Chapter 2
Identification Tools for African Frugivorous Fruit Flies (Diptera: Tephritidae)

Massimiliano Virgilio

Abstract The current classification of African tephritids is the interim result of a continuous process of minor and major changes that, in the last 20 years, has resulted in the description of more than 60 new species from the seven tephritid genera of main economic relevance in Africa (*Bactrocera*, *Capparimyia*, *Ceratitis*, *Dacus*, *Neoceratitis*, *Trirhithrum* and *Zeugodacus*). In this context of dynamic change, rapid and accurate fruit fly identification is critical, particularly with respect to the early detection of pest invasions. Valuable resources for fruit fly identification include: the tephritid reference collections and repositories distributed within and outside the African continent; publicly available online databases; and the single- and multi-entry keys for the morphological identification of African tephritids. Identification through DNA barcoding represents a cost effective tool for the molecular diagnosis of African fruit flies and it has proved particularly useful for the identification of immature stages, of damaged specimens and of incomplete specimens. The molecular diagnosis of tephritids also represents a partial solution to the gradual loss of taxonomical expertise on this and other insect groups. In this chapter the advantages and limitations of the available identification tools and resources are discussed.

Keywords Morphological identification • Natural history collections • Online databases • Identification keys • Molecular diagnosis • DNA barcoding

1 Introduction

Tephritid fruit flies, or 'true' fruit flies (Diptera: Tephritidae), include approximately 500 genera and 4800 valid species, the vast majority (95 %) of which are phytophagous (Aluja and Norrbom 1999). Of all tephritid species 25-30 % are frugivorous. In Africa there are approximately 400 species of frugivorous tephritids of which more

M. Virgilio (✉)
Department of Biology and Joint Experimental Molecular Unit (JEMU), Royal Museum for Central Africa, Tervuren, Belgium
e-mail: massimiliano.virgilio@africamuseum.be

© Springer International Publishing Switzerland 2016

S. Ekesi et al. (eds.), *Fruit Fly Research and Development in Africa - Towards a Sustainable Management Strategy to Improve Horticulture*,
DOI 10.1007/978-3-319-43226-7_2

than 50 are economically important (list provided in Virgilio et al. 2014). The current classification of African tephritids is the interim result of a continuous process of minor and major updating; in just the last 20 years this has included:

- a monograph on the genera *Dacus* and *Bactrocera* from Africa and the Middle East (White 2006) with the genus (Hancock and Drew 2006)
- a revision of the *Ceratitis* subgenera *Acropteromma* and Hoplolophomyia (De Meyer and Copeland 2001), *Ceratalaspis* (De Meyer 1998), *Ceratitis s.s.* (De Meyer 2000), *Pardalaspis* (De Meyer 1996) and *Pterandrus* (De Meyer and Freidberg 2006)
- a revision of the genera *Capparimyia* (De Meyer and Freidberg 2005), *Carpophthoromyia* (De Meyer 2006), *Neoceratitis* (De Meyer and Freidberg 2012), *Perilampsis* (De Meyer 2009) and *Trirhithrum* (White et al. 2003)
- the description of a new pest species: *Bactrocera invadens* Drew et al. (2005)
- a monograph on the genera *Dacus* and *Bactrocera* from Africa and the Middle East (White 2006)
- a revised classification of subgenera and species groups within the genus *Dacus* (Hancock and Drew 2006)
- the description of 17 new *Dacus* species (White and Goodger 2009)
- an analysis of the biodiversity of the western African fauna including the description of a new *Dacus* species (De Meyer et al. 2013)
- the synonimisation of the key pests *Bactrocera invadens* and *Bactrocera. dorsalis* (Hendel) (Schutze et al. 2015)
- a novel generic combination for *Zeugodacus cucurbitae* (Coquillett) (Virgilio et al. 2015)

In this context of dynamic change, rapid and accurate fruit fly identification is critical, particularly with respect to the early detection of pest invasions. For example, in 1995 the incorrect identification of *Bactrocera zonata* (Saunders) as *Bactrocera pallidus* (Perkins and May) in Egypt lead to a three-year delay in the implementation of phytosanitary measures and resulted in serious damage to the agricultural productivity of the whole Alexandria region (Abuel-Ela et al. 1998).

2 Online Databases

The tephritid reference collections and repositories are a valuable resource for fruit fly identification as well as for the training of specialist and non-specialist taxonomists. African researchers can confidently rely on what is a relatively limited number of comprehensive reference collections in the continent which include: the South African National Collection of Insects, Pretoria, South Africa; the collections of the National Museums of Kenya, Nairobi, Kenya; the collection of the Natural History Museum of Zimbabwe, Bulawayo, Zimbabwe; and the collection of the International Institute of Tropical Agriculture (IITA), Cotonou, Benin.

