Agricultural and Forest Entomology (2002) **4**, 127–136

A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) using direct sequencing and a PCR-RFLP-based approach

P. C. Brunner*, C. Fleming† and J. E. Frey*

**Swiss Federal Research Station, PO Box 185, CH-8820 Wädenswil, Switzerland and* †*Agriculture and Food Science Centre, Department of Agriculture and Rural Development and The Queen's University of Belfast, Newforge Lane, Belfast BT9 5PX, Northern Ireland*

Abstract

1 Treatments against pathogens or pests are often very specific and, as a fundamental first step, require the ability to identify taxa correctly and unambiguously. We used PCR amplification techniques to successfully establish a molecular identification key for economically important thrips species.

2 A PCR amplified 433 bp long fragment of the mitochondrial COI coding gene was analysed by automated direct sequencing and RFLP. Sequencing of 264 individual thrips representing 10 named species detected 17 haplotypes. Variation within species was low, whereas among species variation was high resulting in an average sequence divergence of 18.6% and an average pairwise species differentiation (calculated as F_{ST}-value) of 0.9896.

3 Two restriction enzymes (*Alu*I, *Sau*3AI) produced patterns that allowed unambiguous identification of all thrips species.

4 Statistical support for the quality of the key was given by (i) a highly significant permutation approach, assigning individual haplotypes to the correct species groups and (ii) a hierarchical NJ cluster analysis in which all conspecific individual sequences clustered together with maximal (100%) bootstrap support.

5 This study has shown that the use of genetic markers represents a valuable alternative for situations, such as epidemiological research, in which correct identification with classical morphological methods is either very difficult and time consuming or virtually impossible.

Keywords Cluster analysis, direct sequencing, molecular identification, mtDNA, PCR-RFLP, thrips.

Introduction

Thrips (Thysanoptera; Thripidae) are very small insects, widespread throughout the world with a preponderance of tropical species. Of the approximately 5000 species so far identified, a few hundred are crop pests, causing serious damage or transmitting diseases to growing crops and harvestable produce in most countries. Although weak flyers, their fringed wings enable them to remain airborne long enough to travel easily between neighbouring fields and be blown by the wind over far greater distances (Lewis, 1997). Additionally, their minute size and cryptic behaviour make them difficult to detect either in the field or in fresh vegetation transported through international trade of vegetables, fruit and ornamental flowers. Consequently, many species have now spread from their original natural habitats and hosts to favourable new environments, including valuable crops. Here, they often reproduce rapidly to develop damaging infestations (often coupled with resistance to many pesticides) that are costly to control.

As a fundamental first step, plant quarantine diagnosis and treatments against pathogens require the ability to identify taxa rapidly and correctly. Taxa are traditionally distinguished using morphological characters. However, not all species lend themselves to this approach because of insufficient phenotypic variation. Larval thrips, for

Correspondence: Jürg Frey. Tel. + 41 1783 6332; fax: + 41 1783 6434; e-mail: juerg.frey@faw.admin.ch

example, are often mistaken for Collembola (springtails), whereas adults are commonly confused with Staphylinid beetles (Vierbergen, 1995). In most cases identification of larval Thysanoptera to species is impossible without the presence of adults. These are serious drawbacks, given that some thrips species are predatory. For example, *Karnyothrips* species attack immature scale insects and whiteflies as well as mites (Palmer & Mound, 1991), and *Franklinothrips* species are natural enemies of pest thrips (Loomans & Heijboer, 1999).

The use of genetic markers represents a valuable addition or alternative to traditional phenotypic methods of species recognition, as the development of molecular techniques during the last two decades has provided a variety of simple, yet robust and reliable, tools. Since its development, the polymerase-chain-reaction (PCR; Saiki *et al*., 1988) and PCR-based typing methods have made a considerable contribution to the biological sciences. For example, PCR-based methods have provided ideal markers for species identification (Frey & Frey, 1995), and the study of parasite epidemiology (Hide & Tait, 1991; Vago *et al*., 1996). By its nature, PCR is an ideal tool for use in pathogen and parasite identification, especially in a diagnostic context when only minimal amounts of template DNA are available. It is technically simple to employ, requiring only basic laboratory skills, and, once established, it is rapid, sensitive and specific.

The objective of this study was to determine whether current PCR-based DNA techniques are suitable to establish a molecular identification key for agronomically important thrips species.

Materials and methods

Collection

Samples from 10 thrips species (see Table 1) were obtained from laboratory colonies, by direct collection from infested plants and flowers, or by exposure of sticky traps. Only individuals (mainly adults) that could be identified to species level were used for subsequent analyses to establish the molecular key. Within-species genetic polymorphism was assessed by analysing several individuals per species collected from different populations or countries (including England, Israel, North America, Northern Ireland, Scotland, Switzerland and the Netherlands). Additionally, two thrips that could not be unambiguously identified to species levels were included for phylogenetic analyses and to assess limitations (see below).

Molecular protocols

Total genomic DNA was extracted from single thrips using the slightly modified protocol by Kawasaki (1990). Briefly, individual thrips were placed in a 0.5 mL microcentrifuge tube containing 50 µL of lysis buffer. Plastic grinders or toothpicks were used to crush the insects, and the tubes were then incubated at 85 °C for 15 min. The homogenate was stored at −20 °C and used without any further preparation for subsequent PCR amplification.

After evaluation of several mitochondrial and nuclear target sequences (Frey *et al*., unpublished data), a portion of the mitochondrial cytochrome oxidase I gene (COI) was amplified via a standard PCR reaction using the 'universal' primers C1-J-1751 and C1-N-2191 (Simon *et al*., 1994). Detailed description of basic PCR methodology can be found elsewhere (e.g. Innis *et al*., 1990; Palumbi, 1995). However, the quality and amount of PCR product obtained varied considerably between different species due to differences in the primer recognition sequences. Subsequent modification of primers (resulting in the degenerate primers: mtD-7.2F 5′-ATTAGGAGCHCCHGAYATAGCATT-3′, mtD-9.2R 5′-CAGGCAAGATTAAAATATAAACTTCTG-3′) to better match the thrips sequences generated satisfactory PCR results for all species.

