Bulletin of Entomological Research (2006) **96**, 505–521 DOI: 10.1079/BER2006452

Molecular diagnostics of economically important *Ceratitis* fruit fly species (Diptera: Tephritidae) in Africa using PCR and RFLP analyses

N.B. Barr[1,6]*, R.S. Copeland[2,4], M. De Meyer[3], D. Masiga[4,5], H.G. Kibogo[4], M.K. Billah[4], E. Osir[4], R.A. Wharton[2] and B.A. McPheron[1]

[1]Department of Entomology, Pennsylvania State University, University Park, PA 16802, USA: [2]Department of Entomology, Texas A&M University, College Station, TX 77843, USA: [3]Royal Museum for Central Africa, Entomology Section, Belgium: [4]International Centre of Insect Physiology and Ecology, PO Box 30772, Nairobi, Kenya: [5]Department of Biochemistry and Biotechnology, Kenyatta University, Nairobi, Kenya: [6]Center for Plant Health Science and Technology, Pest Detection Diagnostics and Management Laboratory, USDA-APHIS, Moore Air Base, Edinburg, TX 78541, USA

Abstract

The predominantly Afrotropical fruit fly genus *Ceratitis* contains many species of agricultural importance. Consequently, quarantine of *Ceratitis* species is a major concern for governmental regulatory agencies. Although diagnostic keys exist for identification of all described *Ceratitis* species, these tools are based on adult characters. Flies intercepted at ports of entry are usually immatures, and *Ceratitis* species cannot be diagnosed based on larval morphology. To facilitate identification of *Ceratitis* pests at ports of entry, this study explores the utility of DNA-based diagnostic tools for a select group of *Ceratitis* species and related tephritids, some of which infest agriculturally important crops in Africa. The application of the polymerase chain reaction–restriction fragment length polymorphism (PCR–RFLP) method to analyse three mitochondrial genes (12S ribosomal RNA, 16S ribosomal RNA, and NADH-*dehydrogenase* subunit 6) is sufficient to diagnose 25 species and two species clusters. PCR analysis of the internal transcribed spacer region 1 (ITS-1) is able to distinguish three of the five species left unresolved by mitochondrial DNA analysis.

Keywords: mitochondrial DNA, ITS, *Ceratitis*, quarantine

Introduction

The family Tephritidae contains many species that are of agricultural importance. Tephritids ('true fruit flies') can infest and damage the fruit, seeds, and/or vegetative tissue of plant commodities (White & Elson-Harris, 1992). Many tephritids are pests in their endemic range, and several have become invasive species. For example, *Ceratitis capitata* (Wiedemann), the Mediterranean fruit fly, has been extremely successful at invading and establishing in new habitats. Its success is in part due to an extremely polyphagous diet (it attacks nearly 400 plant species worldwide from over 60 families; Liquido *et al*., 1998; DeMeyer *et al*., 2002), a liberal host acceptance behaviour (Carey, 1984; Yuval & Hendrichs,

*Fax: (956) 580 7300
E-mail: Norman.B.Barr@aphis.usda.gov

2000; DeMeyer, 2001a), rapid population growth (type r demographic strategy; Duyck *et al*., 2004) and a tolerance for a wide range of climates (Copeland *et al*., 2002). Although endemic to sub-Saharan Africa, *C. capitata* has successfully colonized tropical, sub-tropical, and mild temperate habitats around the world (Vera *et al*., 2002). Aided by human activity, this range expansion has occurred over the last 200 years.

The potential for future fruit fly invasions is of great concern to regulatory agencies and domestic growers. The interception of fruit shipments with evidence of infestation requires expert identification to the species level in order to determine its fate. This is not a simple task. For example, only recently have updated morphological keys of adult characters been developed for the genus *Ceratitis* MacLeay (Freidberg, 1991; De Meyer, 1996, 1998, 2000; De Meyer & Freidberg, in press), and diagnostic characters are not available for larvae. Ecological data are also difficult to utilize for accurate identification because the systematic relationships of pests and their relatives are often unresolved, the geographic ranges of many species are not documented, and the list of species that can potentially attack any given host plant is unknown. Recent work on *C. capitata* and its relatives has helped to rectify such omissions in their natural history (De Meyer, 2001a; Copeland *et al*., 2002; De Meyer & Copeland, 2005).

Several researchers have addressed this quarantine problem by developing molecular markers for tephritid species diagnosis (Douglas & Haymer, 2001; Kakouli-Durante *et al*., 2001; Baliraine *et al*., 2003; Naelole & Haymer, 2003; Ochando *et al*., 2003). Armstrong *et al*. (1997) developed a polymerase chain reaction–restriction fragment length polymorphism (PCR–RFLP) diagnostic tool to distinguish among species within the genera *Anastrepha* Schiner, *Bactrocera* Macquart and *Ceratitis* that are important to quarantine authorities in New Zealand. More recently, Salazar *et al*. (2002) developed a PCR–RFLP diagnostic for identification of four *Rhagoletis* Loew species in Chile and Muraji & Nakahara (2002) developed a PCR–RFLP protocol for diagnosing *Bactrocera* species. Although PCR–RFLP has been used extensively for population genetics of *C. capitata* (e.g. Gasparich *et al*., 1995, 1997; Silva *et al*., 2003), it has not been used for species level diagnostics of *Ceratitis* in its native Africa.

In this study we explore the utility of molecular data to diagnose *C. capitata* and 29 additional tephritids that, as larvae, infest fruits that grow in Africa. Of the species targeted for diagnosis, 29 are from the tribe Ceratitidini (26 *Ceratitis* spp., *Trirhithrum nigerrimum* (Bezzi), *Capparimyia melanaspis* (Bezzi), and *Carpophthoromyia dimidiata* (Bezzi) and one is from the tribe Carpomyini (a new species of *Notomma*). Most of these species are considered agricultural pests or are listed as quarantine species in parts of the world. Several of these species, like *C. capitata*, are highly polyphagous (e.g. *C. rosa* Karsch, *C. anonae* Graham, *C. cosyra* (Walker), and *C. fasciventris* (Bezzi)) and considered potentially invasive. Despite the prominence of *C. capitata* as a worldwide pest, *C. rosa* (the Natal fruit fly) appears to be more cold tolerant than *C. capitata* and capable of outcompeting and displacing adventive *C. capitata* populations (De Meyer, 2001b; Duyck & Quilici, 2002). Using PCR and RFLP methodologies, we show that three mitochondrial loci (12S ribosomal RNA gene, 16S ribosomal RNA gene, and the NADH-*dehydrogenase* subunit 6 gene (ND6 + $tRNA^{(Thr+Pro)}$)) and one nuclear locus (the ribosomal internal transcribed spacer region 1, ITS-1) are useful for discriminating amongst our target species. In addition, the results are used to construct paper-based and computer-based diagnostic tools and discuss the relative merits and limitations of these tools.

Materials and methods

Insect collections

This study is based on African samples from 29 ceratitidine species and one *Notomma* species (collection code 1000). Species names and sample sizes are reported in table 1. The new species 1195 is tentatively included as a *Ceratitis* species in this study based on molecular phylogenetic analyses by Barr & McPheron (2006) and Barr (2004). Collection information for specimens used for molecular analysis is reported in Barr (2004) and available from the corresponding author. The majority of collections were of flies reared from host fruit collected in Kenya (see Copeland *et al*., 2002) and stored in ethanol prior to genetic analysis; several of these Kenyan collections, however, were pinned prior to analysis. Other samples were from collections made in Ghana (*C. capitata*, $n = 6$; *C. ditissima* (Munro), $n = 13$; *C. anonae*, $n = 5$; *C. colae* Silvestri, $n = 26$), Mali (*C. fasciventris*, $n = 3$; *C. cosyra*, $n = 5$), Malawi (*C. capitata*, $n = 2$; *C. rosa*, $n = 5$), Nigeria (*C. ditissima*, $n = 3$), Reunion (*C. capitata*, $n = 2$; *C. rosa*, $n = 6$), and South Africa (*C. rosa*, $n = 4$). Vouchers for Kenyan collections were submitted to the Frost Museum at the Pennsylvania State University and the Royal Museum of Central Africa, Teruven, Belgium. Vouchers of other collections are stored in the McPheron laboratory at Pennsylvania State University and the Biosystematics Unit of the International Centre of Insect Physiology and Ecology (ICIPE).

DNA extraction

DNA was extracted from each individual using either the total nucleic acid protocol of Han & McPheron (1997) or the DNeasy animal tissue protocol (Qiagen Inc., Valencia, California). Extractions were performed on the head and thorax of flies using the Han & McPheron protocol; wings and abdomens were removed prior to extraction and stored in 95% ethanol. Heads, heads plus thoraces, or entire flies were used to extract DNA with the DNeasy method. For this method, tissue was incubated in tissue lysis buffer between 3 and 48 h and DNA was extracted following the standard protocol for animal tissues (without incubation with RNase). Samples were not homogenized, and the extracted body parts (or body) were stored in ethanol with the rest of the body of the fly. This non-destructive method was performed on both pinned and alcohol preserved samples.

Search strategy for RFLP-based markers

Restriction sites useful for species diagnosis were identified by screening DNA sequence data for endonuclease sites variable among representative species and performing pilot PCR–RFLP studies to measure digestion site variation within species. First, eight mitochondrial genes (12S rRNA, 16S rRNA, NADH-*dehydrogenase* subunit 1, NADH-*dehydrogenase* subunit 4, NADH-*dehydrogenase* subunit 4L, ND6, *cytochrome oxidase* I, *cytochrome oxidase* II) were amplified and sequenced from several species that were considered representative of the larger target list of 30 species (primers and reaction conditions are reported in Barr, 2004). Sequence data for the

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

Table 1. Fruit fly species, sample size, and digestion forms.

