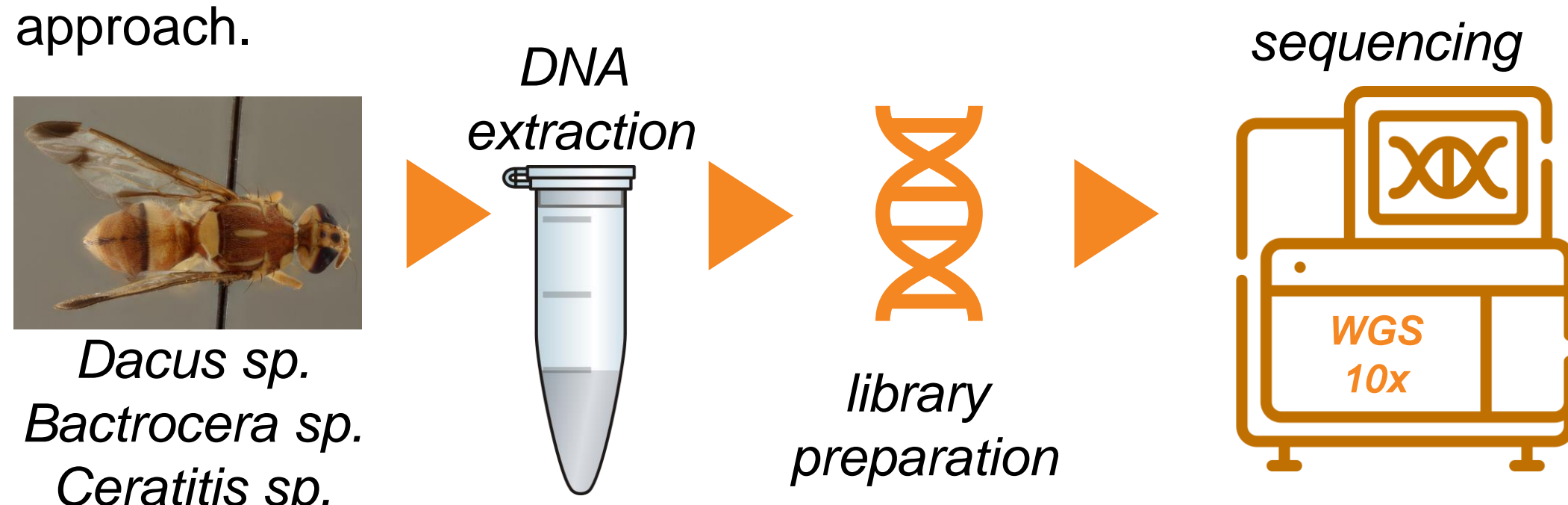


Fig. 2: Graphic visualization of the methodological approach

Preliminary results

DNA quantity over time

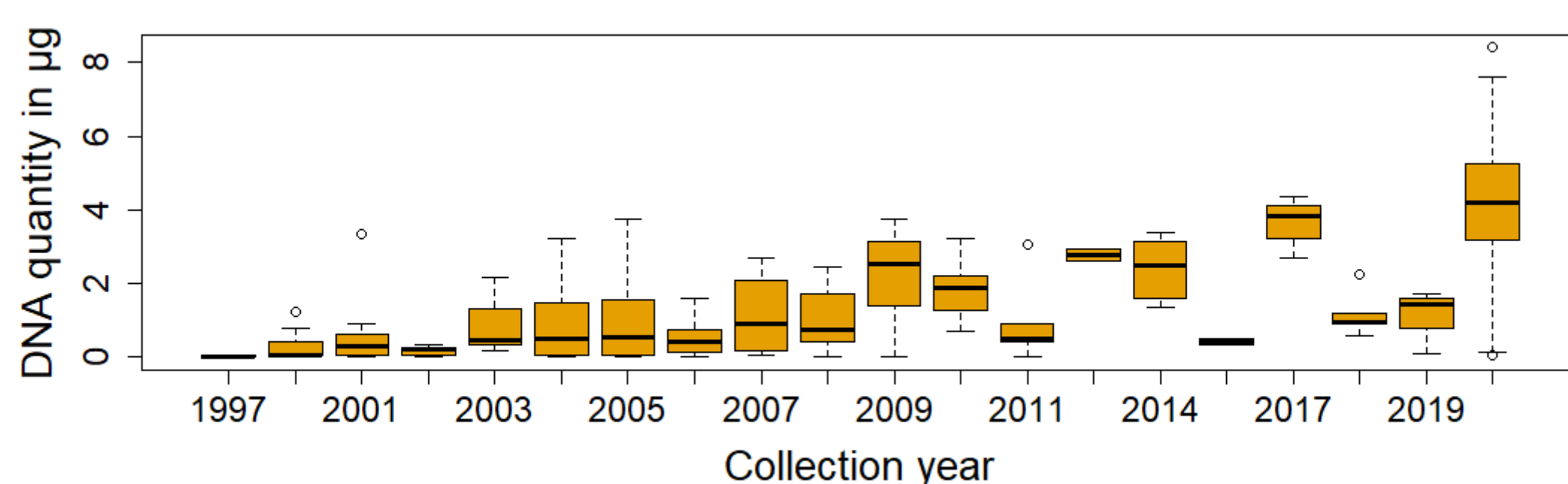


Fig. 3: DNA quantities (calculated from concentration measured with Qubit 4.0) per collection year