Outside of Africa, one of the largest collections of African frugivorous flies is held at the Royal Museum for Central Africa (RMCA), Tervuren, Belgium; this

collection currently includes some 5000 African specimens from approximately 200 species from ten tephritid genera. Detailed information about vouchers available in the RMCA frugivorous tephritid collection can be directly accessed through the 'True fruit flies of the Afrotropical Region' database (http://projects.bebif.be/fruitfly/index.html). This database is part of the Global Biodiversity Information Facility (GBIF, http://www.gbif.org), a platform that aims to provide open access to biodiversity data and is hosted within the Belgian GBIF node (BeBIF, http://www.bebif.be). This database also has information on reference material from African fruit fly species in the genera *Ceratitis, Dacus, Bactrocera, Capparimyia, Trirhithrum, Carpophthoromyia* and *Perilampsis*, that is available from other European, North American and African museums and research institutions. The BeBIF fruit fly database has 150,000 specimens, in excess of 16,000 block records (ie sets of specimens with identical data), material from 60 institutions and private collections, historical collections (eg type collections and collections from USDA expeditions to Africa) and associated data; the associated data include details of approximately 3000 georeferenced localities, 700 host plant records and more than 1500 digital images and maps of sampling locations. Taxon information in BeBIF includes: (a) the current valid taxonomic name of each species and a list of synonyms where applicable; (b) a short taxonomic description of the species based on available taxonomic revisions; (c) a set of images (photographs or drawings) depicting the main morphological characteristics of the species that are taken in a uniform and standardized way in order to facilitate comparison; and (d) a geographical distribution map for each species that is directly linked to specimen information. All relevant data that are linked to individual vouchers or block records are provided and include: place where and the date when the specimen was collected, the name of the collector, the collection where the specimen is deposited and the status of the specimen (type or non-type). Other additional information that can also prove useful for identification are the response to lures and attractants, (which are generally specific at the genus or subgenus level), and the range of host plants attacked by the species.

3 Keys for Morphological Identification

Morphological identification of African fruit flies can be achieved using a range of methods that differ in their technical complexity and reliability. Dichotomous identification keys are generally only accessible by users with existing background knowledge of tephritid morphology and the often-complicated technical terminology used, and who also have access to specialised equipment such as dissection tools and microscopes. Alternative identification tools of more general use include simplified keys for a number of the economically relevant African pests (Ekesi and Billah 2007), identification sheets and online material for the identification of invasive fruit flies in Africa (eg www.africamuseum.be/fruitfly/AfroAsia.htm). These tools, under certain circumstances, can be useful to the large number of untrained

users, such as farmers, who are keen to detect pests on their crops. Of course, although these general tools are rapid, they can also be inaccurate when dealing with the less common species.

Classical single-entry (dichotomous) keys are available for most African Dacini. White (2006) produced a dichotomous key with a revised classification for the African and Middle Eastern species of *Bactrocera* (15 species) and *Dacus* (177 species), a database of digital images for 190 species and a database with notes on the identification of pest species. The revised classification of White (2006) was partly based on a cladistic analysis that explored the subgeneric relationships of a subset of representative species; this facilitated a number of advances including the description of 25 new *Dacus and Bactrocera* species, the synonymisation of 26 species and the removal from synonymy of two species. Since the work of White (2006) more new species and changes in synonymy have occurred which are not in the original dichotomous key. For example 17 new species of *Dacus* have been described and two synonymised by White and Goodger (2009). Similarly, Hancock and Drew (2006) produced a revised classification and a dichotomous key that could be useful for the identification of *Dacus* subgenera and species groups.

Dichotomous identification keys are also available for the genus *Ceratitis*; there are four stand-alone subgeneric keys and revisions including: (a) a key to the subgenus *Ceratitis* (*Pardalaspis*) Bezzi (De Meyer 1996) with ten Afrotropical species and information about species distribution and host plants, (b) a key to the subgenus *Ceratitis* (*Ceratalaspis*) Hancock (De Meyer 1998) with 36 species and illustrations of mesonotal and wing patterns, shape of the aculeus tip, distribution and known host plant data, and tentative species groups within the subgenus, (c) a key to eight species of the subgenus *Ceratitis* Macleay *s.s* with illustrations of cephalic bristles, mesonotal and wing patterns and aculeus shape (De Meyer 2000) and (d) a key to 36 species, of the subgenus *Ceratitis* (*Pterandrus*) Bezzi with information about species distribution and host plant data, tentative species groups within the subgenus and illustrations of male and female terminalia, wing and mesonotal patterns and male leg ornamentation (De Meyer and Freidberg 2006).

The dichotomous key to the genus *Trirhithrum* Bezzi (White et al. 2003) allows identification of 40 *Trirhithrum* species and a further seven taxa of uncertain status. The revision published with the key provides host data, largely from a survey in Kenya. The small genus *Capparimyia* Bezzi was revised by De Meyer and Freidberg (2005) who recognized eight species and provided a dichotomous key with illustrations of mesonotal and wing patterns and male and female terminalia. The 17 species of the genus *Carpophthoromyia* Austen can also be identified using the dichotomous key of De Meyer (2006) that also provides illustrations of wing patterns and both male and female terminalia. The dichotomous key to the genus *Perilampsis* Bezzi (De Meyer 2009) includes 17 species and provides illustrations of wing patterns, female terminalia and information about host specificity. The genus *Neoceratitis* Hendel can be identified using the dichotomous key of De Meyer and Freidberg (2012) and includes six species with illustrations and host information.

One of the main limitations of dichotomous single-entry keys is that species identification is not possible when the user is unable to distinguish between one of

the dichotomous options provided by in the key. This can occur if the specimen is damaged so that the morphological character is not present or easily recognisable, if the user has inadequate taxonomic expertise, or if there is a lack of clarity in the key. The terminology used in some published keys can represent a serious obstacle for non-specialists who are not well acquainted with insect morphology and taxonomy. In fact, many terms used to describe morphological variation, such as small/ large, dark/ pale, thick/ thin etc. could be considered subjective and unclear to non-specialist users (while specialist taxonomists generally find these definitions straightforward because they have the necessary and essential experience that comes from examining large numbers of specimens). In this respect, multi-entry identification keys might overcome some of the technical difficulties associated with dichotomous keys. As the name suggests, multi-entry identification keys allow identification via multiple paths such that the user has the ability to 'skip' problematic questions and score alternative characters.

