EXECUTIVE SUMMARY

- Cost-effective approaches for the high throughput sequencing of insect museum collections have been defined
- A decision map on wet-lab procedures for the whole genome sequencing of specimens from the Diptera collections of RMCA has been produced.
- The DNA of more than > 1300 insect vouchers from the RMCA collections of Tephritidae and Syrphidae (Diptera) has been extracted, quality checked and subjected to Whole Genome Sequencing at genome coverage 10x.
- The resulting DNA was archived at RMCA through dedicated long-term stabilization protocols.
- Genomic vouchers for > 1300 insect specimens from the RMCA insect collections were produced, archived and linked to the collection, DNA, and digital vouchers.
- The backup and long-term storage of WGS data produced has been finalised in the Network-attached storage (NAS) systems of RMCA.
- The results of this project have been valorised through actions targeting both the scientific community (and, *inter alia*, via a publication on a peer-reviewed scientific journal) and the general public (by posting newsflashes on the institutional websites and on social medias)

ABSTRACT

Context

Museum vouchers from biological collections are of particular importance for scientific 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 technological advances over the past few years have led to a substantial reduction in costs, so that routine high throughput sequencing (HTS) of collection vouchers, including their whole genome sequencing (WGS), represents an exciting perspective for the valorisation of museum collections. However, museum specimens are often not directly suitable for genetic/genomic analyses due to low-quality DNA.

Objectives

InsectMOoD aimed at (a) providing a feasibility study for the large-scale WGS of Diptera from museum collections, (b) explore suitable approaches to create open-access, economically affordable and ready-to-use databases and repositories of genomic resources and (c) develop a decision map for the routine archiving of *genomic vouchers* as a complement to the routine archiving of morphological and digital vouchers at the Royal Museum for Central Africa. In this context, a main focus was given to hoverflies (Diptera, *Syrphidae*) and "true" fruit flies (Diptera, *Tephritidae*), two taxon groups for which RMCA has considerable taxonomic expertise and ongoing collection-based research.

Methods and Results

This project provided encouraging indications about the large-scale collection of genomic data from insect museum vouchers and resulted in the production of a consistent amount of "genomic vouchers" (> 1,300). These latter were represented by whole genomes recovered from collection vouchers and by their metadata (including, *inter alia*, information on protocols used for genomic library preparation and high throughput sequencing). The genomic vouchers produced are now linked to the corresponding morphological and digital vouchers as well as to the associated DNA collection vouchers. InsectMOoD indicates the suitability of a two-step approach, which consists in (a) the use of commercial kits and standard genomic library preparation protocols for cost- and time-effective genotyping of subsets of Museum vouchers and (b) more specialised protocols (e.g. including aDNA methodologies) to be used exclusively for the more problematic specimen subsets which did not yield satisfactory results with routine methodologies.

Conclusions

We believe that InsectMOoD provided a remarkable added value to the insect collections of RMCA . This approach allowed delivering a large bulk of easily accessible genetic information available in the framework of ongoing and future research on *Tephritidae* and *Syrphidae*. The optimization of experimental protocols and the collection of the genomic data, coordinated by the Joint Experimental Molecular Unit (JEMU) of RMCA and RBINS, generated guidelines of general interest for the WGS genotyping of material from the biological collections of RMCA and RBINS and further strengthened the expertise of the JEMU in Museomics. The multi-layer collection system advocated by this project will allow upgrading the RMCA standards on the collection of genomic data from museum vouchers.

(French and Dutch versions of this abstract and executive summary provided in Annex 1)

Keywords

Museum collections, Genomic Voucher, Collection management, High Throughput Sequencing, Whole Genome Sequencing

ABSTRACT

Context

Museum vouchers uit biologische collecties zijn van bijzonder belang voor wetenschappelijk onderzoek naar taxonomie, systematiek en biogeografie. Ze bieden middelen om een breed scala aan wetenschappelijke vragen in disciplines zoals ecologie, evolutie en conservatie aan te pakken. De snelle technologische vooruitgang van de afgelopen jaren heeft geleid tot een aanzienlijke verlaging van de kosten waardoor routinematige *high throughput sequencing* (HTS), inclusief *whole genome sequencing* (WGS), van collectievouchers een veelbelovend uitzicht biedt voor de valorisatie van museumcollecties. Museumspecimens zijn echter vaak niet direct geschikt voor genetische/genomische analyses vanwege de lage kwaliteit van het DNA.

Doelstellingen

De doelstellingen van InsectMOoD waren (a) een haalbaarheidsstudie uit te voeren voor grootschalige WGS van Diptera uit museumcollecties, (b) passende methoden te onderzoeken om open-access, economisch betaalbare en gebruiksklare databanken en opslagplaatsen van genomische informatie te creëren en (c) een leidraad op te stellen voor de routinematige archivering van genomische vouchers als aanvulling op de routinematige archivering van morfologische en digitale vouchers in het Koninklijk Museum voor Midden-Afrika. In deze context werd de nadruk gelegd op zweefvliegen (Diptera, *Syrphidae*) en "echte" fruitvliegen (Diptera, *Tephritidae*), twee taxongroepen waarvoor het KMMA over een aanzienlijke taxonomische expertise beschikt en actief onderzoek uitvoert gebaseerd op de collecties.

