Joanna Stojak
PhD student at the Department of Genetics and Evolution of the Mammal Research Institute, Polish Academy of Sciences in Białowieża

Use of molecular methods in identification of species, age and sex of insects useful in forensic entomology

Summary

Forensic entomology uses insects to determine the time, cause and place of death. To this end, two entomological methods are used. The development-based method uses the patterns of insect larvae development under the specific thermal and environmental conditions. The succession-based method analyzes the sequence of insect succession on the body in various environmental conditions. The proper insect species identification is essential in both methods. In this article, the molecular methods of species, age and sex identification are presented such as DNA barcoding or DNA-HRM-PCR.

Keywords forensic entomology, DNA barcoding, mtDNA, molecular methods, identification of species

Introduction

Forensic entomology is a science that uses insects for the purposes of criminal justice. The process of decomposition, that begins at the time of death and continues until complete skeletonization, is a complex and irreversible process carried out by multiple microorganisms and animals.

The information about the time, place and cause of death is usually obtained after a forensic autopsy based on the assessment of early signs of death (e.g. livor mortis, rigor mortis, algor mortis), i.e. up to 72 hours after the cessation of vital signs [1–2]. After this time, methods offered by other fields of science, including forensic entomology, should be used.

Studying the sequence of succession and development of necrophillic insect species on corpses allows the estimation of the time of death, while the knowledge of the ranges of occurrence of each insect species and their environmental preferences allows to determine of the place of death and establish whether the corpse was moved [3–5].

At the early stages of decomposition, corpses are colonized mainly by true flies (*Diptera*), among which the most important family are blow flies (*Calliphoridae*). At the later stages of decomposition, beetles (*Coleoptera*) prevail [6]. It is suspected that this is related to the production of large quantities of ammonia by the *Diptera* larvae, which is toxic to the larvae of beetles [7]. There are examples of cases when the conditions unfavorable for the development of flies (e.g. too low temperature) allowed the beetles of the *Necrodes littoralis* species to carry out corpse decomposition by themselves [8]. During the collection of entomological specimens, species that accidentally appear on the corpse (including wasps, ants, butterflies) or are transferred by beetles via phoresis (mites) must not be ignored, as they are a valuable indication in determining the circumstances of death [9–12].

In the traditional approach, the identification of species is carried out in accordance with a special key for the determination based on morphological characteristics. This is often a laborious and complex task, and if only a fragment of an organism or only a juvenile form is available, it is often impossible to perform. The development of a method to facilitate the identification of species is a priority and should also involve other scientific fields, including the most dynamically developing field of molecular biology.

Identification of species using genetic markers

One of the methods of identification of species known as DNA barcoding was proposed in 2003 by Paul Hebert of the Institute of Biodiversity at the University of Guelph, Canada. The method was used for the analysis of hundreds of adult butterflies occurring in Costa Rica, morphologically very similar to each other, which were thought to represent one

species *Astraptes fulgerator* (Walch, 1775). The analysis revealed, however, that in fact the group was composed of 10 different species [13–14].

DNA barcoding uses short, standardized genetic marker sequences to identify the species from which these sequences originate. The selected sequences should be characterized by a low intraspecies variability (approx. 2%) and, at the same time, a high interspecies variability (approx. 10%) [13]. Designing primers for PCR amplification and the subsequent standardization are facilitated by a large number of repeats in the genome and flanking by conserved domains. Another important aspect is the length of the "barcode" sequence. Short fragments with a small number of insertions/deletions considerably accelerate and facilitate the identification.

In contrast to traditional methods, as little as fragments of organisms (e.g. feathers, bones) or developmental stages that are difficult to identify (e.g. larvae, seeds) are sufficient for the species determination based on the marker sequence [15].

However, finding one marker DNA sequence, which is standard for all organisms, has proven to be too complex and current studies are focused on finding the most appropriate DNA markers for the specific groups of organisms. In case of animals (including insects), a 648 bp fragment of the 5' terminus of the mitochondrial cytochrome c oxidase subunit I (COI) gene has proven to be the most appropriate.

The identification of species may also employ other regions of the mitochondrial DNA (the cytochrome b sequence encoding a protein included in the respiratory chain complex III and the control region, also known as D-loop, constituting the fastest-evolving part of the mitochondrial DNA), as well as nuclear markers.

Mitochondrial DNA is present in the cell in a large number of copies, its circular structure provides greater resistance to degradation and the genes (usually) do not undergo recombination. Therefore, only nuclear markers allow the identification and analysis of hybrids. In such analyses, ribosomal DNA genes (18S, 5.8S and 28S) are widely used. The genes are interspersed by the transcribed spacers ITS1 and ITS2 (internal transcribed spacers), used as marker sequences in fungi [16], plants [17–18] and animals [19–20].