Additionally, there is no comprehensive key to all genera of African fruit flies so that non-specialised users might even find it problematic to assign specimens to the genera for which dichotomous keys are available (but see Hancock and White 1997 for a key to distinguish the genus *Trirhithrum* from others in the *Ceratitis* group of genera). This issue is even more relevant for the genus *Ceratitis*, where additional subgenus identification is also necessary before it is possible to use one of the six dichotomous keys available.

The development of a user-friendly set of multi-entry identification keys for African tephritids began in 1999 with a pilot project supported by the U.S. Agency for International Development (USAID, PCE-G-00-98-0048-00) and by the U.S. Department of Agriculture (USDA) / National Institute of Food and Agriculture (CSREES) / Initiative for Future Agricultural and Food Systems (IFAFS) grants to the Texas A&M University (00- 52,103–9651). This resulted in an initial set of two keys for the identification of *Ceratitis* and *Trirhithrum* species, through the CABIKEY platform. Later on, a project co-funded by the Belgian Directorate-General for Development Cooperation (through a framework agreement with the Royal Museum for Central Africa) and the International Atomic Energy Agency (IAEA – Vienna, project 'Development of a Web Based Multi Entry Key for Fruit Infesting Tephritidae', contract number 16,859) allowed development of a set of multi-entry identification keys for African frugivorous flies (Virgilio et al. 2014). These keys included a 'pre-key' for genus designation (built *ex novo* using a set of 23 characters that were deemed to be informative for separation of genera) as well as seven multi-entry keys for species identification within a genus or a group of genera (*Bactrocera*+*Dacus*+*Zeugodacus*, *Capparimyia*, *Carpophthoromyia*, *Ceratitis*, *Neoceratitis*, *Perilampsis*, *Trirhithrum*) and including a total of approximately 390 taxa. In this set of keys species lists and morphological characters were revised and optimised to include only species with (a) valid names under the International Code of Zoological Nomenclature and (b) characters including at least two states in congeneric species (Virgilio et al. 2014). The keys were based on eight matrices containing scores for a total of 368 characters and were compiled from data sets that were used within the framework of the taxonomic revisions described

above (De Meyer 1996, 1998, 2000, 2006, 2009; White et al. 2003; De Meyer and Freidberg 2005, 2006, 2012; White 2006; White and Goodger 2009). The keys are regularly updated in order to keep pace with changes in the taxonomic status of species and take into account, for example, the recent synonimisation of *B. invadens* and *B. dorsalis* (Schutze et al. 2015), and the novel generic status of *Z. cucurbitae* (Virgilio et al. 2015). To facilitate identification, morphological characters were grouped as sets from the head, thorax, wings, legs and abdomen respectively. Unfolding characters were also included, i.e. those characters that are initially hidden but appear when only a pre-defined subset of species remain to be identified (unfolding keys). Dependencies between characters were also generated; positive dependencies were defined whenever a character was only meaningful in relation to a previously defined character state (eg in the *Ceratitis* key, the morphological character 'number of frontal setae' is positively dependent on the character state -'frontal setae: yes). Conversely, negative dependencies were generated to discard characters that were not meaningful after a previous character state was selected (eg in the *Ceratitis* key, the character 'females: aculeus tip with small notch' is negatively dependent on the character state 'sex: male'). Embedded within the keys are images that illustrate, name and position each character on the insect body. There are also images showing how the same character appears in different species. The initial set of 2300 images and drawings recovered from the databases of the Royal Museum for Central Africa (RMCA) and from the London Natural History Museum (NHM) were rearranged according to species name and body part (head, thorax dorsal, thorax lateral, abdomen, wings, legs), divided into groups and assigned to each combination of character state and species name. This generated a database of approximately 20,000 images that illustrate the phenotypic variability of the same character across species and provides a 'virtual collection' of images that are rapidly accessible. Furthermore, the largest keys (*Bactrocera*/*Dacus*/*Zeugodacus*, *Ceratitis*, *Trirhithrum*) allow the user to distinguish between different subsets of morphological characters including (1) characters that are the most straightforward to identify; (2) all characters except those that are the most difficult to identify; and (3) all characters, including the 'easy', 'average' and 'difficult' ones. The user has the opportunity to first consider only characters that are straightforward to use, and then follow this up by using characters that are increasingly more difficult to interpret. This process facilitates identification and reduces the risk of misidentification, particularly for species that can be identified using straightforward characters only. The keys also allow identification to be restricted to species of economic importance only. The use of this option should speed up identification of the more commonly trapped / intercepted taxa. However, when using this option, identification should be further verified (eg through an in-depth analysis of the species description – see below) as less common species not included in this option could be erroneously identified as species of economic importance (false positives). The keys also provide (a) species descriptions as provided by the published scientific literature, (b) images from the RMCA and NHM tephritid collections and (c) hyperlinks to the Encyclopedia of Life (EOL), the Belgian Biodiversity Platform (BeBIF) and, when available, to the Barcoding of Life Database (BOLD).