Methodes en resultaten

Dit project bracht veelbelovende bevindingen naar boven voor het verzamelen van genomische gegevens van insectenvouchers uit musea op grote schaal en resulteerde in de productie van een consistente hoeveelheid "genomische vouchers" (> 1300). Deze genomische vouchers bestonden uit hele genomen van collectievouchers en bijbehorende metadata (waaronder informatie over protocollen die gebruikt zijn voor genomische *library preparation* en *high throughput sequencing*). De geproduceerde genomische vouchers zijn nu gekoppeld aan de geassocieerde morfologische en digitale vouchers en aan de bijbehorende DNA-verzamelvouchers. InsectMOoD duidt op de doeltreffendheid van een tweefasige aanpak, die bestaat uit (a) het gebruik van commerciële kits en standaard protocollen voor genomische *library preparation* voor efficiënte genotypering in termen van kosten en tijd van subsets van museumvouchers en (b) meer gespecialiseerde protocollen (waaronder aDNA-methodologieën) die uitsluitend worden gebruikt voor de meer problematische specimen-subsets die onvoldoende resultaten opleverden met routinemethoden.

Conclusies

We zijn van mening dat InsectMOoD aanzienlijk heeft bijgedragen aan de insectencollecties van het KMMA. Deze aanpak maakte het mogelijk om veel gemakkelijk toegankelijke genetische informatie beschikbaar te stellen in het kader van lopend en toekomstig onderzoek naar *Tephritidae* en *Syrphidae*. De optimalisatie van experimentele protocollen en het verzamelen van de genomische gegevens, gecoördineerd door de Joint Experimental Molecular Unit (JEMU) van het KMMA en het RBINS, leverde richtlijnen op van algemeen belang voor de WGS-genotypering van materiaal uit de biologische collecties van het KMMA en het RBINS en versterkte de expertise van de JEMU in *Museomics*. Het collectiesysteem met meerdere niveaus dat in dit project wordt voorgesteld, zal bijdragen aan het verbeteren van de standaarden van het KMMA voor het verzamelen van genomische gegevens uit museumvouchers.

Sleutelwoorden

Museumcollecties, genomische voucher, Collectiebeheer, High Throughput Sequencing, Whole Genome Sequencing

RÉSUMÉ

Contexte

Les spécimens de musée provenant de collections biologiques revêtent une importance particulière pour la recherche scientifique sur la taxonomie, la systématique et la biogéographie et fournissent des outils pour aborder un large éventail de questions scientifiques dans des disciplines telles que l'écologie, l'évolution et la conservation. Les progrès technologiques rapides de ces dernières années ont conduit à une réduction substantielle des coûts, de sorte que le séquençage à haut débit (HTS) de routine, y compris le séquençage du génome entier (WGS), représente une perspective passionnante pour la valorisation des collections de musées. Cependant, les spécimens de musée ne sont souvent pas directement adaptés aux analyses génétiques/génomiques en raison de la faible qualité de l'ADN.

Objectifs du projet

InsectMOoD visait à (a) fournir une étude de faisabilité pour le WGS à grande échelle de Diptères provenant de collections de musées, (b) explorer des approches appropriées pour créer des bases de données et des dépôts de ressources génomiques en libre accès, économiquement abordables et prêts à l'emploi et (c) développer un arbre décisionnel pour l'archivage de routine des spécimens génomiques en complément de l'archivage de routine des spécimens morphologiques et numériques au Musée royal de l'Afrique centrale (MRAC). Dans ce contexte, l'accent a été mis sur les syrphes (Diptera, Syrphidae) et les "vraies" mouches des fruits (Diptera, Tephritidae), deux groupes pour lesquels le MRAC dispose d'une expertise taxonomique considérable et mène activement des recherches sur ses collections.

Méthodes et résultats

Ce projet a fourni des indications encourageantes sur la collecte à grande échelle de données génomiques à partir de spécimens de musées et a abouti à la production d'un nombre important de "spécimens génomiques" (> 1,300). Ces derniers étaient représentés par des génomes entiers récupérés à partir de spécimens de collection et des métadonnées associées (y compris, entre autres, des informations sur les protocoles utilisés pour la préparation de la bibliothèque génomique et le séquençage à haut débit). Les spécimens génomiques produits sont désormais liés aux spécimens morphologiques et numériques correspondants ainsi qu'aux spécimens d'ADN associés. InsectMOoD indique la pertinence d'une approche en deux étapes, qui consiste en (a) l'utilisation de kits commerciaux et de protocoles standard de préparation de bibliothèques génomiques pour le génotypage rentable et rapide de sous-ensembles de spécimens et (b) des protocoles plus spécialisés (par exemple, comprenant des méthodologies d'ADNa) à utiliser exclusivement pour les sous-ensembles de spécimens plus problématiques qui n'ont pas donné de résultats satisfaisants avec les méthodologies de routine.

Conclusions

Nous pensons que InsectMOoD a apporté une valeur ajoutée remarquable aux collections d'insectes du MRAC. Cette approche a permis de fournir une grande quantité d'informations génétiques facilement accessibles dans le cadre des recherches en cours et futures sur les Tephritidae et les Syrphidae. L'optimisation des protocoles expérimentaux et la collecte des données génomiques, coordonnées par l'unité moléculaire expérimentale commune (JEMU) du MRAC et de l'IRSBN, ont généré des lignes directrices d'intérêt général pour le génotypage WGS du matériel provenant des collections biologiques et ont renforcé l'expertise du JEMU en matière de muséomique. Le système de collecte multicouche préconisé par ce projet permettra d'améliorer les normes du MRAC sur la collecte de données génomiques à partir de spécimens de musées.

Mots-clés

Collections muséales, Spécimen génomique, Gestion des collections, Séquençage à haut débit, Séquençage du génome entier.