Hereditas 140: 92–98 (2004)

Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration?

JUERG E. FREY and BEATRICE FREY

Agroscope FAW Waedenswil, Swiss Federal Research Station for Horticulture, Department of Horticulture, Molecular Diagnostics Laboratory, Waedenswil, Switzerland

Frey, J. E. and Frey B. 2004. Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? — *Hereditas 140*: 92–98. Lund, Sweden. ISSN 0018-0661. Received June 13, 2003. Accepted November 14, 2003

The mitochondrial genome is increasingly being used as a species diagnostic marker in insects. Typically, genomic DNA is PCR amplified and then analysed by restriction analyses or sequencing. This analysis system may cause some serious problems for molecular diagnosis. Besides the errors introduced by the PCR process, mtDNA sequence variation of amplified fragments may originate from mtDNA heteroplasmy or from nuclear integrations of mtDNA fragments, both of which have been shown to occur in insects.

Here we document abundant variation in PCR-amplified sequences of the mitochondrial cytochrome oxidase I gene of *Thrips tabaci*. We confirm that the most common haplotype is of mitochondrial origin. Some of the observed mutations were introduced by the amplification process. However, the occurrence of some haplotypes at elevated frequencies indicates that within-individual variation of the respective fragment exists at low levels in *T. tabaci*. The frequencies of these sequences are too low to negatively affect mtDNA-based molecular diagnosis of *T. tabaci*. The possible origin of these variant haplotypes is discussed.

Juerg E. Frey, Agroscope FAW Waedenswil, Swiss Federal Research Station for Horticulture, Department of Crop Protection, Molecular Diagnostics Laboratory, Schloss, Lab. 4, CH-8820 Waedenswil, Switzerland. E-mail: juerg.frey@faw.admin.ch

The mitochondrial genome of insects is a circular DNA molecule of ca 18.5 kb in size (Hoy 1994). In general, each mitochondrion is believed to contain between 2–10 copies of the genome. Single cells contain up to several hundred mitochondria and consequently, several thousand copies of mtDNA genes may occur in one cell (Lightowlers et al. 1997; Scheffler 2000). This provides optimal conditions for PCR-based analyses and is an important reason for the attractiveness of mtDNA in molecular diagnostics.

The mitochondrial cytochrome oxidase I (COI) gene was shown to be exceptionally well suited for species discrimination in the order Thysanoptera (Brunner et al. 2002). The degree of conservation in this gene confers low within-species variation coupled with appreciable variation between species which allows highly accurate species identification. We have recently expanded this data set to over 70 species including Lepidoptera, Diptera, Homoptera, Nematodes and Mammalia, all of which can unambiguously be identified to species level using this mtDNA gene fragment (data not shown). The data so far established are based on PCR amplified mtDNA sequences. This approach may cause problems due to uncertainty of the origin of the amplified sequences. Nuclear integration of CO I gene fragments has been observed in vertebrates as well as in arthropods such as aphids and crickets (Zhang and Hewitt 1996; Parfait et al. 1998; Bensasson et al. 2001). Such fragments may co-amplify with the mitochondrial target sequence or even be amplified instead of it (Zhang and Hewitt 1996) and hence negatively affect the quality of the molecular identification. Another source of potential ambiguity in the results of PCR-amplified mtDNA diagnosis may be heteroplasmy, i.e. the occurrence of more than one haplotype within a single organism (Petri et al. 1996; Thomas et al. 1998). For example, Boyce et al. (1989) report several heteroplasmic PCR-product size classes within each individual of bark beetles.

In the process of developing a molecular identification key for thrips species, we encountered within-individual variation in the restriction patterns of PCR-amplified mtDNA fragments of COI. We assessed this mtDNA variation using a cloning approach. We confirm that the most common haplotype is of mitochondrial origin and that haplotype variants can be observed at low frequencies and we discuss their potential origin.