4 Molecular identification through DNA barcoding

DNA barcoding provides a rapid and often effective tool for the molecular diagnosis of species and it has proved to be particularly useful for specimens (or parts of specimens) where distinguishing morphological characteristics are degraded or missing (Hebert et al. 2003; Nagy et al. 2013). DNA barcoding is a distance-based identification method that relies on reference libraries of DNA sequences from unambiguously identified voucher specimens. The most widely used DNA barcode for animal identification is a standardised 648 base-pair region of the mitochondrial cytochrome c oxidase subunit I (COI) while other gene fragments are used for plants (Ribulose-bisphosphate carboxylase [rbcl] and Maturase K [matK]) and fungi (the Internal Transcribed Spacer Region [ITS]). DNA barcoding identification basically relies on (1) calculating the genetic distance between the target DNA sequence of an unidentified specimen (a query) and sequences from the reference library of DNA barcodes and (2) assigning to the query the species name of the most genetically similar reference DNA barcode (ie having the smallest genetic distance from the query) (Hebert et al. 2003; Ratnasingham and Hebert 2007). A number of DNA barcoding bodies and resources are available and include (1) the Consortium for the Barcode of Life (CBOL; http://www.barcodeoflife.org) which promotes DNA barcoding via institutions from over 50 countries and operates out of the Smithsonian Institute's National Museum of Natural History in Washington; (2) the International Barcode of Life (iBOL, http://www.ibol.org) which involves numerous countries in the global barcoding effort and; (3) BOLD (http://www.boldsystems.org) which is an online workbench and the main platform for DNA barcoding identification (reviewed in Taylor and Harris 2012). BOLD is the main barcode repository and provides analytical tools; an interface for submission of sequences to GenBank; species identification tools; and connectivity for external web developers and bioinformaticians (Ratnasingham and Hebert 2007). Each reference DNA barcode in BOLD is linked to specimen information including, *inter alia*, the species name (or its interim), voucher data (catalogue number and institution storage reference), collection records (collector, collection date and location with GPS coordinates) and the name of the person who identified the specimen. The DNA barcoding data associated with animal specimens includes the COI sequence (of at least 500 bp), the polymerase chain reaction (PCR) primers used to generate the amplicon and the sequence forward and reverse trace files (Ratnasingham and Hebert 2007). The DNA barcoding identification tool in BOLD reports the genetic similarity between the query and a list of the best DNA barcode matches in a table of similarity scores (%) and visualizes the distances between the query and its best matches in a neighbor-joining tree reconstruction.

DNA barcoding of fruit flies (eg Meeyen et al. 2014) might indeed represent a feasible and complementary solution to the gradual loss of taxonomical expertise on this and other insect groups (de Carvalho et al. 2007). For immature stages of most fruit fly species and for damaged specimens DNA barcoding is the only available identification tool; for this reason it has potential for routine identification of fruit

Table 2.1 Reference DNA barcodes available in the Barcoding of Life Database (http://www.boldsystems.org, 02/07/2015) across tephritid subfamilies

Subfamilies	Specimens with barcodes	Number of taxonomic entities with barcodes
Dacinae	4530	395
Phytalmiinae	11	6
Tachiniscinae	1	1
Tephritinae	1426	249
Trypetinae	1443	148
Total	7411	799

fly interceptions (Armstrong and Ball 2005; Barr et al. 2012; Boykin et al. 2012). Despite this potential (but see Moritz and Cicero 2004; Cameron et al. 2006; Taylor and Harris 2012; Kvist 2013; Pečnikar and Buzan 2014), DNA barcoding is still not widely used for tephritid identification due to a number of issues associated with the incomplete taxon coverage of the available reference libraries (Virgilio et al. 2010; Kwong et al. 2012; Virgilio et al. 2012; Smit et al. 2013) as well as with difficulties in resolving important species complexes of economic interest (Frey et al. 2013) such as the *Bactrocera dorsalis* (Hendel) (Jiang et al. 2014) or the *Ceratitis* FAR (Virgilio et al. 2012) complexes or even failure to differentiate between closely related species for which there are distinct morphological characters to separate the adults *(eg Ceratitis capitata and Ceratitis caetrata, see Barr et al.* 2012). In 2007 the Consortium for BOLD initiated and supported the Tephritid Barcoding Initiative (TBI), a two year demonstration project to populate the reference database of DNA barcodes for fruit flies and develop protocols for queries in support of pest management, ecology and taxonomy. An analysis of the status of the BOLD libraries (updated 2nd of July 2015) with respect to current taxon coverage for tephritid fruit flies reveals that more than 7000 tephritid vouchers had been barcoded for a total of approximately 800 taxonomic entities including both valid species and interim identifications (the latter representing a relevant 22 % of the tephritid taxa in BOLD). Almost half of the barcoded taxonomic entities (49.4 %) belong to the subfamily Dacinae and include the seven tephritid genera of main economic relevance in Africa (*Bactrocera*, *Capparimyia*, *Ceratitis*, *Dacus*, *Neoceratitis*, *Trirhithrum* and *Zeugodacus*) as well as of the two related genera (*Carpophthoromyia* and *Perilampsis*). These genera alone include 94.9 % of all barcoded Dacinae taxonomic entities (corresponding to 98.9 % of all Dacinae specimens in BOLD) (Table 2.1).