DNA sequences were generated directly using an Applied Biosystems (Foster City, CA) automated sequencer. All DNA was sequenced in both directions to assure accuracy of nucleotide assignments, and aligned with the multiple sequence editor CLUSTAL X (Thomson *et al*., 1997). However, not all laboratories possess or have access to an automated sequencer. Therefore, we evaluated PCR-RFLP (restriction fragment length polymorphism) as a second molecular approach that is widely used, relatively inexpensive and highly reproducible, a prerequisite when applied in different laboratories and on different equipment. The program

Table 1 Values for haplotype diversity (*h*) and nucleotide diversity (π) within thrips species groups, number of individuals analysed per species (*N*), and GenBank (GB) accession numbers with haplotype abbreviations used

Species	*n*	Number of haplotypes	Haplotype diversity (h)	Nucleotide diversity (π)	GB
Frankliniella occidentalis	132	4	0.632 ± 0.035	0.0024 ± 0.0018	AF378685 – AF378688 (Focc 1–4)
Parthenothrips dracaenae	14	1	0	0	AF378681 (Pdra)
Anaphothrips obscurus	12	1	0	0	AF378684 (Aobs)
Thrips palmi	16	2	0.400 ± 0.114	0.0009 ± 0.0010	AF378689 – AF378690 (Tpal1–2)
Thrips tabaci	12	2	0.530 ± 0.076	0.0061 ± 0.0040	AF378692 – AF378693 (Ttab1–2)
Thrips angusticeps	23	2	0.166 ± 0.098	0.0008 ± 0.0009	AF378679 – AF378680 (Tang1–2)
Echinothrips americanus	12	2	0.303 ± 0.148	0.0014 ± 0.0013	AF378677 – AF378678 (Eame1–2)
Hercinothrips femoralis	10	1	0	0	AF378682 (Hfem)
Heliothrips haemorrhoidalis	22	1	0	0	AF378683 ((Hhae)
Taeniothrips picipes	11	1	0	0	AF378691 (Tpic)

GeneJockey II (Biosoft, Cambridge, UK) was used to analyse the thrips' nucleotide sequences and predict potential restriction patterns. Subsequently, candidate enzymes were screened *in vitro* using standard RFLP protocols. Digestion fragments were separated by electrophoresis in 1–2% agarose gels. Gels were stained with ethidium bromide, visualized and photographed under UV light.

Analyses

Statistical analyses were performed with ProSeq v.2.7.1. (D. Filatov; http://helios.bto.ed.ac.uk/evolgen/filatov/proseq.html) and included estimation of within-species groups mitochondrial DNA polymorphism expressed as haplotypic (nucleon) diversity (*h*; Nei & Tajima, 1981) and as nucleotide diversity (π; Nei, 1987). Between-species divergence was estimated as F_{ST} statistics (a measure for population differentiation; Weir & Cockerham, 1984) and as net sequence divergence (nucleotide p-distance) averaged over within-species groups haplotypes. ProSeq also estimates the significance of the observed F_{ST} statistic via a permutation approach, each time randomly assigning individual haplotypes to the species groups. If the observed F_{ST} value exceeds 95% (99% or 99.9%) of the simulated values, the differences between the species groups are significant.

An alternative (apart from statistical tables) to illustrate and investigate within- and among-species polymorphism is a graphical illustration of genetic differentiation. First, genetic distances (Kimura-2-parameter) were calculated for all pairwise comparisons of individual haplotypes and the resulting distance matrix was subjected to a three-dimensional (3D) data reduction procedure using SYSTAT vs. 7.0 (SPSS Inc. Chicago, IL, USA). However, this 3D approach alone is not very popular for two main reasons. First, the illustration becomes confusing when many taxa are considered; second, there is no convincing statistical method available to illustrate significance of differentiation among taxa. Therefore, we also applied phylogenetic reconstruction methods using hierarchic neighbour-joining (NJ) genetic distance clustering, as implemented by the computer program MEGA v.2.0 (Kumar *et al.*, 1994; http://www.megasoftware.net/) and parsimony as implemented by the computer program PAUP* (Swofford, 1999). Sequences from 23 additional thrips species retrieved from GenBank were added to these analyses to increase the data set and, consequently, to gain deeper insight into the general utility and robustness of the molecular key. To test for statistical significance of the generated trees, data were resampled 1000 times to obtain bootstrap *P*-values. The bootstrap (Felsenstein, 1985) involves creating a new data set by sampling *N* characters randomly with replacement, so that the resulting data set has the same size as the original, but some characters have been left out and others are duplicated. The random variation of the results from analysing these bootstrapped data sets can be shown statistically to be typical of the variation obtained from collecting new data sets. In short, bootstrap values are confidence values for the correctness of the cluster (i.e. sequence groupings) to the right of a particular node in the tree.

Results

Nucleotide sequence and RFLP variation

A total of 264 individual thrips representing 10 named species were analysed. For each individual, a 433-bp long fragment of the COI gene was obtained by PCR amplification and direct sequencing. All observed variation was in the form of silent single base-pair substitutions at 3rd base positions (except for position 379 in *F. occidentalis*). A total of 200 (46%) nucleotide positions were polymorphic, and these defined 17 haplotypes (Fig. 1; Table 1). Several individuals (minimum 10 – maximum 132) were analysed per species to assess within-species genetic polymorphism. Except for five species, all individuals from the same species had identical sequences (Table 1). Differences among conspecific haplotypes were only marginal. For example, a maximum of six nucleotide positions was variable in 132 individuals of *F. occidentalis*. Consequently, measures for within-species polymorphism are very low (*h*: 0.166–0.632; π: 0.0008–0.0061; Table 1) even for these variable species.

In sharp contrast to the low within-species genetic variation, differences between species-specific haplotypes were high. Pairwise comparisons of F_{ST} and p-distance values are summarized in Table 2. The most divergent haplotypes (Pdra/Ttab2) differed by 119 substitutions (27.5%). The lowest F_{ST} value (population differentiation) calculated is 0.9713 (*T. angusticeps* vs. *T. tabaci*), and 10 out of 45 possible comparisons attained the maximum value of 1.000. Average F_{ST} value is 0.9896. Similarly, percent sequence divergence (expressed as p-distance) among species was high (16–27.5%) with an average of 19% ($\pm$ 1%). Alignment of thrips sequences with those retrieved from GenBank revealed variability at an additional 81 nucleotide positions. Average sequence divergence between haplotypes of all 33 species was 21% ($\pm$ 1%).

The program GeneJockey II was first used to evaluate potential restriction patterns from the obtained nucleotide sequences. Twelve candidate enzymes were screened *in vitro* using standard RFLP protocols. Of the tested restriction enzymes, two (*Alu*I, *Sau*3AI), producing strong homologous bands as verified by direct sequencing, were chosen for subsequent analyses. These two enzymes produced restriction patterns that allowed unambiguous identification of the thrips species assessed in this study (Fig. 2; Table 3).