Species	*n*	Digestion forms [gene: endonuclease-form]
Ceratitis capitata (Wiedemann)	40	[12S: SspI-A1; MboII-B] [16S: ClaI-A; DraI-A1]
C. caetrata Munro	28	[12S: SspI-A1; MboII-B] [16S: ClaI-A; DraI-A1]
C. pinax Munro	7	[12S: SspI-A1; MboII-B] [16S: ClaI-B; DraI-A1] [ND6: DdeI-B]
C. ditissima (Munro)	32	[12S: SspI-A1; MboII-B] [16S: ClaI-B; DraI-B2] [ND6: NsiI-A]
C. millicentae De Meyer & Copeland	10	[12S: SspI-A1; AseI-A] [16S: ClaI-A; DraI-B2; PsiI-A] [ND6: HinfI-C]
C. anonae Graham	30	[12S: SspI-B; AseI-B] [16S: ClaI-A; DraI-A2; PsiI-C1]
C. fasciventris (Bezzi)	32	[12S: SspI-B; AseI-B; MboII-B; RsaI-A] [16S: ClaI-A/B; DraI-A2; PsiI-C1]
C. rosa Karsch	41	[12S: SspI-B; AseI-B; MboII-B; RsaI-A] [16S: ClaI-A/B; DraI-A2/B2; PsiI-C1]
C. rubivora Coquillett	23	[12S: SspI-B; SwaI-A] [16S: ClaI-B; DraI-B3; PsiI-B1]
Ceratitis sp. CO	12	[12S: SspI-B; AseI-B] [16S: ClaI-A; DraI-A2; PsiI-B1] [ND6: NsiI-B]
C. flexuosa (Walker)	25	[12S: SspI-B; AseI-A/B] [16S: ClaI-A; DraI-A2; PsiI-B1/D] [ND6: MnlI-A; NsiI-A]
C. podocarpi (Bezzi)	25	[12S: SspI-A1; AseI-A] [16S: ClaI-A; DraI-B2] [ND6: HinfI-B1/B2]
C. curvata (Munro)	9	[12S: SspI-A1; AseI-C] [16S: ClaI-A; DraI-B2]
C. querita (Munro)	16	[12S: SspI-B; AseI-A] [16S: ClaI-B; DraI-B3; PsiI-C2]
C. colae Silvestri	26	[12S: SspI-B; AseI-A] [16S: ClaI-A; DraI-A2] [ND6: MnlI-B]
C. perseus De Meyer & Copeland	22	[12S: SspI-A1; AseI-A; MboII-B] [16S: ClaI-B; DraI-B2] [ND6: NsiI-B]
C. oraria De Meyer & Copeland	15	[12S: SspI-A1; MboII-A] [16S: ClaI-B; DraI-A2/B3; MnlI-A] [ND6: DdeI-C]
C. divaricata (Munro)	7	[12S: SspI-A1; MboII-B] [16S: ClaI-B; DraI-A1] [ND6: DdeI-C]
C. cosyra (Walker)	26	[12S: SspI-B; AseI-A; SwaI-B] [16S: ClaI-A/B; DraI-B1/D; PsiI-B1/B2]
C. contramedia (Munro)	16	[12S: SspI-C]
C. stictica Bezzi	7	[12S: SspI-B; MboII-A] [16S: ClaI-B; DraI-A2; PsiI-C1]
C. argenteobrunnea Munro	6	[12S: SspI-A1; AseI-A; MboII-A; RsaI-B] [16S: ClaI-A/B; DraI-A1/B3/C; PsiI-B1; MnlI-A] [ND6: DdeI-A/B; HinfI-A/C]
C. marriotti Munro	27	[12S: SspI-B; AseI-B; RsaI-C] [16S: ClaI-B; DraI-A1; PsiI-B1/C1] [ND6: DdeI-C]
C. venusta (Munro)	27	[12S: SspI-B; AseI-B; MboII-B] [16S: ClaI-B; DraI-A1; PsiI-B1] [ND6: DdeI-A]
C. cristata Bezzi	14	[12S: SspI-A1; RsaI-A] [16S: ClaI-A; DraI-C] [ND6: HinfI-C]
Ceratitis sp. 1195	9	[12S: SspI-A1; AseI-C; MboII-B] [16S: ClaI-B; DraI-B3]
Trirhithrum nigerrimum (Bezzi)	26	[12S: SspI-A1; RsaI-A] [16S: ClaI-A; DraI-C] [ND6: HinfI-A]
Capparimyia melanaspis (Bezzi)	17	[12S: SspI-A1; MboII-A/C] [16S: ClaI-B; DraI-A1/B4; MnlI-B]
Carpophthoromyia dimidiata Bezzi	14	[12S: SspI-A2/D; MboII-A] [16S: ClaI-A; DraI-A2; PsiI-A]
Notomma sp. 1000	8	[12S: SspI-E]

entire mitochondrial genome of *C. capitata* is available from GenBank (Accession number NC_000857; Spanos *et al*., 2000). Restriction sites in the data were located using either OMIGA 2.0 or DS Gene software (Oxford Molecular, Madison, Wisconsin), and variable sites useful for species separation were determined manually. Utility of markers was evaluated by digestion of four to ten individuals from each of the 30 species.

PCR, RFLP and sequence analysis of samples

Based on preliminary results, a PCR–RFLP protocol to distinguish the thirty target species was developed using three genes (12S, 16S and ND6) and 12 restriction enzymes (*Ssp* I, *Ase* I, *Mbo* II, *Swa* I, *Rsa* I, *Dra* I, *Cla* I, *Psi* I, *Mnl* I, *Nsi* I, *Dde* I and *Hinf* I; New England Biolabs, Beverly, Massachusetts). The following primers were used to amplify the genes: 12S (SR-J-14199: 5′-TAC TAT GTT ACG ACT TAT and SR-N-14594: 5′-AAA CTA GGA TTA GAT ACC C); 16S (LR-J-12883 : 5′-CTC CGG TTT GAA CTC AGA TC and LR-N-13398: 5′-CGC CTG TTT ATC AAA AAC AT); and ND6 (TT-J-9886: 5′-TAA AAA CAT TGG TCT TGT AA and N6-N-10447: 5′-TAC TAC AGC AAT TAA TGT AA). The 12S primers were reported by Smith *et al*. (2002), the 16S primers by Han & McPheron (2000) and N6-N-10447 by Gasparich *et al*. (1995). Primers were synthesized at Pennsylvania State University's Nucleic Acid Facility (NAF). PCR was performed on Gene Amp PCR system 9700 thermal cyclers (PE Applied Biosystems, Foster City, California) in 30 or 50 µl reaction volumes using Qiagen *Taq* (Qiagen Inc., Valencia, California) and a final concentration of 0.25 µM of each primer. A standard PCR programme cycle was used for most reactions: 94°C/3 min; 39 cycles of 94°C/1 min, 55°C/1 min, 72°C/1 min; 72°C/10 min. The annealing temperature was 60°C for amplification of the 16S gene and 50°C for samples that produced either faint or no product under the stricter conditions. Digestions were performed in 20 µl reactions following the protocol of the manufacturer (New England Biolabs, Beverly, Massachusetts). Each reaction contained 3–6 µl of PCR product (*c*. 30–60 ng DNA) and 2 (*Cla* I, *Ssp* I, and *Mnl* I), 3 (*Mbo* II and *Swa* I), 4 (*Psi* I, *Ase* I, *Rsa* I, *Dde* I, *Hinf* I) or 5–8 (*Dra* I and *Nsi* I) Units of enzyme. Reaction mixtures were incubated between 3 and 16 h at the appropriate temperature. Restriction fragments were observed on a 1.4% agarose gel in TAE buffer containing ethidium bromide (0.4 µg ml^{-1}).

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

Table 2. Digestion forms for diagnostic tools.

Gene	Enzyme	Form	Fragments (bp)
12S	SspI	A1	189–190, 122–123, 122
		A2	188, 121, 93
		B	244–246, 189–190
		C	147, 145, 97
		D	310, 93
		E	290, 146
	AseI	A	434
		B	317–318, 116
		C	346, 88*
	MboII	A	434–436
		B	274–275, 159
		C	224, 211, (434)†
	SwaI	A	434
		B	317, 118
	RsaI	A	275–292, 141–142
		B	240, 142, 52*
		C	381, 52*
16S	DraI	A1	457–487
		A2	434–435, 70*
		B1	263, 228
		B2	228–241, 150–194, 53–78*
		B3	258–281, 154–200, 70*
		B4	311, 176
		C	200–211, 192–194, 60–85*
		D	524
	ClaI	A	556–561
		B	444–447, 112–114
	PsiI	A	559–561
		B1	389–390, 169–170
		B2	344, 169
		C1	271–291, 168–169, 100–120
		C2	290, 169, 143
		D	247, 169, 143
	MnlI	A	470–473, 88
		B	328, 121–142, 88*
ND6	NsiI	A	598–599
		B	456, 142
	DdeI	A	598–599
		B	411, 187
		C	331–332, 267
	HinfI	A	562–599
		B1	435, 87, 76*
		B2	435, 163
		C	334–340, 259–264
	MnlI	A	478, 120
		B	403, 120, 75*

† This fragment is not supported by sequence data.
* Bands between 50–90 bp can be difficult to view.

If a gene was required for the identification of a species, then it was sequenced from at least one individual of that species. If a species possessed more than one restriction pattern, then an individual representative of each alternate pattern was also sequenced. Sequencing reactions were performed on purified PCR products (QiaQuick purification kit; Qiagen Inc., Valencia, California) using the Big Dye cycle sequencing kit version 3.1 (Applied Biosystems, Foster City, California). The sequencing reaction products were run on an ABI PRISM® Hitachi 3100 Genetic Analyser (Applied Biosystems, Foster City, California) at the Nucleic Acid Facility of the Pennsylvania State University. Each PCR product was sequenced in both directions to corroborate the inferred sequence. All sequences were aligned and edited using the DS Gene programme (Oxford Molecular, Madison, Wisconsin) and submitted to GenBank (see Results).

Using the ITS1-F5 (5′-CAC GGT TGT TTC GCA AAA GTT) and ITS1-B9 (5′-TGC AGT TCA CAC GAT GAC GCA C) primers developed by Douglas & Haymer (2001), the ribosomal internal transcribed spacer region 1 (ITS-1) locus was amplified from five species: *C. capitata*, *C. caetrata* Munro, *C. rosa*, *C. anonae*, and *C. fasciventris*. PCR was performed as described for the mitochondrial genes with the following exceptions: 0.05 μM of each primer was used per reaction and the programme cycle was modified to 94°C/3 min; 10 cycles of 94°C/1 min, 60°C/1 min, 72°C/1 min; 10 cycles of 94°C/30 s, 58°C/30 s, 72°C/30 s; 10 cycles of 94°C/1 min, 57°C/1 min, 72°C/1 min; 72°C/10 min. Amplicon size differences were scored on 1.4% agarose gels stained with ethidium bromide. For comparison with published *Ceratitis* ITS-1 sequences, the PCR products from *C. caetrata*, *C. rosa*, and *C. anonae* were cloned with the TA cloning kit (Invitrogen, Carlsbad, California) and sequenced with universal primers and ITS-1 primers. For the *C. rosa* and *C. anonae* specimens the following internal primers were also used to sequence the insert: ITS1-R1a (5′-AGA AAR GAA ATA CAC GTT AAT GTG) and ITS1-F2a (5′-TGC ATA CAT TGT ATT TGA AAT GC). The sequences were submitted to GenBank (AY792168, DQ645953 and DQ645954, respectively).