There are more than 2200 reference DNA barcodes for the five genera of major economic importance in Africa viz. *Bactrocera*, *Zeugodacus*, *Dacus, Ceratitis* and *Trirhithrum* (see Virgilio et al. 2014 for a list of the main African pests). Economically important *Zeugodacus* and *Bactrocera* species viz. *Z. cucurbitae*, *B. dorsalis*, *B. latifrons*, *B. oleae* and *B. zonata*, are all represented by multiple reference DNA barcodes (with more than 1300 DNA barcodes available in total with an average of 226.8, SD = 220.3 DNA barcodes per species). There are more than 170 reference DNA barcodes in the BOLD libraries (average per species = 10.8, SD = 16.6) for

Table 2.2 Reference DNA barcodes available in the Barcoding of Life Database (http://www.boldsystems.org, 02/07/2015) across genera in the subfamily *Dacinae*

Genera within *Dacinae* available in BOLD (02/07/2015)	Specimens with barcodes	Number of taxonomic entities with barcodes	Number of interim species with barcodes	% interim species	Number of economically important species
Bactrocera	2667	197	66	33.5	5
Zeugodacus	284	1	0	0.0	1
Dacus	423	78	3	3.8	16
Ceratitis	937	57	10	17.5	21
Trirhithrum	82	18	2	11.1	8
Capparimyia	28	7	1	14.3	1
Carpophthoromyia	30	8	2	25.0	0
Neoceratitis	2	2	0	0.0	1
Perilampsis	26	7	0	0.0	0
Acanthiophilus	14	2	1	50	
Acroceratitis	2	2	0	0	
Acrotaeniostola	1	1	0	0	
Bistrispinaria	2	1	0	0	
Capitites	1	1	0	0	
Celidodacus	6	4	1	25	
Clinotaenia	4	3	1	33.3	
Cyrtostola	1	1	0	0	
Dectodesis	9	1	0	0	
Euarestella	2	1	0	0	
Gastrozona	4	1	0	0	
Taeniostola	3	1	0	0	
Urelliosoma	2	1	0	0	
Xanthorrachista	0	0			
Total	4530	395	87	22.0	

fourteen of the sixteen *Dacus* species that are pests in Africa. Despite this, for four species (*D. annulatus*, *D. limbipennis*, *D. lounsbutyii*, *D. persicus*) only a single reference barcode is currently available. Similarly, 18 of the 21 African *Ceratitis* pest species are represented in the BOLD libraries and there are multiple reference sequences (average per species = 31.4, SD = 66.8) for all of them except *C. penniciliata*. Of the eight *Trirhithrum* pest species all but *T. albomaculatum*, *T. basale* and *T. manganum* are represented in the BOLD libraries, all with multiple reference DNA barcodes (average per species = 4.3, SD = 5.2) (Table 2.2).

The completeness of the reference libraries remains a critical issue as, obviously, any query without a conspecific reference DNA barcode in the library cannot be correctly identified (Virgilio et al. 2010; Smit et al. 2013). A distance threshold can be defined such that a query is discarded (ie its identification considered unreliable) whenever the distance between the query and its best DNA barcode match exceeds

the threshold value (according to the Best Close Match criterion, see Meier et al. 2006). This reduces the probability that queries that are not represented in the library by a conspecific will be incorrectly identified with the 'closest' (ie most genetically similar) allospecific match. The outcomes of distance threshold based DNA barcoding can be categorised as: (1) true positives (TP), ie queries that are correctly identified with a genetic distance to their best match that is below the threshold; (2) false positives (FP), ie queries that are misidentified despite the distance to their best match remaining below the threshold; (3) true negatives (TN), misidentified queries that are correctly discarded because the distance to their best match is above the threshold and; (4) false negatives (FN), correctly identified queries that are discarded in error as the distance to their best match is above the threshold. Distinguishing amongst these categories allows the user to quantify the level of accuracy (TP + TN/number of queries), precision (TP/ (TP + FP)), overall identification error (FP + FN/number of queries) and the relative identification error (FP/ TP + FP) of the DNA barcoding identification method. Several criteria for setting the distance thresholds have been proposed (eg Meyer and Paulay 2005). Fixed distance thresholds were common in early barcoding studies (eg Hebert et al. 2003) and were initially implemented in BOLD where a 1 % sequence dissimilarity (ie the fraction of base mismatches between two sequences) represented the cut-off value for identification (Ratnasingham and Hebert 2007). Of course, no single interspecific distance threshold fits all taxonomic groups as coalescent depths amongst species vary due to differences in population size, rate of mutation and time since speciation (Collins and Cruickshank 2013). A number of distance thresholds can be generated directly from the data so that cut-off values change according to the particular reference library / taxon group being considered (Meyer and Paulay 2005; Meier et al. 2006; Puillandre et al. 2011; Virgilio et al. 2012). Initially, a 'ten times' rule was proposed (Hebert et al. 2004) to determine a threshold value as calculated from the distribution of intraspecific distances (but see Hickerson et al. 2006 for criticism). Sonet et al. (2013) developed an R package to calculate *ad hoc* distance thresholds producing identifications with an estimated relative error probability that could be fixed by the user (eg 5 %) (Virgilio et al. 2012). BOLD is now implementing a Barcode Index Number (BIN) algorithm that uses a 2.2 % sequence dissimilarity threshold with subsequent refinement using Markov clustering (Ratnasingham and Hebert 2013). Other statistical approaches, aimed at reducing the limits of distance-based identification have been proposed (eg Nielsen and Matz 2006; Tanabe and Toju 2013; Dowton et al. 2014; Porter et al. 2014). However, performing complex statistics on libraries that include millions of reference barcodes still remains computationally challenging. Furthermore, users willing to adopt alternative approaches and criteria for DNA barcoding identification generally need to build their own reference library, and this is not always possible as not all BOLD reference DNA barcodes are publically available.

