Species identification of forensically important flies using DNA barcoding

Kurt Jordaens[1], Gontran Sonet[1], René Richet[2], Erena Dupont[3], Yves Braet[3] and Stijn Desmyter[3]

[1]Joint Experimental Molecular Unit, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium and Royal Belgian Institute of Natural Sciences, Vautierstraat 29, 1000 Brussels, Belgium
[2]16, Grande Rue; F-03220 Jaligny-sur-Besbre, France
[3]National Institute of Criminalistics and Criminology, Vilvoordsesteenweg 100, 1120 Brussels, BelgiumRoyal Belgian Institute of Natural Sciences, Joint Experimental Molecular Unit (JEMU), Vautierstraat 29, B-1000 Brussels, Belgium

Introduction

Accurate identification of an insect specimen is usually a crucial first step in a forensic entomological analysis because closely related carrion species may differ substantially in growth rate and diapause response. However, species-diagnostic anatomical characters are often difficult to use since 1) they are not known for most immature stages, 2) existing keys may be incomplete and difficult to use for nonspecialists and 3) the available material could be strongly degraded or fragmented. A DNA-based identification tool may largely overcome these problems.

Here, we present, and analyse, the first reference database for West European *Sarcophaga* species (Sarcophagidae) of forensic interest.




Material and Methods

-145 specimens of 50 West European *Sarcophaga* species sequenced for the entire COI gene (whenever possible)
- additional *Sarcophaga* sequences collected from GenBank and BOLD
- total dataset of 270 specimens from 99 species
- Different fragment lengths analysed: entire COI gene - standard barcode fragment - minibarcode fragment
- NJ-tree (K2P-model), optimal intra-interspecific sequence divergence threshold according to Lefébure et al. MPE, 2006, 40: 435-447
- identification (ID) success determined using Best Match (BM) and Best Close Match (BCM) criteria

Results BM criterium

Number of sequences (N seq), number of species (N sp) and the no of species for which no conspecific sequence was available in the dataset (N no con) of the different *Sarcophaga* datasets and the ID success according to the BM criterium

	N seq	N sp	N no con	correct	ambiguous	incorrect
entire COI	173	82	36	135 (78%)	1 (0.6%)	37 (21.4%)
barcode	270	99	42	214 (79.3%)	10 (3.7%)	46 (17%)
mini-barcode	270	99	42	163 (60.4%)	70 (26%)	37 (13.6%)

Results BCM criterium

Number of sequences (N seq), number of species (N sp) and the no of species for which no conspecific sequence was available in the dataset (N no con) of the different *Sarcophaga* datasets and the ID success according to the BCM criterium; TH = threshold

	N seq	N sp	N no con	correct	ambiguous	incorrect but match < TH	incorrect but no match < TH
entire COI	173	82	36	135 (78%)	1 (0.6%)	14 (8.1%)	23 (13.3%)
barcode	270	99	42	214 (79.3%)	10 (3.7%)	32 (11.9%)	14 (5.2%)
mini-barcode	270	99	42	159 (60.4%)	63 (23.3%)	21 (7.8%)	27 (10%)

wrong identifications

incomplete database

General results

1. the entire COI gene and barcode region have a similar ID success (Result tables)
2. the minibarcode fragment has a much lower ID success
3. part of the incorrect IDs can be explained by a lack of conspecifics in the reference database
4. DNA barcoding reveals potential cryptic diversity (Figure)

Conclusions

promising results:
- molecular identification for 80% of the species
- identification of females and larvae possible

but:
- reference database is incomplete
- possibly high within-species variation (geographical / cryptic diversity)
- minibarcode fragment performs worser than longer fragments

Part of the NJ-tree showing potential cryptic diversity within *Sarcophaga dux*