Development of diagnostic tools

The different restriction patterns determined by sequence analysis and scored on gels were encoded as forms (denoted either by a letter code or an alpha-numeric code). These forms are the operational characters in the diagnostic method. Individuals that produced restriction patterns, after PCR–RFLP analysis, that were not distinguishable based on our gel electrophoresis protocol were treated as identical. As a result, patterns with minor restriction site differences were binned into a single form to facilitate future analysis via the PCR–RFLP protocol (tables 1 and 2). The mitochondrial DNA restriction forms identified in this study were then used as the data to construct a paper key and a computational diagnostic tool for species identification.

A computational knowledge database was developed for the mtDNA PCR–RFLP data set using the NetWeaver knowledge engineering tool (NetWeaver, PSU; Saunders *et al.*, 1993; Saunders & Miller, 1997). This knowledge database is comprised of 30 dependency networks (one for each species). Although NetWeaver can incorporate 'fuzzy logic' arguments, the dependency networks used herein are simple arguments of logic that describe the characteristics of a species. Networks can be compared to query information to generate a True–False decision regarding species membership. Each network tests a goal, such as 'Is this individual genetically similar to a genotype from species X?' The NetWeaver program allows multiple goals (in the database) to be tested simultaneously and provides immediate information through its window interface as to which goals have been eliminated from contention and what new information is required to identify the unknown sample. Although the program orders questions according to what will be most informative (given that the queried individual could belong to any of the possible species with equal likelihood) it is possible to enter data in any order (unlike most standard flow charts and keys). However, in both diagnostic tools the hypothesis being tested is that the

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

genotype observed in the unknown individual is the same as the genotype found in one or more of the species in the database (key). More information on the NetWeaver program can be found at http://rules-of-thumb.com.

Transfer of technology and confidence estimation of markers

To investigate the problems of technology transfer and insufficient sampling in our original data set, a study was conducted in Kenya to test the reliability of three of the 16S markers (*Cla* I, *Dra* I and *Psi* I) developed for species diagnosis. The *Cla* I marker was tested for three species: *C. anonae* ($n = 48$), *C. fasciventris* ($n = 72$) and *C. rosa* ($n = 72$); the *Dra* I marker was tested for ten species: *C. capitata* ($n = 70$), *C. cosyra* ($n = 113$), *C. fasciventris* ($n = 72$), *C. anonae* ($n = 48$), *C. rosa* ($n = 72$), *C. colae* ($n = 8$), *C.* new sp. CO ($n = 10$), *C. caetrata* ($n = 8$), *C. ditissima* ($n = 6$), and *T. nigerrimum* ($n = 20$); and the *Psi* I marker was tested for seven species: *C. capitata* ($n = 30$), *C. fasciventris* ($n = 20$), *C. anonae* ($n = 20$), *C. rosa* ($n = 20$), *C. colae* ($n = 10$), *C.* new sp. CO ($n = 10$), and *C. caetrata* ($n = 8$). Specimens were extracted using a phenol chloroform method (Sheppard *et al*., 1992; Han & McPheron, 1997) and the digestions were run on 2.5–3.0% low-melt agarose gels (Flowgen Bioscience Limited, UK). Kibogo (2005) provides a more detailed description of reaction conditions.

The diagnostic tools developed in this study assume that the 30 species have been sufficiently sampled to document all restriction forms (haplotypes) present in each species (or at least that samples are large enough to be statistically confident of capturing this variation). It is possible that additional forms exist within species that also exist in others; such a situation of undocumented shared forms can result in the misidentification of an unknown sample. To quantify confidence levels for data sets, the probability that a form (haplotype) was not sampled given that it does exist was calculated. Using a binomial distribution set at arbitrary frequencies for rare alleles (e.g. 0.1), the probability of not sampling an allele was calculated for sample sizes relevant to our data sets (e.g. $n = 40$ and $n = 139$).

Results

The initial screening stage of this study indicated that three mitochondrial genes (12S rRNA, 16S rRNA and ND6) were more diagnostically-informative than the other five mitochondrial genes that were tested. (These three genes displayed a lower level of intraspecific variation than did the other genes based on preliminary analyses.) Sequences generated from all eight genes were submitted to GenBank: Accession numbers AY788408–AY788456 for COI, AY805308–AY805322 for COII, AY792037–AY792069 for 12S, AY792070–AY792110 for 16S, AY792169–AY792177 for ND1, AY792178–AY792187 for ND4, AY792188–AY792220 for ND4L and AY792132–AY792158 for ND6. PCR–RFLP diagnostic methods were developed and tested using the 12S rRNA, 16S rRNA and ND6 genes. A combination of 12 restriction enzymes with the three genes resulted in 13 markers (see table 2) that are scored by digesting an aliquot of PCR product with a specific enzyme. The efficiency of PCR amplification and restriction digestion was 100% for the 593 individuals included in the marker development phase of the study. PCR produced amplicons (amplified products) of *c*. 435, 560 and 600 bp for 12S, 16S and ND6, respectively. The predicted fragments sizes (based on sequence data) for species and individuals within a species (intraspecific polymorphism) are reported in Barr (2004) and available at the website http://rules-of-thumb.com or from the corresponding author. In general, unexpected artefacts generated during PCR and/or restriction digestion reactions were rare (< 2% of reactions). In most cases these artefacts were not present after PCR and/or restriction digestions were repeated. However, two individuals consistently produced restriction patterns atypical of DNA sequence data. A single *C. colae* sample produced a faint but unexpected 300 bp fragment when the ND6 gene was digested with *Mnl* I. The PCR and digestion was performed twice with the same result. The expected restriction fragments do not add to 300 bp so it is not likely to have been the result of an incomplete digestion. Analysis of ND6 in *C. rosa* has revealed the existence of ND6 pseudogenes in that species (Barr, 2004; Barr & McPheron, 2006). It is possible that a second copy of ND6 is present in *C. colae* that has an alternate recognition site for the *Mnl* I enzyme. Regardless, this *C. colae* marker is only used to distinguish it from *C. flexuosa* (Walker) and the 300 bp fragment in *C. colae* cannot be confused with the alternate ND6-*Mnl* I form present in *C. flexuosa*. Consequently, this marker is included in the diagnosis.

Similarly, a single *Capparimyia melanaspis* individual that has a recognition site for the *Mbo* II enzyme in the 12S gene (based on sequence data) produces both the digested and undigested (434 bp) DNA fragments after restriction digestion. All other individuals in this species do not have a recognition site for *Mbo* II and produce only a 434 bp fragment after restriction digestion (see fig. 1). This artefact may be the result of incomplete digest but increasing enzyme (three-fold) and incubation time (48 h) did not affect the digestion. This unexpected 434 bp band does not interfere with diagnosis of *C. melanaspis* because the band is characteristic of the species.

Based on resolution of fragments run alongside a 50 bp ladder on a 1.4% agarose gel, the RFLP banding patterns were grouped (or binned) into forms demarcated by letter and number codes (table 2 and figs 1–3). Bands below 50 bp were not seen under the gel conditions described here and are excluded from the form descriptions. In addition, because smaller bands are more difficult to see, even bands between 50 and 90 bp may not be visible if insufficient DNA is digested. A potential technical problem with the PCR–RFLP protocol is that several forms are defined by an absence of restriction sites (e.g. *Nsi* I of ND6). To reduce the probability of misdiagnosis because an enzyme fails to digest the DNA, a control containing the restriction site in question should be included in the analysis of such markers.

Most forms are distinguishable when run alongside a 100 bp ladder. Forms that require resolution at the 50 bp level to differentiate them or could be mistakenly scored under poor gel conditions are coded with the same letter but have a different number (e.g. 16S *Dra* I forms B2 and B3). Therefore, the data are binned in two ways: (i) according to a finer resolution (e.g. distinction between 16S *Dra* I forms B2 and B3 requires a 50 bp stepped ladder to score) called 'exact' forms; and (ii) according to a less fine resolution (e.g. distinction between 16S *Dra* I forms A and B requires a 100 bp stepped ladder to score) called 'general' forms. For diagnostic purposes the 'exact' forms can be entered as the 'general' letter (e.g. enter B instead of B1) when the more 'exact' form is uncertain. However, the use of a 'general'

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

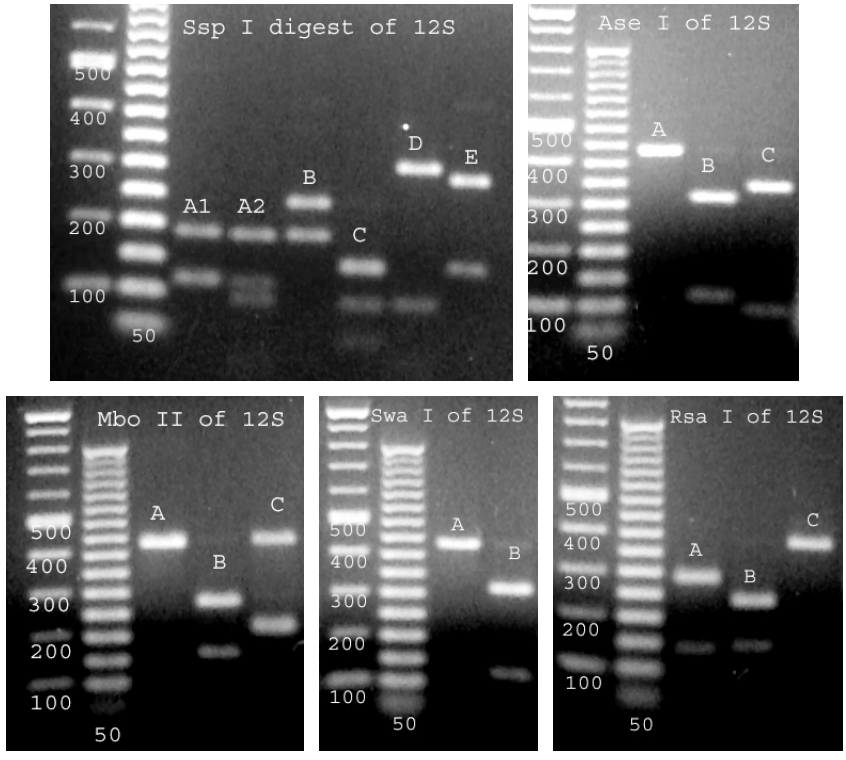


Fig. 1. Representative gels for 12S gene digestion profiles. Specimens used to represent forms:
SspI: A1 (*Ceratitis capitata*, NC_000857); A2 (*Carpophthoromyia dimidiata*, AY792045); B (*Ceratitis fasciventris*, AY792057); C (*Ceratitis contramedia*, AY792042); D (*C. dimidiata*, AY792046); E (*Notomma* sp. 1000, AY792062).
AseI: A (*Ceratitis argenteobrunnea*, AY792038); B (*C. fasciventris*, AY792057); C (*Ceratitis* sp. 1195, AY792061).
MboII: A (*C. dimidiata*, AY792046); B (*Ceratitis* sp. 1195, AY792061); C (*Capparimyia melanaspis*, AY792048).
SwaI: A (*Ceratitis rubivora*, AY792066); B (*Ceratitis cosyra*, AY792053).
RsaI: A (*C. fasciventris*, AY792057); B (*C. argenteobrunnea*, AY792038); C (*Ceratitis marriotti*, AY792060).

form when a more specific form exists can result in additional steps (restriction sites) to be tested.