Producing fruit fly DNA barcodes is a relatively straightforward process when starting from common, recently collected and adequately preserved fruit fly specimens. In these cases, DNA barcodes can generally be obtained using universal DNA primers (Folmer et al. 1994) and standard or slightly modified protocols for DNA

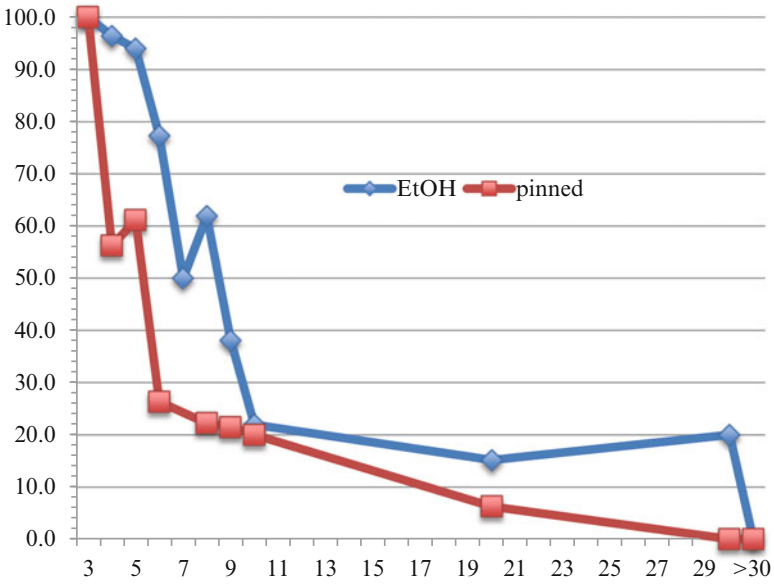

Fig. 2.1 Percentage of DNA barcodes obtained using standard protocols on EtOH – preserved and pinned specimens of different ages (Virgilio and De Meyer, unpublished)

extraction, amplification and sequencing (Barr et al. 2012). However, obtaining DNA barcodes from the less common African fruit flies or for species not commonly found in crop production areas is relatively difficult as many of these species are not regularly trapped / reared in the context of monitoring programmes or sampling campaigns (Virgilio et al. 2011). In this respect, Natural History collections are considered as a valuable source of already referenced vouchers that can be used for DNA barcoding. However, producing DNA barcodes from Natural History collections can also be problematic if the DNA has become degraded during storage (Zimmermann et al. 2008). A screening based on approximately 400 tephritid vouchers from the RMCA collections (Virgilio and De Meyer, unpublished data) confirms that (a) as the age of the specimen increases, standard protocols for DNA extraction, amplification and Sanger sequencing become less and less efficient at producing DNA barcodes and (b) ethanol-preserved specimens tend to be more resistant to DNA degradation than pinned specimens. This screening revealed that, using standard protocols, DNA barcodes could be produced from less than 20 % of voucher specimens when those specimens were more than 10 years old (Fig. 2.1).

A survey of the collections of the RMCA revealed that 51 % of the 16,000 African tephritid vouchers were more than 15 years old (Fig. 2.2) suggesting that standard protocols for Sanger sequencing would be unlikely to produce DNA barcodes (Zimmermann et al. 2008).

An alternative approach for recovery of DNA from pinned museum specimens is the use of internal DNA primers and overlapping amplicons to reconstruct the full

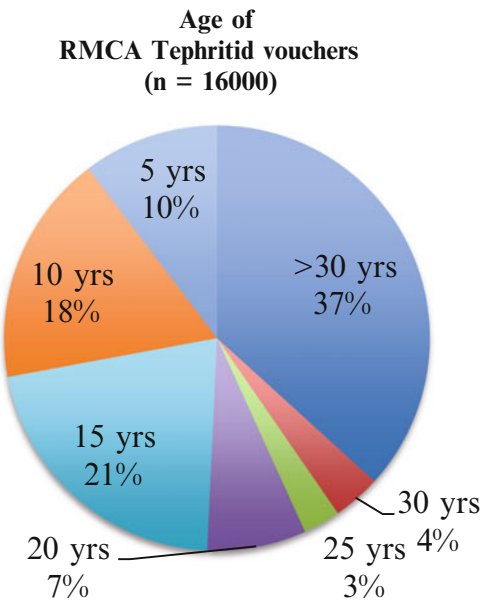

Fig. 2.2 Proportions of vouchers of different age classes in the RMCA collections

DNA barcode (Mitchell 2015). Van Houdt et al. (2010) and Smit et al. (2013) developed sets of internal primers specifically for tephritid fruit flies in the Natural History Collections. Van Houdt et al. (2010) used two overlapping amplicons successfully for reconstructing the full DNA barcode of specimens that were up to 15 years old and three overlapping amplicons for specimens that were up to 25 years old. However, this approach can be costly and time consuming so it is generally only used for rare collection material.

A more recent and cost-effective approach is the use of high throughput sequencing (next generation sequencing, NGS) that allows millions of DNA fragments from thousands of DNA templates to be sequenced in parallel. The NGS strategy for the mass production of DNA barcodes is promising (Meier et al. 2016) and allows DNA barcode amplicons to be individually tagged (using a set of oligonucleotides with a known sequence) so that multiple individuals can be processed in a single sequencing run and the individual DNA barcodes recovered through bioinformatics (Sucher et al. 2012; Shokralla et al. 2014; Shokralla et al. 2015).