Quality assessment

The quality (i.e. the 'robustness') of a molecular key has to be addressed by applying statistical tests to assess the probability of misidentification/misassignment of individuals to a particular species. This was done in two ways. First, we estimated the significance of the observed F_{ST} statistic (i.e. 'the goodness' of species differentiation; Table 2) via a permutation approach. Calculation of 5000 permutations resulted in $P < 0.001$, indicating that correct assignment of individual haplotypes to species groups was excellent and that the differences between species were highly significant

```
F. occidentalis       CCT CGA CTW AAT AAC ATA AGA TTT TGA CTT CTT CCA CCC TCT TTA ACA TTG TTA ATT ATA  60
P. dracaenae          ..C ... T.A ... ..T ... ..G ... ... ... T.A ..T ..G ... ... .TT ..T A.. ... T..
A. obscurus           ... ..G T.G ... ..T ..G ... ... ... ... T.G ..T ..T ... A.T .TG ..A ... ... T..
T. palmi              ..A ... T.A ... ..T ... ... ... ... ... ... ... ..T ... ... ..C C.C ... ... ..G
T. tabaci             ... ... T.A ... ..T ... ... ..C ... ... T.A ..C ..T ... C.G GG. ..A ... ... ...
E. americanus         ... ..T T.A ... ..T ... ... ... ... ... T.A ..T ..T ... ... ..T C.T ... ... T..
T. angusticeps        ..A ... A.G ... ..T ... ... ... ... ..A T.A ..T ..A ..A ... .T. A.T C.T T.A ...
H. femoralis          ..A ... T.A ... ... ... ... ..C ... T.. T.A ... ..A ..A A.T TTT ..A ... ... T..
H. haemorrhoidalis    ... ... ..A ... ..T ... ... ... ... T.. T.A ... ..A ... A.T CT. ..A ... ... T..
T. picipes            ..A ... T.A ..C ... ... ... ... ... ... ... ... ..T ... C.. ..T ..A ... ... ...

F. occidentalis       GGT TTA TCA AAA GAT GGT GCA GGA ACA GGA TGA ACA GTT TAC CCA CCT TTG TCA ACY TTT  120
P. dracaenae          A.A A.G CTC GGT ... ..A C.T ..T ... ..T ..G ... C.A ..T ... ... C.A ... ..T .A.
A. obscurus           ..G ..T ... GG. ..A ..A ... ... ..T ... ... ..G ... ..T ..T ... C.T ... ..A ...
T. palmi              ... ... .AT ... ..A ..A ... ... ... ... ... ... ... ..T ... ... ..A ... ..A ...
T. tabaci             ..A C.T .AT ... ..A ..A ..G ... ..G ... ... ... ..A ..T ... ... ..A ..W ..G ...
E. americanus         ... ... ... ... ..A ..A ... ... ... ... ... ... A.. ..T ... ..Y C.A ... ..A ...
T. angusticeps        ..A C.T ..T ... ..A ... ... ... ... ... ... ... ... ..T ..T ... C.T ... ..A ...
H. femoralis          ... A.. .TT GG. ..A ... ..T ... ..G ..G ... ..T ... ... ... ..A ..A ... ..T ...
H. haemorrhoidalis    ..A A.G ATG GG. ... ..A ... ... ... ..T ... ... ... ..T ..T ... C.T ... ..A ...
T. picipes            ..A ... AG. ... ..A ..A ..G ... ... ... ... ... ..C ... ... ... ..A ... ..A ...

F. occidentalis       TAT CAC TCT GGA CCA TCA GTA GAT TTA ACT ATY TTT TCC CTT CAT TTA GCA GGT ATT TCT  180
P. dracaenae          .T. ..T ..A ... ATT ..C ..T ..C ... ..A ..T ... ..T ... ... ... ..T ..A ... ...
A. obscurus           .T. ..T ... ..T ATT ... ... ... ... ... ..T ..C ..T T.A ..C C.G ... ... G.A ..A
T. palmi              ... ..T G.. ... ATT ... ... ... ... ..A ..T ... ..T ..C ... ... ..T ..G G.A ..C
T. tabaci             ... ..T ..A ... ..T ... ... ..C ... ..A ..T ... ..T T.A ..C C.T ... ..G ... ...
E. americanus         ... ..T ..A ... AG. ..T ... ..C ..T ..A ..T ... ..T T.A ..C C.. ... ..A ... ..A
T. angusticeps        ... ... ..A ..T ATT ..G ..T ... ... ..A ..T ... ..T T.A ... ... ... ..R ... ..C
H. femoralis          ..C ..T ..A ... ATT ... ..T ... ... ..A ..T ... ..A T.A ... ... ... ..A ... ..C
H. haemorrhoidalis    ... ..T ..A ... ATT ..C ..G ..C ... ..A ..T ... ..A ... ... C.T ... ... ... ..A
T. picipes            ... ..T ..A ... AT. ... ..T ..C ... ... ..T ... ..T ... ..C C.. ..G ... ... ..A

F. occidentalis       TCA ATT CTA GGA GCT TTA AAT TTT ATT ACA ACG ATT TTA AAT TTA AAG ATC AAA AAA TTA  240
P. dracaenae          ..T ... T.G ... ..A G.T ... ... ... ... ... ... G.C ... ... CGA .A. ... C.. ..T
A. obscurus           ... ... T.. ... ..A ... ... ... ... T.T ..A ... ... ... ... ..A .GA ... G.T ...
T. palmi              ... ... T.. ... ..A ... ... ..C ..C ..T ..A ... ... ... ... ..A .AT ... ..T C.T