Amplification of 16S rRNA was successful for 97.6% of the specimens tested in Kenya (i.e. 457/468). Of the *c.* 730 endonuclease digestions performed in Kenya only one species, *C. rosa*, produced forms not included in the original PCR–RFLP database. One *C. rosa* individual produced an unexpected 16S-*Dra* I form, and two *C. rosa* individuals produced an unexpected 16S-*Cla* I form. (The unexpected *Cla* I form is identical to a rare form documented in the original dataset for *C. fasciventris*.) Both of these forms were sequenced from *C. rosa* (accession numbers AY792097 for the new *Cla*I form and DQ015704 for the new *Dra*I form) and included in the diagnostic tools presented in this study. The probability of not including a rare haplotype in our sample was calculated for several sample sizes. Based on our binomial model a sample size of 40 flies is sufficiently large (based on $P = 0.05$) to detect infrequent haplotypes (e.g. 10–25%) but not to detect very rare haplotypes (e.g. $\leq 1\%$). The marker with the largest sample size in our study (i.e. *Ceratitis cosyra*, $n = 139$: 16S-*Dra*I) was sufficiently large to detect a haplotype with a frequency $\geq 2.1\%$.

Each species was tested for the minimum number of genes and restriction sites required to differentiate it from all other species. For example, the *Notomma* species only requires digestion of the 12S gene with *Ssp* I to distinguish it from the other 29 species. As a consequence, the full spectrum of restriction forms across all genes and restriction enzymes is not presented. This method of utilizing the minimum number of steps to obtain a diagnosis, however, is successful at distinguishing most of the 30 species. This method is incorporated in the diagnostics tools developed for these species: a paper key that requires 'exact' digestion forms (as reported in table 2) is reported in appendix 1, a paper key that allows 'general' forms to be used is reported in appendix 2, and a computational tool developed with NetWeaver is available either from the website http://rules-of-thumb.com or by request from the corresponding author. Note that the flow charts in the appendices are decision tools and not diagrams of the molecular procedure (see the aforementioned website for molecular protocols).

The only species that failed to separate based on PCR–RFLP analysis of mitochondrial DNA are the *C. capitata*–*C. caetrata* cluster and the *C. anonae*–*C. fasciventris*–*C. rosa* cluster. Under some circumstances, both *C. rosa* and *C. fasciventris* can be distinguished from *C. anonae* because they each possess an intraspecific polymorphism that is not shared with *C. anonae*. However, this is not a frequent haplotype and, in most cases, individuals from the three species are indistinguishable.

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

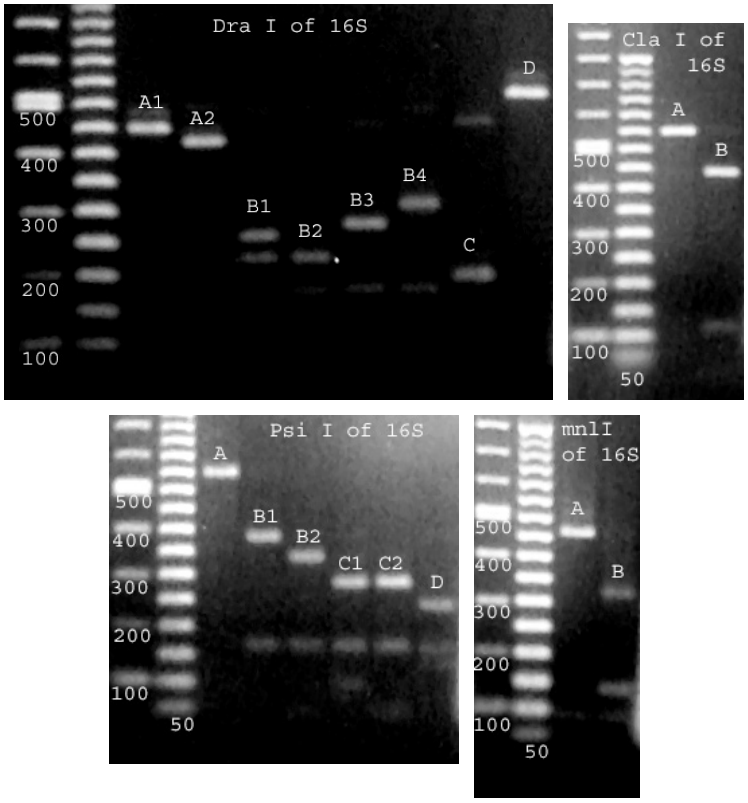


Fig. 2. Representative gels for 16S gene digestion profiles. Specimens used to represent forms:
DraI: A1 (*Ceratitis capitata*, NC_000857); A2 (*Ceratitis oraria*, AY792103); B1 (*Ceratitis cosyra*, AY792078); B2 (*Ceratitis millicentae*, AY792091); B3 (*Ceratitis querita*, AY792095); B4 (*Capparimyia melanaspis*, AY792102); C (*Ceratitis cristata*, AY792081); D (*C. cosyra*, AY792080).
ClaI: A (*C. capitata*, NC_000857); B (*C. oraria*, AY792103).
PsiI: A (*C. millicentae*, AY792091); B1 (*C. cosyra*, AY792078); B2 (*C. cosyra*: AY792080); C1 (*Ceratitis anonae*, AY792070); C2 (*C. querita*, AY792095); D (*Ceratitis flexuosa*, AY792106).
MnlI: A (*C. oraria*, AY792103); B (*C. melanaspis*, AY792102).

Ceratitis capitata and *C. caetrata* have length and sequence differences at the ITS-1 locus. The ITS-1 fragment amplified from *C. caetrata* ($n = 28$) was consistently *c.* 30 bp larger than those amplified for *C. capitata* ($n = 40$). Sequence data of the *C. caetrata* ITS-1 locus (fig. 4) supports this size difference when compared to the *C. capitata* sequences reported by Douglas & Haymer (2001; AF189691 and AF307848, *c.* 1 kb amplicon length). Although useful, this difference is small given the large size of the *C. caetrata* amplicon containing the ITS-1 locus (1028 bp). Therefore, a control of *C. capitata* should be run alongside an unknown individual for comparison (see fig. 4).

The ITS-1 locus is also useful for the separation of the *C. anonae*–*C. fasciventris*–*C. rosa* cluster. The ITS-1 marker was originally developed by Douglas & Haymer (2001) to discern between *C. rosa* and *C. capitata*. Their study, however, included *C. capitata*, *C. rosa* (from South Africa) and *C. fasciventris* (from Kenya – see De Meyer, 2001b as this species was considered a variation of *C. rosa* at the time of the Douglas & Haymer study). They reported size and sequence differences among these three species. The present analysis also found size differences among these species. The amplicon size from *C. fasciventris* ($n = 29$, the specimens from Mali were not included) was similar to that predicted by Douglas & Haymer (2001; AF189689, *c.* 890 bp). *Ceratitis anonae* ($n = 30$; DQ645954) produced an amplicon similar in size to *C. fasciventris*. *Ceratitis rosa*, however, is polymorphic for ITS-1 ($n = 41$; see fig. 4). Amplification of ITS-1 from 14 of the tested *C. rosa* individuals generates an amplicon like that described by Douglas & Haymer (AF189690, *c.* 1100 bp). These flies were collected from Reunion, South Africa, Malawi and two collection sites in Kenya (codes 1896 and 2135). Most of the *C. rosa* individuals in our analysis, however, produced a much larger (*c.* 1400 bp; DQ645953) amplicon. These flies were from Kenyan collections (excluding codes 1896 and 2135; see Barr (2004)) and a single individual from Malawi.

Discussion

This study demonstrates the utility of PCR–RFLP analysis of mitochondrial genes for the identification of immature

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

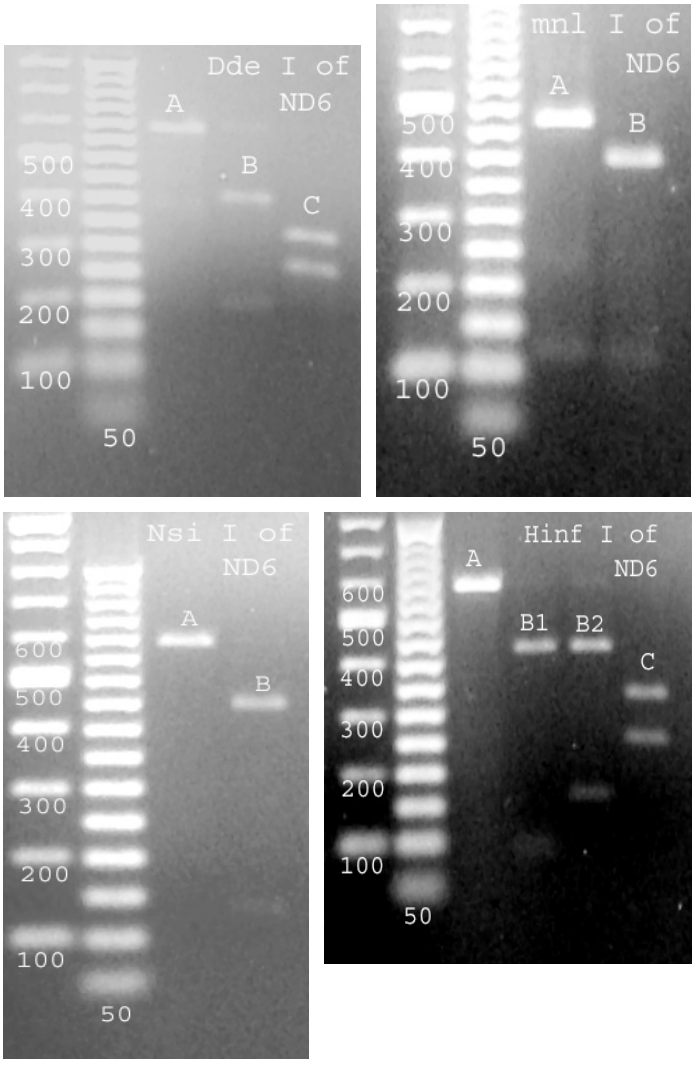


Fig. 3. Representative gels for ND6 + tRNA gene digestion profiles. Specimens used to represent forms:
DdeI: A (*Ceratitis venusta*, AY790597); B (*C. pinax*, AY790573); C (*C. marriotti*, AY790556).
MnlI: A (*C. flexuosa*, AY790574); B (*C. colae*, AY790607).
NsiI: A (*C. flexuosa*, AY790574); B (C. sp. CO, AY790606).
HinfI: A (*Trirhithrum nigerrimum*, AY790610); B1 (*C. podocarpi*, AY790551); B2 (*C. podocarpi*, AY790552); C (*C. millicentae*, AY790593).