References

Abuel-Ela RG, Hashem AG, Mohamed SMA (1998) *Bactrocera pallidus* (Perkin and May) (Diptera: Tephritidae), a new record in Egypt. J Egyptian German Soc Zool Entomol 27:221–229

Aluja M, Norrbom AL (1999) Fruit flies (Tephritidae) phylogeny and evolution of behavior. CRC Press, Boca Raton, pp 967

Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. Phil Trans Roy Soc London Ser B 360:1813–1823

Barr NB, Islam MS, De Meyer M, McPheron BA (2012) Molecular identification of *Ceratitis capitata* (Diptera: Tephritidae) using DNA sequences of the COI barcode region. Anns Entomol Soc Amer 105:339–350

Boykin LM, Armstrong K, Kubatko L, De Barro P (2012) DNA barcoding invasive insects: database roadblocks. Invertebr System 26:506–514

Cameron S, Rubinoff D, Will K (2006) Who will actually use DNA barcoding and what will it cost? System Biol 55:844–847

Collins RA, Cruickshank RH (2013) The seven deadly sins of DNA barcoding. Mol Ecol Res 13:969–975

de Carvalho MR, Bockmann FA, Amorim DS, Brandão CRF, de Vivo M, de Figueiredo JL, Britski HA, de Pinna MCC, Menezes NA, Marques FPL, Papavero N, Cancello EM, Crisci JV, McEachran JD, Schelly RC, Lundberg JG, Gill AC, Britz R, Wheeler QD, Stiassny MLJ, Parenti LR, Page LM, Wheeler WC, Faivovich J, Vari RP, Grande L, Humphries CJ, DeSalle R, Ebach MC, Nelson GJ (2007) Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. Evol Biol 34:140–143

De Meyer M (1996) Revision of the subgenus *Ceratitis* (*Pardalaspis*) Bezzi, 1918 (Diptera, Tephritidae, Ceratitini). Syst Entomol 21:15–26

De Meyer M (1998) Revision of the subgenus *Ceratitis* (*Ceratalaspis*) Hancock (Diptera: Tephritidae). Bull Entomol Res 88:257–290

De Meyer M (2000) Systematic revision of the subgenus *Ceratitis* Macleay *s.s.* (Diptera, Tephritidae). Zool J Linn Soc 128:439–467

De Meyer M (2006) Systematic revision of the fruit fly genus *Carpophthoromyia* Austen (Diptera, Tephritidae). Zootaxa 1235:1–48

De Meyer M (2009) Taxonomic revision of the fruit fly genus *Perilampsis* Bezzi (Diptera, Tephritidae). J Nat Hist 43:2425–2463

De Meyer M, Copeland R (2001) Taxonomic notes on the subgenera *Ceratitis* (*Hoplolophomyia*) and *Ceratitis* (*Acropteromma*) (Diptera, Tephritidae). Cimbebasia 17:77–84

De Meyer M, Freidberg A (2005) Revision of the fruit fly genus *Capparimyia* (Diptera, Tephritidae). Zool Scripta 34:279–303

De Meyer M, Freidberg A (2006) Revision of the subgenus *Ceratitis* (*Pterandrus*) Bezzi (Diptera: Tephritidae). Israel J Entomol 36:197–315

De Meyer M, Freidberg A (2012) Taxonomic revision of the fruit fly genus *Neoceratitis* Hendel (Diptera: Tephritidae). Zootaxa 3223:24–39

De Meyer M, White IM, Goodger KFM (2013) Notes on the frugivorous fruit fly (Diptera: Tephritidae) fauna of western Africa, with description of a new *Dacus* species. Europ J Taxon 50:1–17

Dowton M, Meiklejohn K, Cameron SL, Wallman J (2014) A preliminary framework for DNA barcoding, incorporating the multispecies coalescent. System Biol 63:639–644

Drew RAI, Tsuruta K, White IM (2005) A new species of pest fruit fly (Diptera : Tephritidae : Dacinae) from Sri Lanka and Africa. Afr Entomol 13:149–154

Ekesi S, Billah MK (2007) A field guide to the management of economically important tephritid fruit flies in Africa. ICIPE Science Press, Nairobi

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol & Biotechnol 3:294–299

Frey J, Guillén L, Frey B, Samietz J, Rull J, Aluja M (2013) Developing diagnostic SNP panels for the identification of true fruit flies (Diptera: Tephritidae) within the limits of COI-based species delimitation. BMC Evol Biol 13:1–19

Hancock DL, Drew RAI (2006) A revised classification of subgenera and species groups in Dacus Fabricius (Diptera: Tephritidae). Instrumenta Biodiversitatis VII:167–205