MATERIAL AND METHODS

Thrips tabaci specimens were collected from our laboratory culture that is being reared on onion

shoots. Frozen individuals were homogenized in 100 µl extraction buffer (Dilworth and Frey 2000). A 440 base pair fragment of the mitochondrial cytochrome oxidase I gene was amplified with one µl of the homogenate in a PCR reaction containing 10 µl of HotStarTaq Amplification Mix (Qiagen) and 0.3 µM of each primer (C1-J-1751, GGATCACCTGATATAGCATTCCC; and C1-N-2191, CCCGGTAAAATTAAAATATAAACTTC, respectively; Simon et al. 1994) in a 20 µl reaction. The reaction was performed in a Perkin-Elmer 9600 cycler using 15 min initial denaturation, followed by 40 cycles of 40 s denaturation at 95°C, 40 s annealing at 52°C, and 40 s extension at 72°C, and finalized by 7 min at 72°C.

For species identification and assessment of the occurrence of amplification fragment variation, the amplified PCR product was digested for 2 h at 37°C with 1.7 U of *Alu*I (Roche, Basel) and electrophoresed on a 25% Long Ranger hydrolink gel (BioConcept, Allschwil, Switzerland) in 0.6 × TBE buffer (Tris-Borate-EDTA; Sambrook et al. 1989) at 60 V for 2 h. The restriction fragments were visualized and photographed under UV in a MultiImage Light Cabinet (Alpha Innotech Corporation).

For the analysis of mtDNA variation, the amplification products of individual thrips were cleaned with Qiagen PCR purification columns, quantified on a Eppendorf Biophotometer and cloned using the TA-Cloning® Kit (Invitrogen) according to the manufacturer's instructions. White (positive) colonies were picked, placed in 100 µl of double distilled water and boiled for 10 min at 95°C to prepare PCR-ready DNA. One µl of this solution was used to amplify the cloned fragments using the same parameters as above. The amplification products were again cleaned (Qiagen PCR cleanup kit) and 35 ng of the amplified fragment, 2 µl BigDye™ (Applied Biosystems), 0.2 µl of primer in a total of 8 µl was used for the sequencing reactions. All fragments were sequenced at least twice using both the forward and the reverse primer.

The error rate introduced by the Taq-polymerase was assessed in a system check, re-cloning one of the PCR products obtained by the methods outlined above that was previously cloned and sequenced. The DNA of these colonies was extracted as described above and sequenced using the same BigDye dye-terminator kit. Sequence alignments were done using the Sequencher™ software (V. 3.1, Gene Codes Corporation).

To establish a mitochondrial origin of the consensus sequence, RNA was extracted from a pool of 50 *T. tabaci* individuals using the Rneasy Mini Kit (Qiagen, Basel, Switzerland) and Dnase treated with the Rnase-free Dnase set (Qiagen, Basel, Switzerland) according to the manufacturers recommendations. The RNA was then reverse-transcribed using the Omniscript™ Reverse Transcriptase kit (Qiagen, Basel, Switzerland) and an oligo-dT primer. Finally, the resulting cDNA was amplified using the high-fidelity ProofStart DNA Polymerase (Qiagen, Basel, Switzerland).

RESULTS

The PCR amplifications always produced only a single band of 480 bp suggesting a single genetic origin of the template sequence. Digestion of this fragment with the restriction enzyme *Alu*I produced patterns suggesting two restriction sites. Sequening confirmed this assumption, as we found the first *Alu*I restriction site at position 283, the second at position 322, respectively. However, the restriction patterns often suggested the co-occurrence of 2 different haplotypes, i.e. one with a single restriction site at position 322, the other one with two restriction sites as described above (Fig. 1). Such patterns may sometimes be caused by incomplete digestion, an explanation that seems to fit here as almost all individuals show undigested amplification fragments (Fig. 1). However, in individual number one, there is no indication of undigested amplification product while it still shows the pattern expected from a haplotype mixture. Although the haplotype lacking the second restriction site was not found in the cloned fragments, it was present in another study on thrips