stages of 25 species and two species clusters of fruit flies. In addition, PCR analysis of the ITS-1 locus is useful for species diagnosis (e.g. *C. rosa*) within the genus. The molecular diagnostic tools reported here possess several properties that make them useful for identification of intercepted *Ceratitis* fruit flies. Firstly, PCR and RFLP methods are robust and reproducible for the markers described here (i.e. there is high probability of success for PCR amplification and RFLP digestion) thereby reducing the risk of negative results because of technical difficulties. Secondly, interpretation of the molecular characters reported here does not require expert knowledge of taxonomic terminology or familiarity with the taxa being diagnosed. In addition, the binning of restriction patterns (DNA fragments) into categories should reduce error introduced during scoring of gels. Lastly, the molecular protocol can be completed in 24–48 h. This is an improvement because a longer time is required to rear immatures to the adult stage (assuming that the larvae even survive the rearing process).

A protocol for using our PCR–RFLP diagnostic tool is provided on the website http://rules-of-thumb.com. Ideally this methodology should be integrated with the current set of tools available for diagnosis and is not intended to supplant established methods of diagnosis. Like morphological keys, the diagnostic tools that result from this study are structured so that the most informative markers are required (or requested) first; this logical progression of inquiry facilitates the identification of the target species.

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

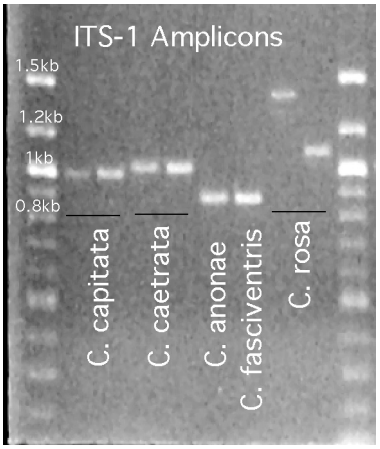


Fig. 4. Representative amplicons of the ITS-1 locus.

The mitochondrial DNA (mtDNA) PCR–RFLP method presented here can be supplemented with additional genetic markers. For example, the ITS-1 locus permits diagnosis of *C. rosa* from its close relatives (*C. fasciventris* and *C. anonae*) and *C. capitata* from *C. caetrata*. Based on our results, the applicability of the ITS-1 locus should be explored further for tephritid molecular diagnostics. Although specimens of *C. anonae* and *C. fasciventris* shared identical ITS-1 amplicons, an RFLP approach has not yet been explored for these species.

Our observation of multiple ITS-1 forms in the species *C. rosa* is intriguing from a taxonomic perspective. The prevalence of the shorter (*c*. 1.1 kb) and longer (*c*. 1.4 kb) ITS-1 forms is associated with geography: the shorter form is fixed in South Africa, Reunion and two of the eight Kenyan collections; likewise, the longer form is fixed in six of the eight Kenyan collections. The only collection containing both forms is from Malawi. Sequence analysis of the ND6 gene (Barr, 2004) indicates that *C. rosa* has a high genetic diversity in Malawi when compared to Kenya and South Africa, and, therefore, may represent a bridge between (or sink for) Kenyan and South African forms. We are currently investigating the population genetic structure of this species throughout Africa.

The development of molecular diagnostics for tephritid species is an important area of research because many are agricultural pests and are morphologically difficult to identify, especially in the immature stages (White & Elson-Harris, 1992; Armstrong *et al*., 1997; Salazar *et al*., 2002). In terms of economic importance the present study included many fruit flies that are pests or have the potential to become invasive species. Most notable is *C. capitata* which is regarded as one of the most destructive international fruit pests. Although the 30 target species included in this study are Afrotropical species, international trade with Africa is expected to increase in the future and, with it, the risk of introductions. Morphological keys based on larval characters are not available for most tephritids (White & Elson-Harris, 1992) and can be technically difficult to use (Steck *et al*., 1990). Thus far, studies of *Ceratitis* larvae have not resulted in characters useful for species diagnosis. Even when larval data are available, they are generally limited to the third instar (White & Elson-Harris, 1992). This is problematic because it requires larvae to be reared to the third instar or to the adult stage in order to identify which species is infesting a fruit commodity that has been intercepted at a port. During this time the commodity can lose value regardless of whether the maggot is of quarantine importance. The methods described here allow an unknown sample to be analysed rapidly, regardless of its developmental stage.

Although the paper-based keys (appendices 1 and 2) developed from this study are sufficient for diagnostic analysis, the computational tool offers several advantages over a paper-based tool. Firstly, the computational knowledge base allows a user to reduce the number of goals (i.e. species) to be tested. A paper key will always test the query against a fixed number of goals (species). This may not be necessary (or time efficient) if the number of possible goals is limited *a priori* by geographic or host data. Before or during an interrogation, using the NetWeaver-based tool, a user can reduce the number of goals and search those that are of interest. Secondly, the computational tool allows the user to determine the order in which data are entered. Although the program, by default, orders questions according to what will be most informative, there are situations (such as when data are missing) when the user will prefer to determine the order. This is not possible with a paper key. Thirdly, both 'exact' and 'general' digestion forms can be entered into the computational tool. The paper keys are designed to allow the user to enter either 'exact' (appendix 1) or 'general' (appendix 2) forms, but in some cases the digestion profile for a query will consist of both types. The computational tool can accommodate such data. Lastly, the computational tool can be easily modified (updated) when additional data are available.

The diagnostic tools presented here rely heavily on samples from Kenya and may only be applicable to individuals collected from this geographical area. To apply this tool to infested fruit from outside the sampled range, one assumes that the variation in Kenya (and other sites included in the study) is indicative of all populations of the species. Biogeographic, allozyme, and microsatellite studies have indicated that the subgenus *Ceratitis* s.s. originated in or near Kenya (Malacrida *et al*., 1998; Baliraine *et al*., 2004; De Meyer *et al*., 2004). The genetic diversity of a population (or species) should be higher for an older population than a younger population. Therefore, if an east African origin is true, there is a greater probability that most of the intraspecific variation present in the species of *Ceratitis* s.s. (i.e. *C. capitata*, *C. caetrata*, and *C. pinax* Munro) was captured. Evidence from a PCR–RFLP study, however, indicates that *C. capitata* is most variable in western Africa (Gasparich *et al*., 1997). A more extensive survey in the region should shed light on the true origin of *Ceratitis* s.s. and other subgenera.

Related to the problem of geographic sampling is the issue of insufficient sample size. By using a binomial distribution set at an arbitrary frequency (e.g. 0.1), we calculated the probability that we failed to sample an additional, rare haplotype assuming that it exists in the population. For example, when applied to 12S-*Ssp* I for *C. capitata* ($n = 40$) the probability is only < 0.05 when the hypothetical frequency for the haplotype is > 0.07. To ensure that rare alleles (frequency of 0.01) are not missed in an analysis (with a probability of 0.05), *c*. 300 individuals must be analysed per species. It should be noted that this binomial model is based on biologically unrealistic assumptions (e.g. that species are each genetically uniform over geographic space) and it does not take sampling error into account. Our results indicate that rarity (i.e. small sample size) does not equate with

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

smaller population variation. For example, *C. argenteobrunnea* Munro ($n = 6$) has multiple forms for the 16S markers, whereas *C. capitata* ($n \geq 40$) and its close relative *C. caetrata* ($n = 28$) are ostensibly fixed for single forms.

Another assumption made in the diagnostic tool is that only individuals from the 30 target species will be tested. For example, *C. punctata* (Wiedemann), an economically important polyphagous species that could be confused with species in our study based on geography and host plants, was excluded because samples were unavailable. Although it is possible for another species (e.g. *C. punctata*) to share a digestion profile with a species already in the diagnostic tool, this does not affect the tool's ability to discern amongst species in the key. For example, if a user wants to know if an individual is not *C. capitata* and the digestion profile is unlike *C. capitata*, then the key would remove *C. capitata* from contention and provide the answer sought. This is the result regardless of whether *C. punctata* and *C. capitata* have identical profiles.

It should be noted that the mitochondrial markers developed in this study are capable of amplifying mtDNA genes from across the entire family Tephritidae (and perhaps other families). Indeed the genus *Notomma* (subfamily Trypetinae) was included in this study and our three genes were successfully amplified from several *Bactrocera* (subfamily Dacinae), *Dacus* (subfamily Dacinae), *Anastrepha* (subfamily Trypetinae), and *Rhagoletis* (subfamily Trypetinae) species (unpublished). Therefore it is possible for all tephritid fruit flies to be analysed with this method. Unfortunately, many of the diagnostic markers presented here could be homoplasic and not divide species according to their phylogenetic relationships. For example, preliminary analysis of species from two other economically important genera, *Dacus* Fabricius (i.e. *D. demmerezi* (Bezzi), *D. vertebratus* Bezzi, and *D. bivittatus* (Bigot)) and *Bactrocera* (i.e. *B. cucurbitae* (Coquillett), *B. dorsalis* (Hendel), and *B. amplexa* (Munro)), failed to find a marker that could unambiguously separate *Ceratitis* from these genera (data not shown). This does not preclude the ability of these markers to diagnose fruit flies outside of *Ceratitis*. Rather than having a single character to distinguish all *Bactrocera* from all *Ceratitis*, it is possible to have a combination of several characters (i.e. genetic profile or fingerprint) that can diagnose all species in these genera; each profile would be consistent for each species but not follow evolutionary classification. This approach to tephritid identification could be taken by continually adding species (and markers) to the current data matrix. Until such a matrix is available, it is important that quarantine specialists use additional characters to identify species from non-*Ceratitis* genera prior to using our molecular diagnostic tool. White & Elson-Harris (1992) provide a key to identify economically important genera based on third instar larval morphology. This key is used by port identifiers in the USA to distinguish amongst *Ceratitis*, *Trirhithrum* Bezzi, *Dacus* and *Bactrocera* (G. Steck, personal communication). In addition, Armstrong *et al*. (1997) included 13 *Bactrocera* species, four *Anastrepha* species, and *Ceratitis capitata* in a molecular diagnostic tool for New Zealand. This and similar tools (e.g. Muraji & Nakahara, 2002) can help determine if the genetic profile of an intercepted fly is at least consistent with the genetic profile of an economically important species outside of *Ceratitis*. Alternatively, DNA sequencing could be applied to an unknown fly to search a DNA database (e.g. GenBank) for its most similar species or group (see Armstrong & Ball, 2005). This, however, is more expensive and time limiting than other diagnostic methods.