T. tabaci             ... ... T.. ..T ..C ... ... ... ... ..T ..A ... A.T ... C.T ..A GCA ... ..C C.T
E. americanus         ... ... T.. ... ... ..C ... ... ... ... ..A ... ... ... ... ..A .AT ... ... C.T
T. angusticeps        ... ... T.. ... ..A G.T ... ... ... ... ..T ... ... ..C ... ..A .AT ... ... A.C
H. femoralis          ..T ... T.. ... ... G.. ... ..C ... T.. ..T ... ..C ... C.. CGT .A. ... .T. C..
H. haemorrhoidalis    ... ... T.. ... ..A G.. ... ..C ... T.. ..T ... C.. ... ... CGT .A. ... G.. C..
T. picipes            ... ... T.. ..G ..A C.. ... ... ..C ... ..A ... A.T ... C.. ..A .GA ... G.C C..

F. occidentalis       ACA ACG GAA AAG ATA ACT TTA TTT GTT TGA TCA GTT ATT TTA ACA GCT ATT TTA TTA TTA  300
P. dracaenae          .A. GTT ... TCT C.. T.. ... ... AC. ... ..T ... T.A ... ... ... ..C ..G ..G C..
A. obscurus           .G. TTT ... ..A .CT .GA C.T ... ... ... ... ... ... ..G ..G ... ..C C.. ... C.T
T. palmi              T.. .GA ... ..A C.. .GA ... ..Y ..G ... ... ... ..A ... ... ..A ... C.T C.G C.T
T. tabaci             T.. G.A ... ..A ..T .GA C.. ... ..C ... ... ... ... ... ... ..C ... C.T C.T C.T
E. americanus         T.. .GT ... ..A .AT T.. ... ... ... ... ... ..A ..C ... ..A ... ... C.T ... C.T
T. angusticeps        T.T .TA ... ..A G.T .GY ... ... A.. ... ... ... T.A ... ... ... ... ... C.T C.T
H. femoralis          T.C .TA ... .GT T.. ... ... ... TC. ... ..T ..A T.A ... ... ..A ... ... C.T C.T
H. haemorrhoidalis    .A. .TA ... ... T.. ..A ... ... ACA ... ... ... T.. A.C ..T ..A ... ... C.. ...
T. picipes            T.. G.A ... ..A ... .G. ... ... ... ... ... ..A ..A C.. ... ..A ... C.. ... ...

F. occidentalis       TTG TCG TTA CCA GTT TTA GCA GGA GCT ATT ACR ATA TTA TTA ACA GAC CGA AAC TTA AAT  360
P. dracaenae          ..A ..T ... ..T ... ... ... ... .GA ... ..T ... A.T ... ..G ..T ... ..T ... ..C
A. obscurus           ..A ..T ... ..T ..A ... ..T ... ..C ... ..T ... C.T ... ... ..T ... ..T ... ...
T. palmi              ..A ..T ... ..T ... C.T ... ... ..C ..C ..A ... C.T ... ... ... ... ..T ... ..C
T. tabaci             ..A ..T ..R ... ..K ... ..G ... ... ..C ..A ... C.T ... ..T ... ... ... ... ...
E. americanus         C.A ..A ... ..T ..C ..R ... ... ... ... ..T ... C.T ... ... ..T ... ..T ... ..C
T. angusticeps        ..A ..T C.T ..T ..A ... ... ... ... ... ..A ... C.T ... ... ... ..T ... C.. ...
H. femoralis          ..A ..A ... ... ... C.. ..T ... ..A ..C ..A ... C.G ..G ... ..T ... ..T ... ...
H. haemorrhoidalis    ..A ..A ... ..T ... ... ... ... .GA ... ..A ... ... ... ... ... ... ... ..C
T. picipes            ..A ..T ... ..C ... C.. ... ... ... ... ..T ... C.T C.. ... ..T ... ..T C.. ...

F. occidentalis       ACA TCC TTC TTT GAT CCG RGA GGA GGT GGR GAC CCA GTT TTA TAC CAG CAC TTG TTT TGA  420
P. dracaenae          ..T ..T ..T ... ... ..A A.T ..G ..G ..G ..T ... A.. ... ... ..A ..T ..A ... ...
A. obscurus           ..T ... ..T ... ..C ..A A.. ..G ..A ..T ..T ... ACC C.T ..T ..A ... ..A ... ...
T. palmi              ... ..A ... ... ... ..A A.G ... ..A ..A ..T ... ..A C.T ... ..A ..T C.A ..C ...
T. tabaci             ..C ..T ..T ... ..C ..T A.. ..R ..A ... ... ..T ... ... ..T ..A ... C.T ... ...
E. americanus         ... ..T ..T ... ..C ..T G.. ... ..A ..A ..T ..T A.. C.T ..T ..A ..T ..A ..C ...
T. angusticeps        ... ..A ..T ... ... ..T A.. ... ..A ..A ..T ... ... ... ... ..A ..T ..A ... ...
H. femoralis          ..C ..T ..T ... ... ..T A.. ... ..G ..T ... ... A.. ... ... ..A ..T ..A ..C ...
H. haemorrhoidalis    ..T ..A ..T ... ... ..T A.. ... ..A ..T ... ... A.. ... ..T ..A ... ..A ... ...
T. picipes            ... ..T ..T ... ... ..T A.. ..G ..A ..A ..T ... ... ... ... ..A ..T ..A ... ...

F. occidentalis       TTT TTT GGG CAC C  433
P. dracaenae          ... ..C ..T ..T .
A. obscurus           ... ... ..T ..T .
T. palmi              ... ... ..T ... .
T. tabaci             ... ... ..T ... .
E. americanus         ... ... ..T ..T .
T. angusticeps        ... ... ..T ..T .
H. femoralis          ... ... ..T ... .
H. haemorrhoidalis    ... ... ..A ..T .
T. picipes            ... ... ..T ..T .
```