However, in the era of the intense development of high-throughput sequencing techniques, when learning the genomic sequence of any organism appears to be easier and less expensive, the method of DNA barcoding faces a major challenge. Developing the database of barcode sequences of all living organisms will take many years of research and, to date, the analysis of DNA in the sample permits establishing only that the DNA originates from the species indicated by the database or another species, whose barcode sequence has not been determined.

Despite the need to refine the data, DNA barcoding is correctly considered as a fully functional tool for taxonomic analysis, useful in forensic entomology. However, when selecting the method of identification, it should be noted that the traditional methods based on the observation of morphological characteristics are still the least expensive, the fastest and usually do not involve the loss of the specimens.

DNA barcoding is becoming increasingly popular in identifying the species of insects that are important for forensic entomology, especially blow flies (*Calliphoridae*) and their larvae, while beetle species, that are equally important and much more difficult to identify, are wrongfully ignored.

Use of DNA barcoding in the identification of species important for forensic entomology

Schroeder et al. [21], in their analysis of COI and COII (cytochrome oxidase subunit II) fragments of 349 bp from three Calliphoridae species, amplified and subsequently digested with the restriction enzyme DraI, demonstrated that the sequence of *Lucilia sericata* (Meigen, 1826) differs from that of *Calliphora vicina* (Robineau-Desvoidy, 1830) and *Calliphora vomitoria* (Linnaeus, 1758) by 34 and 30 nucleotides, respectively. *Calliphora vicina* and *Calliphora vomitoria* differ by 15 nucleotides. The phylogenetic analyses by Wells and Sperling [22] based on the COI fragment revealed that the genetic distance within the individual species of Calliphoridae is 1% and less, but the distance between the species is 3% and more.

Meier et al. [23] investigated 1333 COI sequences from 449 Diptera species in order to determine the sequences that would best identify each species and to create ready-to-use tools for the potential identification of individuals found at the sites of corpse discovery. It was reported that 21% of species did not have any specific barcode sequence.

The COI analysis carried out by Rolo et al. [24] in 95 individuals of 7 species of the Muscidae and Calliphoridae families (*Musca autumnalis*, *Hydrotaea dentipes*, *Eudasyphora cyanella*, *Calliphora vicina*, *Calliphora vomitoria*, *Lucilia caesar*, *Pollenia rudis*) confirmed the usefulness of the method for the correct identification with 99–100% effectiveness. Satisfactory results were also obtained by Gil Arriortua et al. [25] who analyzed a 307 bp sequence of the cytochrome b gene in 185 individuals of *Calliphora vicina*, *Calliphora vomitoria*, *Lucilia sericata*, *Lucilia caesar*, *Lucilia ampullacea* and *Chrysomya albiceps*.

Equally popular are the analyses of the COI fragment of the Sarcophagidae true flies, whose identification is problematic not only in case of larvae (and their developmental stage), but also adult individuals [26]. Jordaens et al. [27] sequenced COI in 126 individuals from 56 species of Sarcophagidae from Western Europe, while Meiklejohn et al. [28–29], based on the

analysis of COI in 588 individuals, identified barcode sequences for 16 species of Sarcophagidae in Australia and developed an effective method of identification of *Sarcophaga impatiens* larvae (99.95% effectiveness).

Only two papers on DNA barcoding of beetles relevant for forensic entomology have been published, to date. In one of these papers, Schilthuizen et al. [30] analyzed beetles of the Cholevinae subfamily (*Coleoptera*: *Leiodidae*: *Cholevinae*). Due to the small size and difficulties in the determination of either larvae or adult forms, entomologists rarely use these beetles to estimate the time of death, even though they are frequently found on corpses. In the study, 86 individuals (only males, as in this case, the identification of species is very simple and limited to the isolation of the copulator) belonging to several genera (*Catops*, *Fissocatops*, *Apocatops*, *Choleva*, *Nargus*, *Ptomaphagus*, *Sciodrepoides*), collected in the Netherlands and France, were used. The conducted COI analyses proved that the molecular method significantly facilitates and accelerates the identification of species of the Cholevinae subfamily individuals: the interspecies genetic distance was as high as 9%, while the intraspecies genetic distance was 4% or less. Only *Catops nigricans* and *Catops fuscus* were not separated using this method.

Identification of species, age and sex using other molecular methods

An interesting method of identification of species, uniting several techniques, is DNA-HRM-PCR. It utilizes real-time PCR and the analysis of high-resolution DNA denaturation curves (high resolution melting, HRM). DNA-HRM-PCR enables a rapid detection and analysis of single-nucleotide polymorphisms (SNPs) in the obtained PCR products, based on the measurement of changes in fluorescence caused by the detachment of a dye that intercalats into DNA during the separation of the double strand (the so-called DNA melting profile is created). The method was used by Malewski et al. [31] for the identification of 16 Calliphoridae species occurring in Poland, relevant for forensic entomology. Using the PCR, two short COI fragments of 119 and 70 bp were amplified. The HRM analysis of several individuals of the same species demonstrated only minor differences in the denaturation curves. Moreover, the sensitivity and the accuracy of the method permitted the authors to analyze synthesized oligonucleotides of *Lucilia sericata* from Poland, France, England, India and the U.S. with respect to the geographical variability of genetic data, which very often complicates the correct identification.