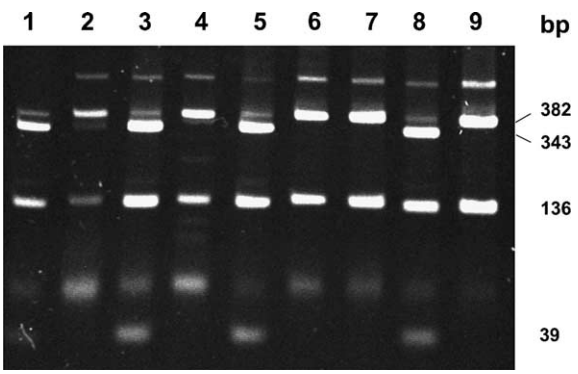

Fig. 1. *Alu*I restriction patterns of a PCR-amplified fragment (518 base pairs total length) of the mitochondrial cytochrome oxidase I gene for nine different *Thrips tabaci* individuals showing evidence of within-individual mtDNA-variation with respect to two restriction sites. One restriction site is located at position 382 and produces two fragments of 382 and 136 base pairs length; the second restriction site is located at position 343 and occurs together with restriction site 382, thus producing three fragments of 343, 136 and 39 base pairs. Co-occurrence of fragments 382 and 343 in individuals 1, 2, 3, 5 and 8 indicates heteroplasmy and/or nuclear pseudogenes.

species conducted in our laboratory (GenBank accession number AF378692) suggesting that this haplotype occurs in low frequency in the heteroplasmic state in some thrips individuals.

To assess the occurrence of mutations, a total of 64 clones from four individual thrips was sequenced on both strands. In the 27459 base pairs sequenced we found a total of 33 point mutations in twenty-three different haplotypes (Table 1a). Two haplotypes were found at a frequency of over 64% (Table 2). They differed at two synonymous positions (Table 1a; GenBank accession nrs. AF395332 and AF395333, respectively). The base composition of the consensus sequence of the two most common haplotypes is 29.0% A, 37.1% T, 19.0% C, and 14.8% G. This bias towards high AT content is typical for mitochondrial sequences (Tamura 1992; Zhang and Hewitt 1996; Honda et al. 2000). The consensus sequence aligned well with the corresponding sequences from other thrips species sequenced in our laboratory as well as with sequences from GenBank (Brunner et al. 2002). Only 11 synonymous differences were found in a comparison to published haplotypes of the corresponding mtDNA sequence from *T. tabaci* (GenBank accession numbers AF378692 and AF378693, respectively), corresponding to 2.54% nucleotide difference. Together, this provides good evidence that the consensus sequence is in fact of mitochondrial origin (Sunnucks and Hales 1996). To obtain further evidence for a mitochondrial origin of the consensus sequence, we performed a PCR-amplification on the cDNA of reverse-transcribed mRNA from a pool of 50 *T. tabaci* using the same primers as for direct amplifications. The DNA-sequences from this PCR fragment were 100% homologous to the consensus sequence of the DNA-based direct amplifications (data not shown).

Because no Taq polymerase works totally error-free, some of the observed mutations may have been introduced by this enzyme. It was thus important to establish how many of the mutant haplotypes were caused by sequencing errors and how many have a real genetic basis. According to the manufacturer (Qiagen), the error-rate of the HotStar Taq is ca 4×10^{-5} or 4 errors in 100 000 base pairs. To be sure that no other error sources, such as the process of in vitro amplification in the cloning step, further increase this number, we decided to perform an error test by re-cloning a cloned and sequenced fragment of a fifth *T. tabaci* individual. The haplotype of this individual was different at one position from the two previously reported ones (Table 1a and b; GenBank accession nrs. AF395332 – AF395334). We sequenced 19460 bp of 46 clones from this error test (Table 1b) and found 16 mutations in 14 different haplotypes. The most common haplotype of the error test (63% of all clones; Table 2) was identical to that obtained after direct sequencing of the PCR product. The error-rate in the error test turned out to be slightly lower than expected, i.e. 2.06×10^{-5}.