Figure 1 Consensus sequence of 433 bp from the mitochondrial cytochrome oxidase I gene for 10 thrips species. Dots indicate nucleotides identical throughout the species compared. Polymorphic sites are indicated in bold and nucleotide symbols follow IUB code: R = G/A, Y = C/T; W = A/T; K = G/T.

Table 2 Above diagonal; F_{ST} values for pairwise comparisons of thrips species. Differences between all groups of sequences are significant for all pairwise comparisons ($P < 0.001$) as estimated via a permutation approach. Below diagonal; average sequence divergence (nucleotide p-distance).

Species	*F.o.*	*P.d.*	*A.o.*	*T.p.*	*T.t.*	*T.a.*	*E.a.*	*H.f.*	*H.h.*	*T.p.*
Frankliniella occidentalis	–	0.9911	0.9906	0.9841	0.9753	0.9744	0.9796	0.9910	0.9897	0.9882
Parthenothrips dracaenae	0.250	–	1.0000	0.9957	0.9898	0.9867	0.9907	1.0000	1.0000	1.0000
Anaphothrips obscurus	0.232	0.254	–	0.9946	0.9862	0.9847	0.9891	1.0000	1.0000	1.0000
Thrips palmi	0.192	0.247	0.206	–	0.9786	0.9764	0.9828	0.9953	0.9953	0.9947
Thrips tabaci	0.212	0.274	0.212	0.185	–	0.9713	0.9750	0.9882	0.9885	0.9844
Thrips angusticeps	0.202	0.226	0.215	0.182	0.206	–	0.9735	0.9855	0.9847	0.9860
Echinothrips americanus	0.197	0.220	0.176	0.160	0.185	0.179	–	0.9912	0.9912	0.9884
Hercinothrips femoralis	0.223	0.215	0.259	0.215	0.226	0.203	0.220	–	1.0000	1.0000
Heliothrips haemorrhoidalis	0.197	0.194	0.210	0.229	0.230	0.201	0.220	0.176	–	1.0000
Taeniothrips picipes	0.176	0.229	0.206	0.166	0.174	0.199	0.165	0.212	0.212	–

and diagnostic. Second, we used two graphical illustrations of genetic differentiation to illustrate and investigate within- and among-species polymorphism. Figure 3 shows the 3D picture of the Kimura-2-parameter distance matrix calculated for all pairwise comparisons of individual haplotypes. It is obvious, at least qualitatively, that the haplotypes belonging to the same species cluster closely together, whereas different species are clearly separated. This method of visualizing relationships between taxa generates confidence in taxa differentiation and has the advantage of not being hierarchical (see method below), i.e. relationships of taxa are not restricted to two dimensions. However, a major drawback is the lack of convincing statistics to support the grouping of haplotypes. In contrast, the widely used NJ clustering can address this problem by evaluating bootstrap replication of the original data set (Fig. 4). Similar to the 3D approach, all individuals that belong to the same species cluster together with maximal bootstrap values (100%). This means that in all the 1000 bootstrap replications none of the individual thrips haplotypes was assigned to the wrong species.

Interestingly, the phylogenetic aspect of the NJ tree suggested that the variation at COI, in addition to its powerful utility at the species level, also resolved a deep phylogenetic split (100% bootstrap support) based on morphological evidence, i.e. between the two suborders Tubulifera and Terebrantia. However, phylogenetic levels between these two extremes were not resolved. For example, species belonging to the genus *Oncothrips* within the suborder Tubulifera did not form a monophyletic cluster. Similarly, *Thrips* species from this study did not form a distinct cluster within the Terebrantia. Bootstrapping indicated that only the sister-group relationship of *O. habrus* and *O. tepperi* was unambiguously supported 97%; see also Crespi *et al.*, 1998 for additional comments on this finding). Although this observation is important and has to be discussed in detail in the future (e.g. comparison of phylogenetic trees derived from nuclear and mitochondrial genes and re-evaluation of possibly homoplasious morphological characters used in thrips taxonomy), this level of phylogenetic reconstruction is not of particular relevance to the establishment of an identification key for the species discussed here.

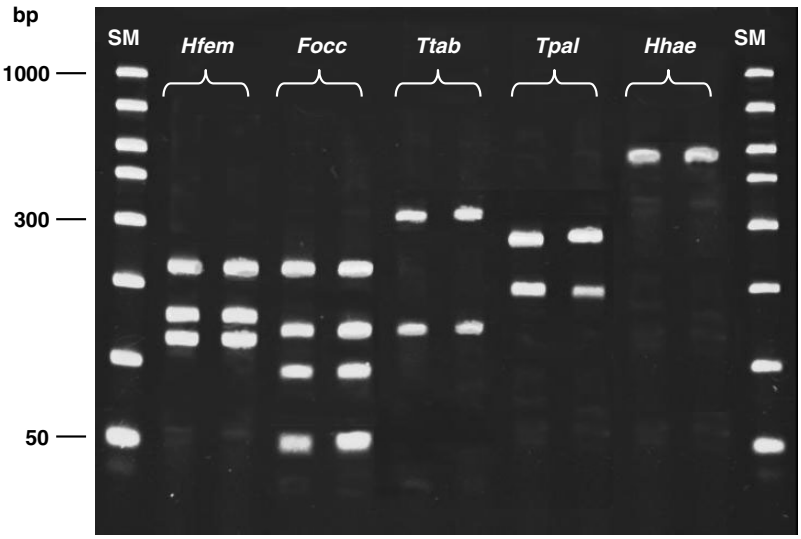

Figure 2 Example of an RFLP pattern (digested with *AluI*) separated on an agarose gel. Lanes labelled SM are size ladders in bp. Abbreviation are: Hfem, *Hercinothrips femoralis*; Focc, *Frankliniella occidentalis*; Ttab, *Thrips tabaci*; Tpal, *Thrips palmi*; Hhae, *Heliothrips haemorrhoidalis*.

Table 3 Calculated restriction fragment sizes (bp) following digestion of PCR-amplified COI region of 10 thrips species with two restriction enzymes: AluI, recognition sequence AG⊥CT; Sau3AI, recognition sequence ⊥GATC. Underlined cut sites are present only in a particular haplotype of a species. Fragment sizes are 'as appearing on gel', i.e. including flanking primers.

	F. occidentalis		*P. dracaenae*		*A. obscurus*		*T. palmi*		*T. tabaci*	
Enzymes	Cut site	Fragment size (bp)	Cut site	Fragment size (bp)	Cut site	Fragment size (bp)	Cut site	Fragment size (bp)	Cut site	Fragment size (bp)
*Alu*I	193	218	169	194	319	344	169	291	325	350
	286	135	286	177		141		194		135
	325	93		117						
		39								
*Sau*3AI	227	252	268	293	268	293	268	293	268	293
	268	104	345	77	345	77	372	104		192
	372	88	372	70	390	70	390	70		
		41	390	27		45		18		
				18						

	E. americanus		*T. angusticeps*		*H. femoralis*		*H. haemorrhoidalis*		*T. picipes*	
Enzymes	Cut site	Fragment size (bp)	Cut site	Fragment size (bp)	Cut site	Fragment size (bp)	Cut site	Fragment size (bp)	Cut site	Fragment size (bp)
*Alu*I	193	218	257	282	193	218	–	485	325	350
	286	135	286	135	319	126				135
	325	93	325	39		141				
	29									
*Sau*3AI	133	135	268	293	268	293	268	293	268	293
	268	158	372	104	345	77	372	104	345	77
	345	77	390	70	372	88		88	372	70
	390	70		18		27			390	27
		45								18

Discussion

DNA sequences generated by PCR have tremendous utility in the identification of species. Over the past two decades the methodology of molecular systematics has developed to allow reliable construction and interpretation of phylogenetic trees from DNA data. Confidence in phylogenetic reconstruction can be established statistically using procedures like permutations tests (Faith, 1991) or bootstrap resampling (Felsenstein, 1985). Much of this methodology is now widely applied in forensic studies and conservation genetics to unambiguously identify species or evolutionary significant units (Hillis *et al*., 1994; Baker & Palumbi, 1995).