Despite several limitations inherent to diagnostic tool development, this study is a major contribution to *Ceratitis* pest identification and control efforts. The method is fast (compared to the alternative of rearing larvae), inexpensive (requiring standard equipment of molecular biology) and technically easy to perform. The computational diagnostic tool is easily modified and can accommodate additional forms, species, and markers. The problem of insect molecular diagnostics is one that requires extensive sampling and ongoing revision. This study is ambitious in terms of the number of species included and was intended to establish a framework onto which additional studies should build; consequently, future studies should focus on increasing collection locations and sample sizes of the species included in this study. In addition, the diagnostic tool can be improved by analysing all 13 markers for the 30 species. The original diagnostic is based on the minimum number of markers required to distinguish species. As a result, species are analysed for a subset of the 13 markers. For example, the *Notomma* species is diagnosed based on a single marker (*Ssp* I of 12S). Adding information for the other markers will improve the tool by making the identification process more conservative because all of the markers must be consistent between species and query profiles to result in a match.

Acknowledgements

The authors thank Andrea Forbes, Katrina Shoemaker and Nicole Shakir for helping with nucleic acid extractions and other laboratory work. Michael Saunders and Bruce Miller provided excellent support during the development of the *Ceratitis* NetWeaver database. The authors also appreciate the assistance of Raul Ruiz, Eric Deutsch, and Dietmar Schwartz regarding the testing of the ITS-1 marker and reviewing an earlier version of the manuscript. An anonymous reviewer provided useful comments on an earlier version of this paper. Millicent Okumu and Peris Machera provided excellent technical help in overseeing the rearing of tephritids from fruits collected in Kenya. The Director General, ICIPE, is thanked for providing office and laboratory space and a field vehicle for fruit collection and insect rearing activities. This work was supported by USAID grant no. PCE-G-00-98-0048-00, USDA/CSREES/IFAFS grant no. 00-52103-9651, and the Pennsylvania Agricultural Experiment Station.

References

Armstrong, K.F. & Ball, S.L. (2005) DNA barcodes for biosecurity: invasive species identification. *Philosophical Transactions of the Royal Society B* **360**, 1813–1823.

Armstrong, K.F., Cameron, C.M. & Frampton, E.R. (1997) Fruit fly (Diptera: Tephritidae) species identification: a rapid molecular diagnostic technique for quarantine application. *Bulletin of Entomological Research* **87**, 111–118.

Baliraine, F.N., Bonizzoni, M., Osir, E.O., Lux, S.A., Mulaa, F.J., Zheng, L., Gomulski, L.M., Gasperi, G. & Malacrida, A.R. (2003) Comparitive analysis of microsatellite loci in four fruit fly species of the genus *Ceratitis* (Diptera: Tephritidae). *Bulletin of Entomological Research* **93**, 1–10.

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

Baliraine, F.N., Bonizzoni, M., Guglielmino, C.R, Osir, E.O., Lux, S.A., Mulaa, F.J., Gomulski, L.M., Zheng, L., Quilici, S., Gasperi, G. & Malacrida, A.R. (2004) Population genetics of the potentially invasive African fruit fly species, *Ceratitis rosa* and *Ceratitis fasciventris* (Diptera: Tephritidae). *Molecular Ecology* **13**, 683–695.

Barr, N.B. (2004) Molecular diagnostics and systematics of the genus *Ceratitis* (Tephritidae). PhD thesis, The Pennsylvania State University, Pennsylvania.

Barr, N.B. & McPheron, B.A. (2006) Molecular phylogenetics of the genus *Ceratitis* (Diptera: Tephritidae). *Molecular Phylogenetics and Evolution*. **38**, 216–230.

Carey, J.R. (1984) Host-specific demographic studies of the Mediterranean fruit fly, *Ceratitis capitata*. *Ecological Entomology* **9**, 261–270.

Copeland, R.S., Wharton, R.A., Luke, Q. & De Meyer, M. (2002) Indigenous hosts of *Ceratitis capitata* (Diptera: Tephritidae) in Kenya. *Annals of the Entomological Society of America* **95**, 672–694.

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

De Meyer, M. (1998) Revision of the subgenus *Ceratitis* (*Ceratalaspis*) Hancock (Diptera: Tephritidae). *Bulletin of Entomological Research* **88**, 257–290.

De Meyer, M. (2000) Systematic revision of the subgenus *Ceratitis* MacLeay *s. s.* (Diptera: Tephritidae). *Zoological Journal of the Linnean Society* **128**, 439–467.

De Meyer, M. (2001a) Distribution patterns and host-plant relationships within the genus *Ceratitis* MacLeay (Diptera: Tephritidae) in Africa. *Cimbebasia* **17**, 219–228.

De Meyer, M. (2001b) On the identity of the Natal fruit fly *Ceratitis rosa* Karsch (Diptera, Tephritidae). *Entomologie* **71**, 55–62.

De Meyer, M. & Copeland, R.S. (2005) Description of new *Ceratitis* MacLeay (Diptera: Tephritidae) species from Africa. *Journal of Natural History* **39**, 1283–1297.

De Meyer, M. & Freidberg, A. (in press) Revision of the subgenus *Ceratitis* (*Pterandrus*) Bezzi (Diptera: Tephritidae). *Proceedings of the 1st Tephridologists Meeting*, Israel.

De Meyer, M., Copeland, R.S., Lux, S.A., Mansell, M., Quilici, S., Wharton, R., White, I.M. & Zenz, N.J. (2002) Annotated check list of host plants for afrotropical fruit flies (Diptera: Tephritidae) of the genus *Ceratitis*. *Documentations Zoologiques, Musee Royal de l'Afrique Centrale* **27**, 1–91.

De Meyer, M., Copeland, R.S., Wharton, R.A. & McPheron, B.A. (2004) On the geographical origin of the medfly, *Ceratitis capitata* (Wiedemann) (Diptera: Tephritidae). *Proceedings of the 6th International Symposium on Fruit Flies of Economic Importance*, Stellenbosch, South Africa, pp. 45–53.

Douglas, L.J. & Haymer, D.S. (2001) Ribosomal ITS1 polymorphism in *Ceratitis capitata* and *Ceratitis rosa* (Diptera: Tephritidae). *Annals of the Entomological Society of America* **94**, 726–731.

Duyck, P.F. & Quilici, S. (2002) Survival and development of different life stages of three *Ceratitis* spp. (Diptera: Tephritidae) reared at five constant temperatures. *Bulletin of Entomological Research* **92**, 461–469.

Duyck, P.F., David, P. & Quilici, S. (2004) A review of relationships between interspecific competition and invasions in fruit flies (Diptera: Tephritidae). *Ecological Entomology* **29**, 511–520.

Freidberg, A. (1991) A new species of *Ceratitis* (*Ceratitis*) (Diptera: Tephritidae), key to species of subgenera *Ceratitis* and *Pterandrus*, and record of *Pterandrus* fossil. *Bishop Museum Occasional Papers* **31**, 166–173.

Gasparich, G.E., Sheppard, W.S., Han, H.-Y., McPheron, B.A. & Steck, G.J. (1995) Analysis of mitochondrial DNA and development of PCR-based diagnostic markers for Mediterranean fruit fly (*Ceratitis capitata*) populations. *Insect Molecular Biology* **4**, 61–67.

Gasparich, G.E., Silva, J.G., Han, H.-Y., McPheron, B.A., Steck, G.J. & Sheppard, W.S. (1997) Population genetic structure of the Mediterranean fruit fly (Diptera: Tephritidae) and implications for worldwide colonization patterns. *Annals of the Entomological Society of America* **90**, 790–797.

Han, H.-Y. & McPheron, B.A. (1997) Molecular phylogenetic study of Tephritidae (Insecta: Diptera) using partial sequences of the mitochondrial 16S ribosomal DNA. *Molecular Phylogenetics and Evolution* **7**, 17–32.

Han, H.-Y. & McPheron, B.A. (2000) Nucleotide sequence data as a tool to test phylogenetic relationships among higher groups of Tephritidae: a case study using mitochondrial ribosomal DNA. pp. 115–134 *in* Aluja, M. & Norrbom, A.L. (*Eds*) *Fruit flies (Tephritidae): phylogeny and evolution of behavior*. Boca Raton, Florida, CRC Press.

Kakouli-Durante, T., Casey, D.G. & Burnell, A.M. (2001) Development of a diagnostic DNA probe for the fruit flies *Ceratitis capitata* and *Ceratitis rosa* (Diptera: Tephritidae) using amplified fragment-length polymorphism. *Journal of Economic Entomology* **94**, 989–997.

Kibogo, H.G. (2005) Identification of fruit fly species in the genera *Ceratitis* and *Trirhithrum* (Diptera: Tephritidae). MSc thesis, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya.

Liquido, N.J., Barr, P.G. & Cunningham, R.T. (1998) MEDHOST, an encyclopedic bibliography of the host plants of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), version 1. Diptera dissemination disk 1 *in* Thompson, F.C. (*Ed.*) *Fruit fly expert identification system and systematic information database*.

Malacrida, A.R., Marinoni, F., Torti, C., Gomulski, L.M., Sebastiani, F., Bonvicini, C., Gasperi, G. & Gugliemino, C.R. (1998) Genetic aspects of the worldwide colonization process of *Ceratitis capitata*. *Journal of Heredity* **89**, 501–507.

Muraji, M. & Nakahara, S. (2002) Discrimination among pest species of *Bactrocera* (Diptera: Tephritidae) based on PCR–RFLP of the mitochondrial DNA. *Applied Entomology and Zoology* **37**, 437–446.

Naelole, C.K.M. & Haymer, D.S. (2003) Use of oligonucleotide arrays for molecular taxonomic studies of closely related species in the oriental fruit fly (*Bactrocera dorsalis*) complex. *Molecular Ecology* **3**, 662–665.

Ochando, M.D., Reyes, A., Callejas, C., Segura, D. & Fernandez, P. (2003) Molecular genetic methodologies applied to the study of fly pests. *Trends in Entomology* **3**, 73–85.