We found distinctly yet not significantly more mutations in the four thrips individuals as compared to those of the thrips clones of the error test, i.e. 33 mutations in a total of 64 thrips clones corresponding to 51.6% vs 16 mutations in 46 clones of the error test corresponding to 34.8% ($\chi^2 = 0.843$, 1 df, $P = 0.359$; Table 1, 2).

DISCUSSION

In a previous study we established a DNA-based species identification key for Thysanoptera using a 440 base pair, PCR amplified fragment of the mitochondrial cytochrome oxidase I gene (Brunner et al. 2002), and we showed that two different haplotypes occur in *T. tabaci* with respect to this fragment. To simplify the use of this key, we also analysed the patterns of several restriction enzymes. The present study was motivated by the observation of within-individual variation in *Alu*I restriction analyses of PCR-amplified mitochondrial cytochrome oxidase I fragments as shown in Fig. 1. Here we describe the within-individual genetic variation in this PCR-amplified fragment in *T. tabaci* and discuss its possible origin.

Variation in PCR-amplified fragments can be of diverse origin. It may be caused by mitochondrial heteroplasmy, by nuclear integration of mitochondrial sequences, or by simple Taq polymerase errors. Heteroplasmy occurs either because there is more than one haplotype within a single mitochondrion or, alternatively, because different mitochondria belong to different haplotype groups that are distributed within or between cells (De Stordeur 1997; Lightowlers et al. 1997). Nuclear copies of CO I were described from many organisms including insects (Jacobs and Grimes 1986; Sunnucks and Hales 1996; Bensasson et al. 2001). Because mitochondrial genes far outnumber single copy nuclear genes, such nuclear copies should generally represent no threat to regular PCR diagnostics. In some cases, however, multiple nuclear copies of mitochondrial genes have been observed and therefore, this possibility cannot be a priori discarded (Gellissen and Michaelis 1987; Sunnucks and Hales 1996; Bensasson et al. 2001). Taq polymerase errors may be single base changes, insertions or deletions. Because these errors generally occur only isolated and in low portions of the

Table 1. *Haplotype table with all polymorphic sites and the position in base pairs from the start of the consensus sequence. Lower case: synonymous mutation; upper case: non-synonymous mutation.*

Position rel. to consensus		14	16	21	31	34	57	84	97	105	106	112	121	139	153	156	169	171	192	210	216	222	223	233	239	246	271	281	329	330	341	343	344	345	349	351	364	366	370	372	377	378	382	394	396	399	403	405	416	424
Consensus	Haplo-type No.	A	A	T	A	C	T	C	G	C	T	R	T	A	T	C	A	G	C	C	T	T	T	A	T	A	R	A	A	C	A	T	G	A	A	A	T	T	T	A	A	G	G	T	T	T	T	A	T	T
a) variant haplotypes from the four individual T. tabaci females B1, G5, IA and SW, respectively.																																																		
mB1c#05	1	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	G	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mB1c#19	2	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	G
mB1c#20	3	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	:	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mB1c#22	4	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	:	-
mB1c#24	5	-	-	-	g	t	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mB1c#26	6	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mG5c#05	7	-	g	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-
mG5c#12, mG5c#19	8	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-
mG5c#14	9	G	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mG5c#18	10	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	G	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mG5c#20	11	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	c	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mG5c#22	12	-	-	-	-	-	C	-	-	-	-	g	-	-	-	-	-	-	t	T	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mIAc#02	13	-	-	C	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	C	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mIAc#05	14	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-
mIAc#06	15	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-
mSWc#06	16	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	c	-	-	-	-	-	-	-	-	-	-	-	-	-
mSWc#11	17	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mSWc#17	18	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mSWc#18, mSWc#26	19	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	t	-	-	-	-	-	-	-	-	-	a	-	G	-	-	-	-	-	-	-	-	:	-	-	-	-	-	-	-	-	-	-	-
mSWc#20	20	-	-	-	-	-	-	-	-	-	c	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
mSWc#22	21	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	-	-	-
mSWc#46, mSWc#49, mSWc#50	22	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-
b) variant haplotypes from the re-cloned sequence of the error test (? = not sequenced).																																																		
cm0611#02	23	-	-	-	-	-	-	T	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-
cm0611#08	24	-	-	-	-	-	-	-	-	-	-	a	-	-	C	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cm0611#10	25	-	-	-	-	-	-	-	-	T	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	t	-	-	-	-	-	-	-
cm0612#14	26	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cm0612#20	27	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	A	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cm0612#23	28	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	g	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cm0612#24	29	-	-	-	-	-	-	-	-	-	-	a	-	g	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cm0612#28	30	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	:
cm0612#29	31	-	-	-	-	-	-	-	-	-	-	a	-	-	-	T	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cm0612#31	32	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	-	-	-	-	-	-
cm0613#33	33	-	-	-	-	-	-	-	a	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	?
cm0613#33	34	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cmA3/C#5	35	-	-	-	-	-	-	-	-	-	-	a	c	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
cmA3/C#8	36	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	C	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 2. *Number of clones sequenced and recovered haplotypes for the four T. tabaci individuals and the error test.*