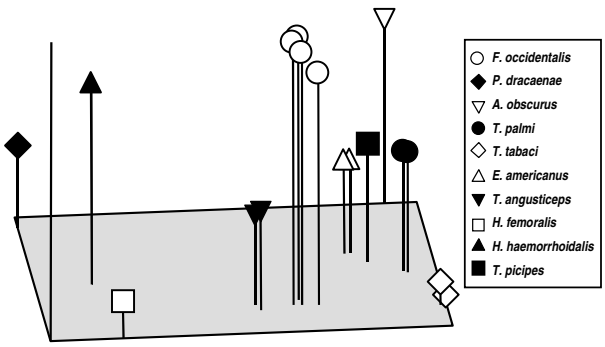

Figure 3 Three-dimensional clustering of pairwise genetic distances (Kimura-2-parameter) among 17 different haplotypes detected by direct sequencing in 10 thrips species.

However, certain constraints have to be considered in the use of PCR and molecular systematics for the establishment of a molecular identification key. These constraints fall into two major categories: (i) technical problems in the collection of molecular data using PCR and (ii) limitations on the analysis of species-level systematic relationships using molecular data.

PCR and artefects

Polymerase errors occur at a low frequency when the *Taq* polymerase is used in PCR reactions. This is because *Taq* polymerase has no proof-reading function (i.e. no exonuclease activity). As a result, when an incorrect nucleotide is added to the growing DNA strand during the extension step of PCR, it is not removed or replaced with the correct nucleotide. These polymerase errors are rare (2.1×10^{-4} errors/bp; Keohavang & Thilly, 1989) and random in their distribution along the DNA strand produced with PCR. If PCR products are cloned, the sequence of each clone is an exact match to the single sequence that was inserted into the vector. As a result, a polymerase error will be carried through into the resulting sequence data. However, when PCR products are analysed by restriction digestion or by sequencing of the whole product (such as the

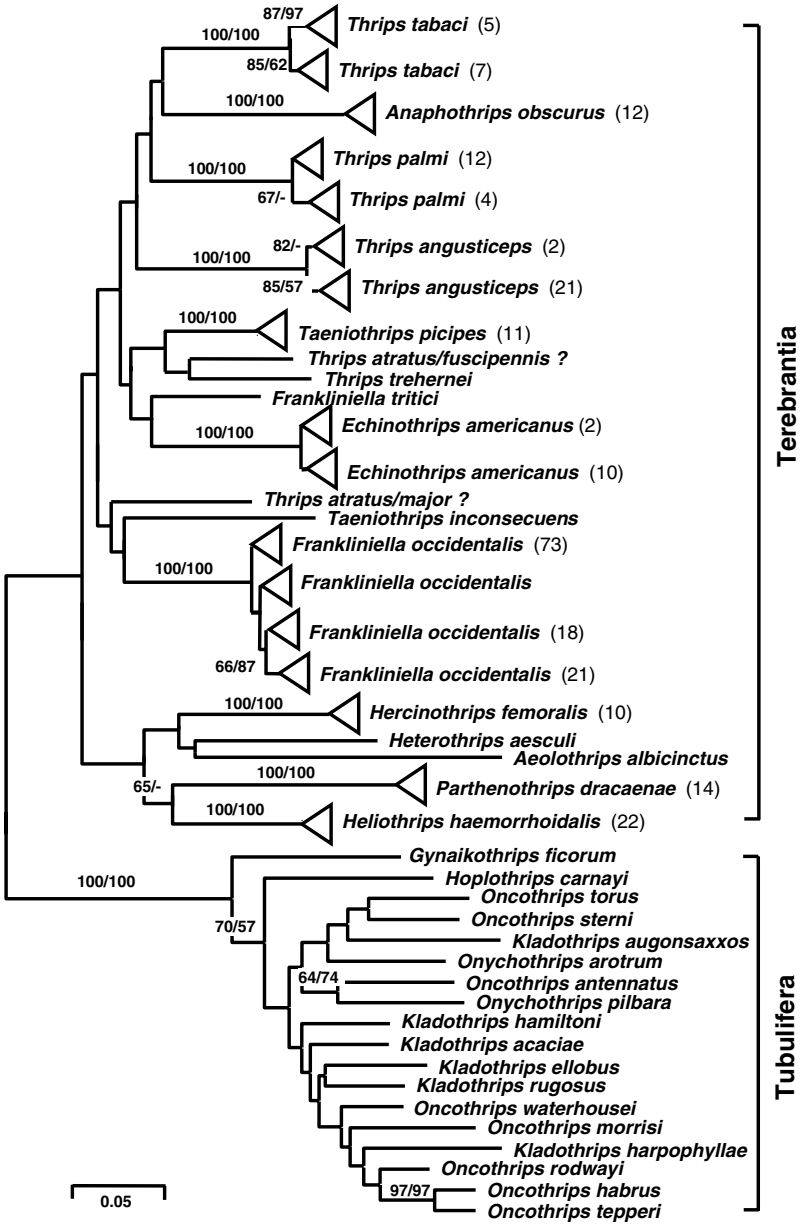

Figure 4 Neighbour-joining dendrogram showing genetic relationship based on COI sequences among 33 morphologically identified and two ambiguous (indicated with question marks) thrips species. For clarity, individuals with identical haplotypes are collapsed and represented as triangles. Numbers on branches are bootstrap values > 50% obtained with Kimura-2-parameter distance followed by values obtained with parsimony analysis (1000 replications). Samples sizes are indicated in brackets.

direct sequencing technique used here), low frequency errors at a particular position are far outnumbered by other templates that have the correct nucleotide. Our results corroborate this in that no 'aberrant' haplotypes among all assessed individuals was observed. For example, among the 132 *F. occidentalis* examined, at least 18 individuals shared the same haplotype. This strongly suggests that differences among haplotypes of the same species – although small – are real and not caused by PCR errors. It is very unlikely that the same error at the same nucleotide position would have occurred independently in many individuals.