Salazar, M., Theoduloz, C., Vega, A., Poblete, F., Gonzalez, E., Badilla, R. & Meza-Basso, L. (2002) PCR–RFLP identification of endemic Chilean species of *Rhagoletis* (Diptera: Tephritidae) attacking Solanaceae. *Bulletin of Entomological Research* **92**, 337–341.

Saunders, M., Coulson, R.N. & Folse, L.J. (1993) Natural resource management and agriculture, applications and artificial intelligence. pp. 149–162 *in* Kent, A. & Williams,

J.G. (*Eds*) *Encyclopedia of microcomputers vol. 12, multistrategy learning to operations research, microcomputers applications*. New York, Marcel Dekker, Inc.

Saunders, M.C. & Miller, B.J. (1997) A graphical tool for knowledge engineers designing natural resource management software: NetWeaver. *1997 ACSM/ASPRS Annual Convention and Exposition Technical Papers, Seattle, Washington*: *Resource Technology*, Vol. 4, 380–389.

Sheppard, W.S., Steck, G.J., and McPheron, B.A. (1992) Geographic populations of the medfly may be differentiated by mitochondrial DNA variation. *Experientia* **48**, 1010–1013.

Silva, J.G., Meixner, M.D., McPheron, B.A., Steck, G.J. & Sheppard, W.S. (2003) Recent Mediterranean fruit fly (Diptera: Tephritidae) infestations in Florida – a genetic perspective. *Journal of Economic Entomology* **96**, 1711–1718.

Smith, P.T., McPheron, B.A. & Kambhampati, S. (2002) Phylogenetic analysis of mitochondrial DNA supports the monophyly of Dacini fruit flies (Diptera: Tephritidae). *Annals of the Entomological Society of America* **95**, 658–664.

Spanos, L., Koutroumbas, G., Kotsyfakis, M. & Louis, C. (2000) The mitochondrial genome of the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Molecular Biology* **9**, 139–144.

Steck, G.J., Carroll, L.E., Celedonio, H. & Guillen, J.C. (1990) Methods for identification of *Anastrepha* larvae (Diptera: Tephritidae), and key to 13 species. *Proceedings of the Entomological Society of Washington* **92**, 333–346.

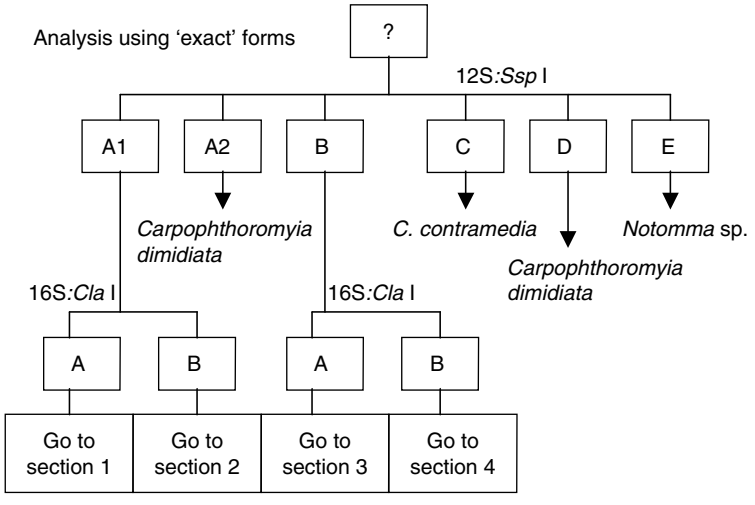
Vera, M.T., Rodriguez, R., Segura, D.F., Cladera, J.L. & Sutherst, R.W. (2002) Potential geographical distribution of the Mediterranean fruit fly, *Ceratitis capitata* (Diptera: Tephritidae), with emphasis on Argentina and Australia. *Environmental Entomology* **31**, 1009–1022.

White, I.M. & Elson-Harris, M.M. (1992) *Fruit flies of economic significance: their identification and bionomics*. 600 pp. Wallingford, Oxon, CAB International.

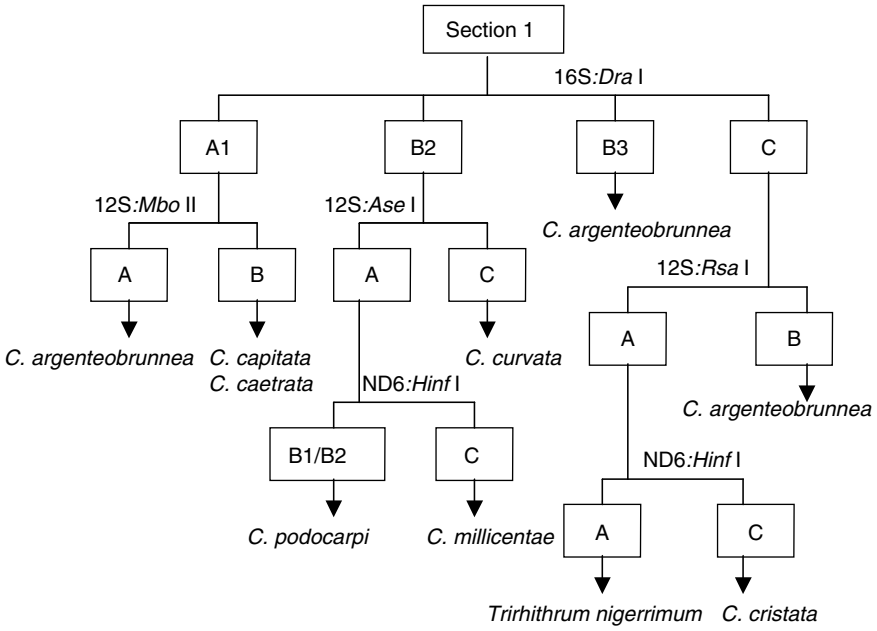
Yuval, B. & Hendrichs, J. (2000) Behavior of flies in the genus *Ceratitis* (Dacinae: Ceratitidini). pp. 429–457 *in* Aluja, M. & Norrbom, A.L. (*Eds*) *Fruit flies (Tephritidae): phylogeny and evolution of behavior*. Boca Raton, Florida, CRC Press.

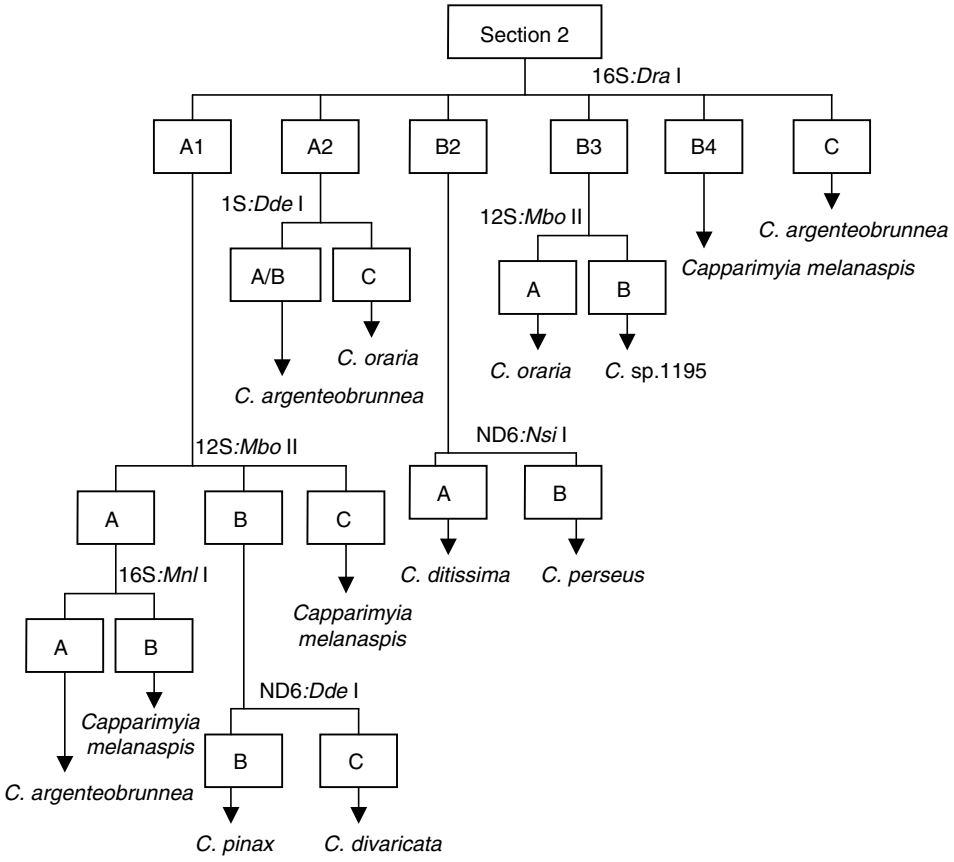
Appendix 1

Flow chart for interpreting PCR–RFLP results using 'exact' forms. (*C.* = *Ceratitis*.)

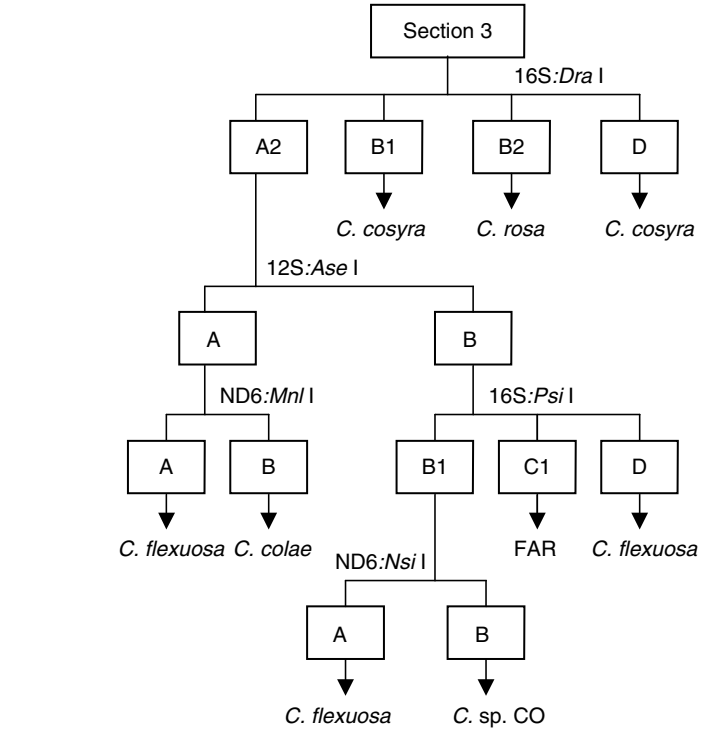


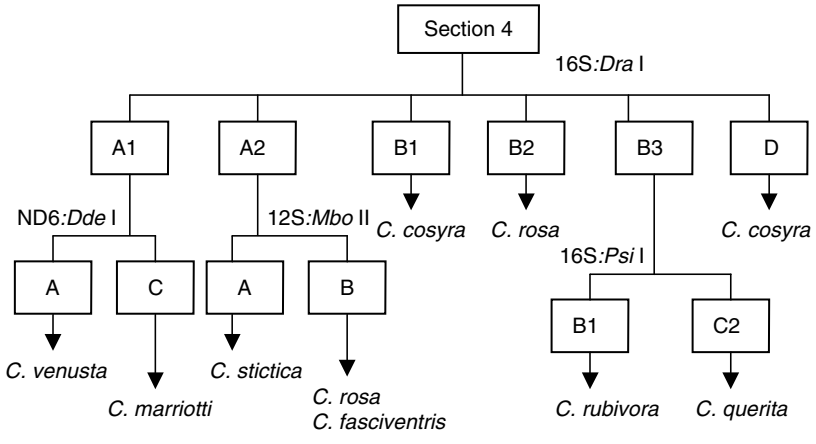
https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press