What is more, assessing the size of genomes using flow cytometry (FCM) helps to identify not only the species, but also the sex of an individual. The FCM technique is based on the measurement of the fluorescence of cells, allowing the determination of the physico-biological properties of their components, such as nucleic acids. The method suits perfectly the species that are closely related to each other, enabling visualization of very subtle differences that would be difficult to pinpoint using other molecular methods. Importantly, flow cytometry is also ideal for identifying the species and the sex at larval stages. This was confirmed in the study by Picard et al. [32], in which the genome sizes of both sexes of 17 *Diptera* species relevant for forensic entomology were determined. These analyses demonstrated that the average genome size was from 425.8 Mb in case of *Chrysomya rufifacies* female to 1197.4 Mb in case of *Haematobia irritans* female.

Molecular methods are also applicable in determining the age of preimaginal forms of insects, especially that of problematic pupae in which the hard cocoon shell impedes the observation of the subtle changes occurring in the individual, and age determination based on morphology requires particular knowledge of the processes occurring during metamorphosis. These calculations are enabled by the analysis of age-dependent gene expression (developmental stage) in differentially expressed genes (DEGs), involving the creation of species-specific profiles presenting the cascade of hierarchically activated genes that regulate the development of the individual. The subsequent correlation of the transcriptome of the insect with the individual and the points of the full gene expression profile allows the determination of the developmental stage and thus, age estimation. The principle of these analyses is the qRT-PCR method (quantitative real-time PCR) which allows determining the quantity of the product in real time during its synthesis (simultaneous amplification and quantity monitoring after every cycle) [33]. To date, DEGs profiles have been established for two species relevant for forensic entomology. In the case of *Lucilia sericata*, the identified DEGs did not allow the correct age estimation of pupae [34, 35], while a study of DEGs in *Calliphora vicina* demonstrated that the method can be an alternative to traditional methods for estimating the age of insects [33].

The same method was also employed in the assessment of the age of eggs deposed by *Lucilia sericata*, using the expression of three genes: *bcd* (observed at an early developmental stage), *sll* (determining the dorsoabdominal coloration pattern, with a high expression observed in salivary glands) and *cs* (determining the correct chitin synthesis during the development of larval cuticle). Due to the variations in the level of expression of the three genes, depending on the time elapsed since the oviposition, the oval stage lasting several hours may be subdivided into shorter periods, thereby permitting a precise age determination [34].

When estimating the time of death and determining the age of preimaginal forms, the environmental conditions

are of key importance. A too low temperature may cause a temporary delay in the larval development (diapause) which, when disregarded in the calculations, generates serious errors. To date, the lack of morphological differences between a normally developing larva and a larva in diapause has prevented the development of an effective distinction method. Comparing the gene expression profiles of larvae in and out of diapause has allowed finding genes that are characteristic of this process. In *Sarcophaga crassipolpis*, increased expression is observed for genes such as *hsp23* and *hsp70* [36, 37], in *Lucilia cuprina*, e.g. *hsp23* and *hsp24* [38], and in *Calliphora vicina*, e.g. *hsp23*, *hsp24*, *hsp70* and AFBP [39].

The analysis of age-dependent gene expression is commonly used in developmental molecular biology, and the employed techniques and tools can be easily introduced in the laboratories conducting forensic DNA analyses. The use of microarrays and fully automated methods increases analysis throughput and reduces the risk of error, allowing for an accurate prediction of error level in the obtained results and meeting the principles of the Daubert standard [34].

Conclusions

The exact and correct identification of species useful in forensic entomology is of great importance, as any error in the analysis may lead to raising charges against an innocent person with the true perpetrator remaining at large. Forensic entomology is practiced in many countries, with the highest level of advancement in the United States, France and Canada. Only 70 fully trained researchers working exclusively in this specialty may be found worldwide [40]. The results obtained from the observation of corpse decomposition (carried out by forensic anthropologists, e.g., at the University of Tennessee, U.S.A.) can be transferred to real criminal cases, and the National Oceanic and Atmospheric Administration (NOAA), providing (for a small fee) temperature data for almost all regions, is a source of information extremely valuable in estimating the time of death [41].

In Poland, the possible use of necrophillic insects in determining the time of death is investigated by six centers located in Poznan/Toruń, Warsaw, Gdańsk, Kraków, Łódz and Katowice. There is no doubt that molecular methods substantially raise the significance of entomological analyses and constitute their valuable complement.

Acknowledgments

I would like to thank the anonymous reviewer for the valuable comments that helped to raise the scientific value of this publication.

Translation *Rafał Wierzchosławski*