A more serious but less common problem when using mtDNA as a marker is the potential to amplify a nuclear insertion of a mtDNA sequence (i.e. a pseudogene; Lopez *et al*., 1994). The insertion and duplication of mtDNA sequences into the nuclear genome can be an aid (Zischler *et al*., 1995) or a hindrance (Collura & Stewart, 1995) to phylogenetic analysis. If unrecognized as a pseudogene, these paralogous sequences could generate misleading phylogenies depending on the evolutionary timing of the insertion event. Pseudogenes have been described for insects (Zhang & Hewitt, 1996; Bensasson *et al*., 2000; references therein). To date, however, no mtDNA pseudogenes have been reported in the nuclear genome of Thysanoptera. Direct sequencing of the fragment used in this study produced nucleotide sequences with a perfect reading frame (i.e. no STOP-codons; Fig. 1). All variations (except for position 379 in *F. occidentalis*) were in the form of 3rd base substitutions and therefore are silent substitutions not changing the resulting amino acids. This strongly suggests

that the sequenced fragment was indeed the mitochondrial coding COI gene, as a non-coding nuclear pseudogene is expected to mutate randomly.

Tree-based approach

The phylogenetic (or tree-based) approach for species identification used in this study hinges on several assumptions and has its limitations. These difficulties are common to the use of molecular systematics for taxonomic classification (e.g. Avise, 1989; Davis & Nixon, 1992).

First, a particular DNA sequence has to be amplified from a test product without artefacts (see Discussion above) and has to be of sufficient length (i.e. contain enough information) for phylogenetic reconstruction. Bootstrap values on the NJ-tree (Fig. 4) indicate that the COI fragment analysed contains enough information for species identification. All conspecific individuals unambiguously cluster together with maximum bootstrap support. The analysis also supports the deep phylogenetic split between Terebrantia and Tubulifera, and corroborates the sister-taxon relationship of these two probably monophyletic suborders (Crespi *et al*., 1996). However, variation in the COI gene obviously is too 'unspecific' (i.e. too low or too high and, thus, blurring the true phylogenetic signal) to resolve relationships among species. Hence, for species identification (the purpose of this study) COI is an ideal molecular marker, but for phylogenetic studies other markers or a combination of COI with other markers should be considered (see also Crespi *et al*., 1998).

Second, a critical consideration in using molecular systematics for species identification is the assumption that the taxonomy of the group in question is complete (e.g. Brunner *et al*., 2001). If this basic biological information is lacking, questions about the adequacy of type sequences (i.e. species-specific sequences) and the possibility of paraphyletic relationships among recognized species cannot be answered. We expect that this problem will become more and more apparent in thrips studies with the advance of molecular approaches. For example, Crespi *et al*. (1998) examined two Australian gall thrips species using sequences of the COI gene. They found that each species apparently represented a pair of sibling species. This observation indicates that an absence of distinguishing morphological traits (i.e. apparent morphological identity) should not be taken uncritically. Sequence divergence levels of 8.7 and 15.9% within these two pairs of sibling species are well within the range of between-species divergence found in this study.

Finally, an accurate phylogenetic reconstruction of DNA sequences can be misinterpreted if the database for type sequences is incomplete. This is because of the hierarchic structuring of trees, a major disadvantage compared to a 3D graphic (see Discussion above). For example, two test sequences of thrips that could not be identified unambiguously on the basis of morphological characters were added to the NJ analysis. One thrips, identified as being either *Thrips atratus* or *T. fuscipennis*, clustered with *T. trehernei* (Fig. 4). Thus, in the absence of other information (e.g. bootstrap values or additional individuals of the species in question), the branching order of this tree suggests that this test sequence is from the same species as 'type sequence *T. trehernei*'. In this case, the inclusion of type sequences from *T. atratus* and *T. fuscipennis* would alter this conclusion.

It should be noted, however, that even without species *T. atratus* and *T. fuscipennis,* it is possible to conclude that the two test sequences are not closely related and, hence, are not the same species (i.e. *T. atratus*). In many cases, this alone may be an important conclusion. If there is doubt about the completeness of the database or the taxonomy of a group, a conservative strategy applied in this study is to identify species only when a test sequence groups within a set of type sequences (i.e. sequences obtained from several individuals collected from different locations) from a particular species.

Conclusions

PCR amplification techniques have been used successfully to identify a variety of organisms and pathogens. The ability to screen crude material insufficient for morphological identification (e.g. parts of animals or single eggs from aphid species; Frey *et al*., unpublished) is crucial for plant quarantine diagnosis and treatments against pathogens or pests that are very specific. However, the development, accuracy and effective use of PCR-based diagnostic markers is dependent upon an understanding of the assumptions and limitations of the techniques used to generate the markers and the use of appropriate controls to test them.

As demonstrated in this study, restriction analysis of PCR-amplified mtDNA is a relatively simple, and still regularly used technique that might be employed in species identification. However, only a fraction (i.e. the cutting sites) of the information present in the amplified DNA fragment is assessed by RFLP. For example, a RFLP analysis with *Sau*3AI cannot distinguish between *P. dracaenae* and *T. picipes* (Table 3), although their sequence divergence is 22.9% (Table 2). Recent authors have generally debated the accuracy and role of RFLP analyses in an area of increasing access to DNA sequence data (Lamb *et al*., 1994; Walker *et al*., 1995; Grant *et al*., 1998). In contrast, direct sequencing makes use of the maximal information content (individual nucleotide sites) and should therefore be preferred whenever possible.

Polymorphism, or variation within species, is common in all kinds of data. However, polymorphism is often ignored by systematists and comparative biologists. Polymorphism may have a profound impact on phylogeny reconstruction or species delimitation. This study demonstrated conclusively that genetic polymorphism can easily be incorporated in a molecular-based identification key. Furthermore, by taking into account within-species polymorphism (i.e. adding as many type sequences as possible) the robustness of the key will even be enhanced and assignment of test sequences will be more meaningful (see example above). We therefore plan to extend our molecular identification key by including more species and we will place the restriction maps on the web. This will allow the use of the key in laboratories that do not dispose of sequencing capabilities.

In conclusion, molecular markers provide powerful tools for species identification and are a valuable alternative to traditional morphological methods. Molecular methods can be used with reliability and accuracy provided that they undergo appropriate development and testing. Thus, a general approach to the development of a species diagnostic key should involve three major steps: (1) evaluation of potential target DNA sequences and appropriate diagnostic molecular markers and techniques; (2) establishment of diagnostic patterns and adaptation of the method to accommodate all species in question; (3) quality assessment (i.e. evaluating the robustness) of the molecular key with regard to within- and among-species polymorphism.