DNA quality over time

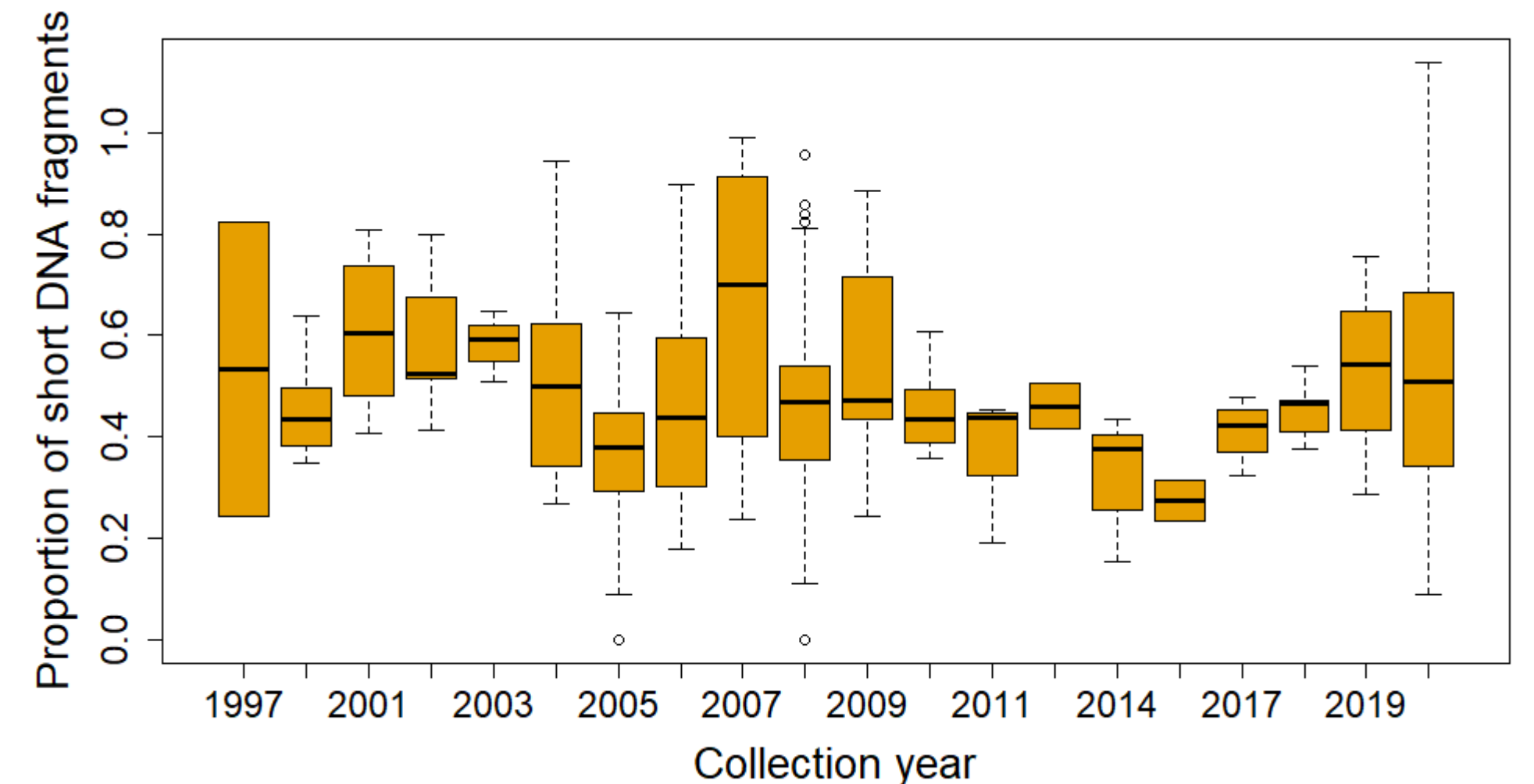


Fig. 4: The proportion of DNA fragments between 35 and 350 bp

High throughput sequencing (HTS) performances

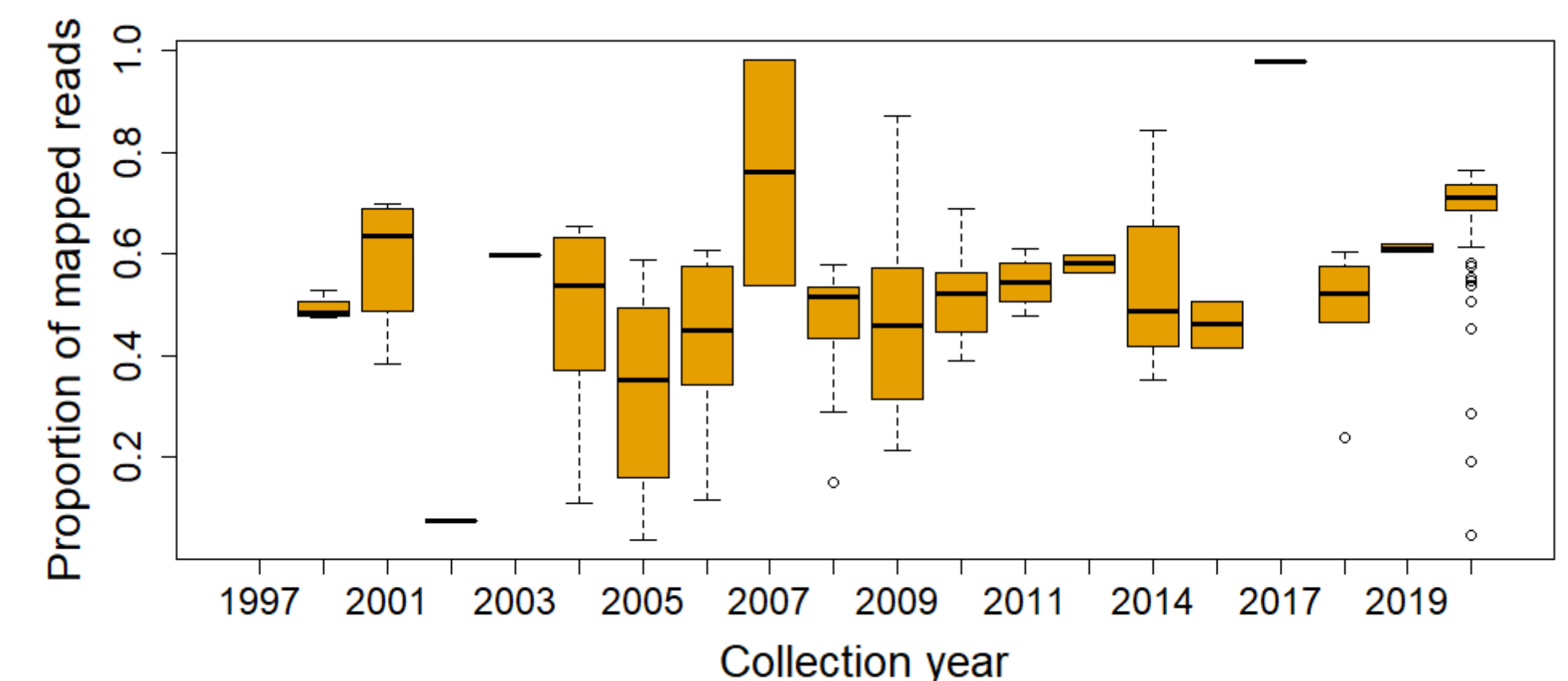


Fig. 5: The proportion of mapped reads against reference genome of the closest Tephritid representative

Conclusions

Preliminary results suggest an increasing trend of DNA quantity over time, while the DNA quality and HTS performances show a constant trend.

The process of including more samples and performing statistical analyses is currently ongoing.

Perspectives

The WGS data and metadata obtained for more than 300 fruit fly vouchers have been archived on the RMCA servers and [DARWIN](#) (open-source natural history collections data management system). The setup of an open-access database for voucher archiving (including voucher metadata, genomic data, digital imaging and linked to the Global Genome Biodiversity Network – GGBN and Global Biodiversity Information Facility - GBIF) has been initiated and will be finalized by 2023 in the framework of the projects InsectMOoD and SYNTHESYS+ (respectively funded by the Belgian Science Policy and by the European Union).