Hancock DL, White IM (1997) The identity of *Tririthrum nigrum* (Graham) and some new combinations in *Ceratitis* MacLeay (Diptera: Tephritidae). The Entomologist 116:192–197
Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Roy Soc B 270:313–321
Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2:e312
Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. Syst Biol 55:729–739
Jiang F, Jin Q, Liang L, Zhang AB, Li ZH (2014) Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: a case study in *Bactrocera* spp. Mol Ecol Res 14:1114–1128
Kvist S (2013) Barcoding in the dark?: a critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Mol Phylog Evol 69:39–45
Kwong S, Srivathsan A, Meier R (2012) An update on DNA barcoding: low species coverage and numerous unidentified sequences. Cladistics 28:639–644
Meeyen K, Nanork Sopaladawan P, Pramual P (2014) Population structure, population history and DNA barcoding of fruit fly *Bactrocera latifrons* (Hendel) (Diptera: Tephritidae). Entomol Sci 17:219–230
Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst Biol 55:715–728
Meier R, Wong W, Srivathsan A, Foo M (2016) $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. Cladistics 32:100–110
Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3:e422
Mitchell A (2015) Collecting in collections: a PCR strategy and primer set for DNA barcoding of decades-old dried museum specimens. Mol Ecol Res 15:1102–1111
Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. PLoS Biol 2:e35
Nagy ZT, Backeljau T, De Meyer M, Jordaens K (2013) DNA barcoding: a practical tool for fundamental and applied biodiversity research. In: ZooKeys p. 410
Nielsen R, Matz M (2006) Statistical approaches for DNA barcoding. Syst Biol 55:162–169
Pečnikar ŽF, Buzan E (2014) 20 years since the introduction of DNA barcoding: from theory to application. J Appl Gen 55:43–52
Porter TM, Gibson JF, Shokralla S, Baird DJ, Golding GB, Hajibabaei M (2014) Rapid and accurate taxonomic classification of insect (class Insecta) cytochrome c oxidase subunit 1 (COI) DNA barcode sequences using a naïve Bayesian classifier. Mol Ecol Res 14:929–942
Puillandre N, Lambert A, Brouillet S, Achaz G (2011) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol 2:1864–1877
Ratnasingham S, Hebert P (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 7:355–364
Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS One 8:e66213
Schutze MK, Aketarawong N, Amornsak W, Armstrong KF, Augustinos AA, Barr N, Bo W, Bourtzis K, Boykin LM, CÁCeres C, Cameron SL, Chapman TA, Chinvinijkul S, ChomiČ A, De Meyer M, Drosopoulou E, Englezou A, Ekesi S, Gariou-Papalexiou A, Geib SM, Hailstones D, Hasanuzzaman M, Haymer D, Hee AKW, Hendrichs J, Jessup A, Ji Q, Khamis FM, Krosch MN, Leblanc LUC, Mahmood K, Malacrida AR, Mavragani-Tsipidou P, Mwatawala M, Nishida R, Ono H, Reyes J, Rubinoff D, San Jose M, Shelly TE, Srikachar S, Tan KH, Thanaphum S, Haq I, Vijaysegaran S, Wee SL, Yesmin F, Zacharopoulou A, Clarke AR (2015) Synonymization of key pest species within the Bactrocera dorsalis species complex (Diptera: Tephritidae): taxonomic changes based on a review of 20 years of integrative morphological, molecular, cytogenetic, behavioural and chemoecological data. Syst Entomol 40:456–471

Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M (2014) Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Mol Ecol Res 14:892–901

Shokralla S, Porter TM, Gibson JF, Dobosz R, Janzen DH, Hallwachs W, Golding GB, Hajibabaei M (2015) Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. Sci Reps 5:9687

Smit J, Reijnen B, Stokvis F (2013) Half of the European fruit fly species barcoded (Diptera, Tephritidae); a feasibility test for molecular identification. ZooKeys 365:279–305

Sonet G, Jordaens K, Nagy ZT, Breman F, de Meyer M, Backeljau T, Virgilio M (2013) *Adhoc*: an R package to calculate ad hoc distance thresholds for DNA barcoding identification. ZooKeys 365:329–336

Sucher NJ, Hennell JR, Carles MC (2012) DNA fingerprinting, DNA barcoding, and next generation sequencing technology in plants. Methods Mol Biol 862:13–22

Tanabe AS, Toju H (2013) Two new computational methods for universal dna barcoding: a benchmark using barcode sequences of bacteria, archaea, animals, fungi, and land plants. PLoS One 8:e76910

Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. Mol Ecol Res 12:377–388

Van Houdt JKJ, Breman FC, Virgilio M, De Meyer M (2010) Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. Mol Ecol Res 10:459–465

Virgilio M, Backeljau T, Nevado B, De Meyer M (2010) Comparative performances of DNA barcoding across insect orders. BMC Bioinf 11:206

Virgilio M, Backeljau T, Emeleme R, Juakali JL, De Meyer M (2011) A quantitative comparison of frugivorous tephritids (Diptera: Tephritidae) in tropical forests and rural areas of the Democratic Republic of Congo. Bull Entomol Res 101:591–597

Virgilio M, Jordaens K, Breman FC, Backeljau T, De Meyer M (2012) Identifying insects with incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case. PLoS One 7:e31581

Virgilio M, White IM, De Meyer M (2014) A set of multi-entry identification keys to African frugivorous flies (Diptera, Tephritidae). ZooKeys 428:97–108

Virgilio M, Jordaens K, Verwimp C, White IM, De Meyer M (2015) Higher phylogeny of frugivorous flies (Diptera, Tephritidae, Dacini): localised partition conflicts and a novel generic classification. Mol Phylog Evol 85:171–179

White IM (2006) Taxonomy of the Dacina (Diptera:Tephritidae) of Africa and the Middle East. Afr Entomol Memoir 2:1–156

White IM, Goodger KFM (2009) African *Dacus* (Diptera: Tephritidae); new species and data, with particular reference to the Tel Aviv University collection. Zootaxa 2127:1–49

White I, Copeland R, Hancock D (2003) Revision of the afrotropical genus *Trirhithrum* (Diptera: Tephritidae). Cimbebasia 18:71–137

Zimmermann J, Hajibabaei M, Blackburn D, Hanken J, Cantin E, Posfai J, Evans T (2008) DNA damage in preserved specimens and tissue samples: a molecular assessment. Front in Zool 5:18

Massimiliano Virgilio is a molecular taxonomist and coordinator of the Joint Experimental Molecular Unit (JEMU, http://jemu.myspecies.info) of the Royal Museum for Central Africa (Tervuren, Belgium). His main interests are in morphological and molecular taxonomy, phylogeny and population genetics of African frugivorous tephritids. He is co-author of a number of publications on DNA barcoding and on the resolution of cryptic species complexes. In collaboration with Marc De Meyer and Ian White, he developed and regularly maintains a set of multi-entry keys for the morphological identification of African fruit flies (http://fruitflykeys.africamuseum.be/en/index.html).