Acknowledgements

We would like to thank Beatrice Frey, Ronald Hunter and Franz Schwaller for technical assistance. Two anonymous referees and Allan Watt improved the manuscript with careful reviews. We are also grateful to Laurent Excoffier and David Paetkau for valuable comments on species definitions and assignment tests. We appreciate additional insights into thrips biology by Carolien Zijlstra and suggestions on the use of the program ProSeq by Dimitry Filatov. This study resulted from COST action 823 and was supported by grant BBW/OFES-NB. C98.0026 to J.E.F.

References

Avise, J.C. (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, **43**, 1192–1208.

Baker, C.S. & Palumbi, S.R. (1995) Population structure, molecular systematics and forensic identification of whales and dolphins. *Conservation Genetics: Case Histories from Nature* (ed. by J. Avise and J. L. Hamrick), pp. 10–49. Chapman & Hall, New York.

Bensasson, D., Zhang, D.-X. & Hewitt, G.M. (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genome. *Molecular Biology and Evolution*, **17**, 406–415.

Brunner, P.C., Douglas, M.R., Osinov, A., Wilson, C.C. & Bernatchez, L. (2001) Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution*, **55**, 573–586.

Collura, R.V. & Stewart, C.-B. (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature*, **378**, 485–489.

Crespi, B.J., Carmean, D.A., Mound, L.A., Worobey, M. & Morris, D. (1998) Phylogenetics of social behavior in Australian gall-forming thrips: evidence from mitochondrial DNA sequences, adult morphology and behavior, and gall morphology. *Molecular Phylogenetics and Evolution*, **9**, 163–180.

Crespi, B., Carmean, D., Vawter, L. & von Dohlen, C. (1996) Molecular phylogenetics of Thysanoptera. *Systematic Entomology*, **21**, 79–87.

Davis, J.I. & Nixon, K.C. (1992) Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology*, **41**, 421–435.

Faith, D.P. (1991) Cladistic permutation tests for monophyly and nonmonophyly. *Systematic Zoology*, **40**, 366–375.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Frey, J.E. & Frey, B. (1995) Molecular identification of six species of scales (*Quadraspidiotus* spp.) by RAPD-PCR: assessing the field-specificity of pheromone traps. *Molecular Ecology*, **4**, 777–780.

Grant, W.S., Clark, A.-M. & Bowen, B.W. (1998) Why restriction fragment length polymorphism analysis of mitochondrial DNA failed to resolve sardine (*Sardinops*) biogeography: insights from mitochondrial DNA cytochrome *b* sequences. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 2539–2547.

Hide, G. & Tait, A. (1991) The molecular epidemiology of parasites. *Experientia*, **47**, 128–142.

Hillis, D.M., Huelsenbeck, J.P. & Cunningham, C.W. (1994) Application and accuracy of molecular phylogenies. *Science*, **264**, 671–677.

Innis, M.A., Gelfand, D.H., Sninsky, J.J. & White, T.J. (1990) *PCR Protocols A Guide to Methods and Applications*. Academic Press, Inc, New York.

Kawasaki, E.S. (1990) Sample preparation from blood, cells, and other fluids. *PCR Protocols: A Guide to Methods and Applications* (ed. by M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. L. White), pp. 146–152. Academic Press, San Diego.

Keohavang, P. & Thilly, W.G. (1989) Fidelity of DNA polymerases in DNA amplification. *Proceeding of the National Academy of Sciences USA*, **86**, 9253–9257.

Kumar, S., Tamura, K. & Nei, M. (1994) MEGA: Molecular evolutionary genetics analysis software for microcomputers. *Computer Applications in the Biosciences*, **10**, 189–191.

Lamb, T., Lydeard, C., Walker, R.B. & Gibbons, J.W. (1994) Molecular systematics of map turtles (*Graptemys*): a comparison of mitochondrial restriction site versus sequence data. *Systematic Biology*, **43**, 543–559.

Lewis, T. (1997) Flight and dispersal. *Thrips as Crop Pests* (ed. by T. Lewis), pp. 175–196. Cambridge University Press, Cambridge.

Loomans, A.J.M. & Heijboer, A. (1999) *Franklinothrips vespiformis* (Crawford): a polyphagous predator preying on thrips eggs. *Proceedings of Experimental and Applied Entomology, the Netherlands Entomological Society (N E V)*, **10**, 143–150.

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W. & O'Brian, S.J. (1994) *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, **39**, 174–190.

Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M. & Tajima, F. (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, **105**, 207–217.

Palmer, J.M. & Mound, L.A. (1991) Thysanoptera. *The Armored Scale Insects, Their Biology, Natural Enemies and Control* (ed. by D. Rosen), pp. 67–76. Elsevier, Amsterdam.

Palumbi, S.R. (1995) Nucleic acids II: The polymerase chain reaction. *Molecular Systematics* (ed. by D. Hillis, C. Moritz & B. K. M. Able), pp. 205–248. Sinauer Associates, Inc, Sunderland, MA.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988) Primerdirected enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighing, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Swofford, D.L. (1999) *PAUP: Phylogenetic Analysis Using Parsimony* 4.0b. Sinauer Associates, Sunderland, MA.

Thomson, J.D., Gibons, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Vago, A.R., Macedo, A.M., Oliveira, R.P., Andrade, L.O., Chiari, E., Galvão, L.M.C., Reis, D.A., Pereira, M.E.S., Simpson, A.J.G., Tostes, S. & Pena, S.D.J. (1996) Kinetoplast DNA signatures of *Trypanosoma cruzi* strains obtained directly from infected tissue. *American Journal of Pathology*, **149**, 2153–2159.

Vierbergen, G. (1995) International movement, detection and quarantine of Thysanoptera pests. *Thrips Biology and Management* (ed. by B. L. Parker, M. Skinner and T. Lewis), pp. 119–132. Plenum Press, New York.

Walker, D., Burke, V.J., Barak, I. & Avise, J.C. (1995) A comparison of mtDNA restriction site vs. control region sequences in phylogeographic assessment of the musk turtle (*Sternotherus minor*). *Molecular Ecology*, **4**, 365–373.

Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Zhang, D.-X. & Hewitt, G.M. (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution*, **11**, 247–251.

Zischler, H., Geisert, H., von-Haeseler, A. & Pääbo, S. (1995) A nuclear 'fossil' of the mitochondrial d-loop and the origins of humans. *Nature*, **378**, 489–492.

Accepted 8 December 2001