https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

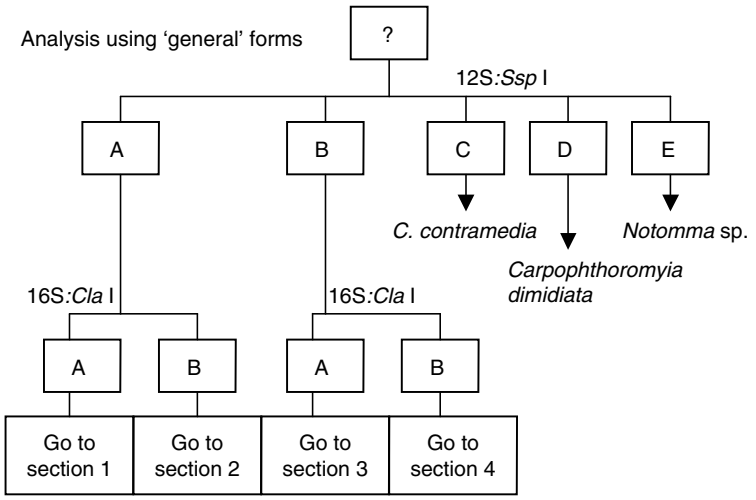


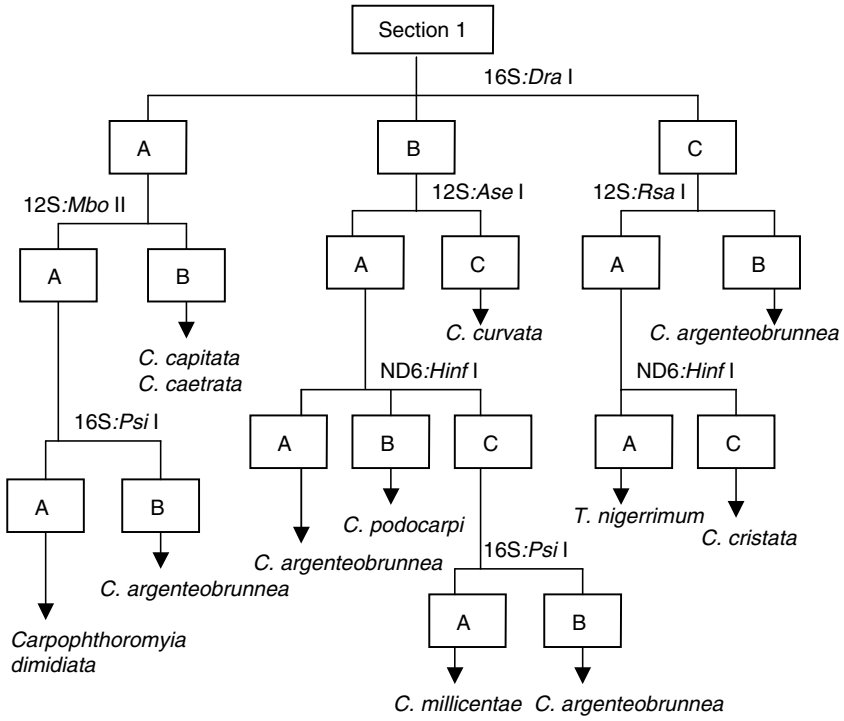


https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press

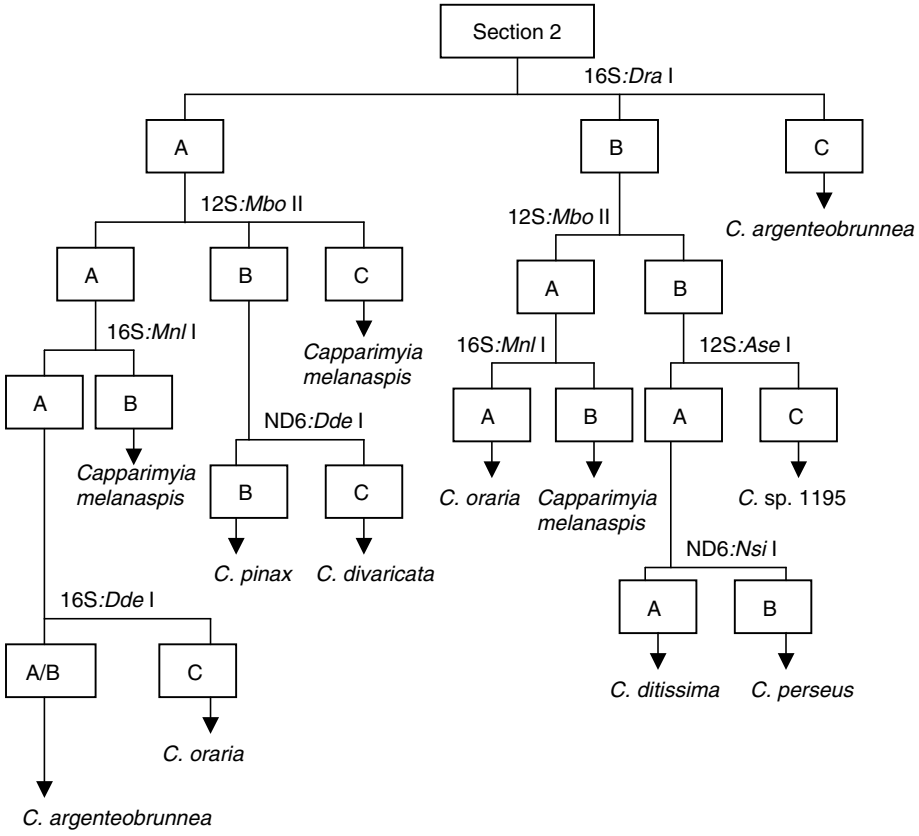
Appendix 2

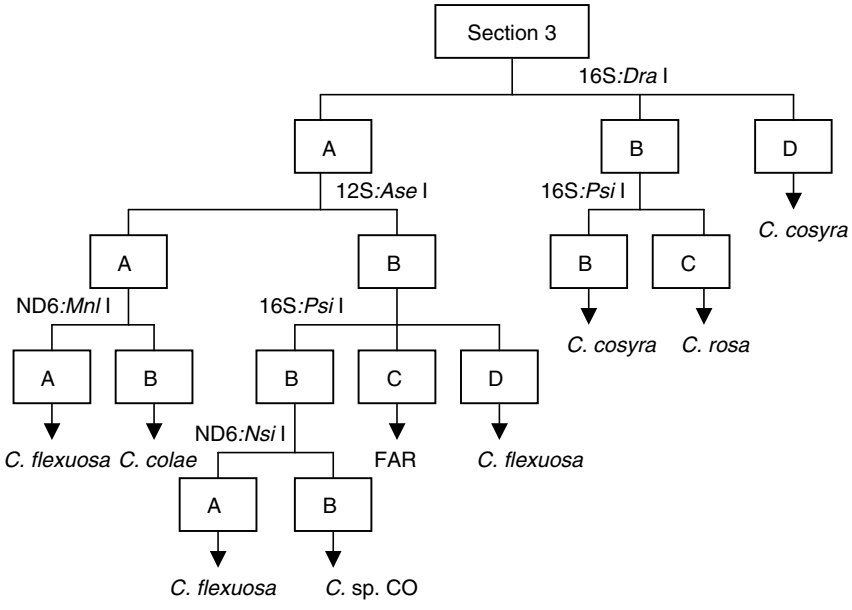
Flow chart for interpreting PCR–RFLP results using 'general' forms. (*C.* = *Ceratitis*.)



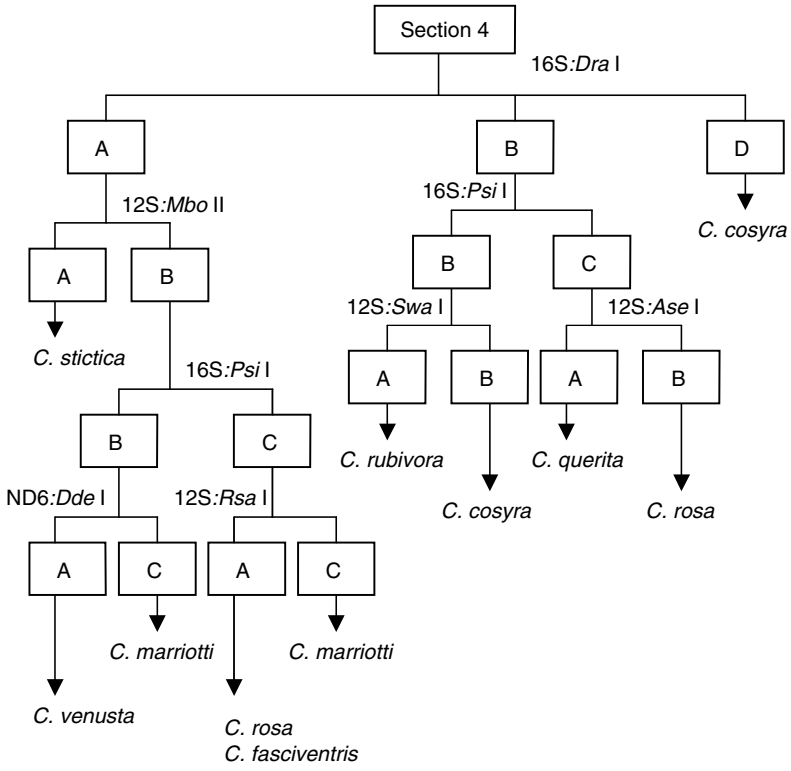


https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press





https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press



(Accepted 13 June 2006)

https://doi.org/10.1079/BER2006452 Published online by Cambridge University Press