Individual	No. of clones sequenced	No. of different haplotypes	Frequency of most common haplotype (%)
B1	16	7	62.5
G5	23	8	69.6
SW	18	9	55.6
IA	7	4	57.1
All *T.tabaci* clones	64	22	67.2
Clones of error test	46	15	69.6

amplified fragments, it is generally assumed that strand-slippage and Taq polymerase errors are not a serious threat to accurate PCR-based molecular diagnostics (Gyllensten 1989). We exemplify this based on our analysis system below.

We found three different haplotypes in more than one clone: haplotype 8 and haplotype 19 were each found twice, haplotype 22 even three times among the clones of the respective individual (Table 1a). This corresponds to 8.7% (haplotype 8), 11.1% (haplotype 19) and 16.7% (haplotype 22) of the clones and thus of the population of haplotypes within the respective individuals. Haplotypes that are found repeatedly among a limited number of clones of individual thrips must have originated from the template DNA rather than from errors in the PCR process. We used at least 1 ng thrips DNA per PCR reaction. DNA content per cell is in general between 2–10 pg and very rarely exceeds 20 pg in arthropods (Kirby 1992). The amount of DNA in insect cells depends on ploidy levels of different tissues and tends to be between 0.1 pg to 2 pg, of which 20–80% may be repetitive DNA (Hoy 1994). Assuming an average of 2 pg DNA per cell, at least 500 cells with at least 1000 copies of nuclear DNA and at least 100 times more of mtDNA are available as PCR templates in our reactions. Even if the Taq polymerase introduces an error in the first cycle of PCR and if this error is exponentially amplified together with the "normal" templates, the relative frequency of this "haplotype" among the PCR products would be a maximum of 1 in 1000. The probability of finding such a haplotype twice and three times among 18 clones, as in individual SW, is thus 3.2×10^{-13}. Therefore, Taq polymerase errors can be excluded as the source for the elevated frequency variants, as suggested by Gyllensten (1989).

If Taq polymerase errors can be excluded, the observed elevated frequency haplotypes must originate either from mitochondrial heteroplasmy or from nuclear integrated gene fragment copies. Alternatively, a combination of these is possible. Each of the three haplotypes found at elevated frequencies (i.e. 8, 19, 22) contains a non-synonymous mutation. In addition, haplotype 19 contains a synonymous mutation and a single-base deletion that leads to a stop codon few base pairs downstream. Non-synonymous mutations are generally thought to be rare in mitochondrial sequences (Sunnucks and Hales 1996). In fact, among 180 individuals of the thrips species *Frankliniella occidentalis* (Thysanoptera: Thripidae), we only found one individual with a non-synonymous mutation in a sequenced region of the CO I gene (Brunner et al. 2002). However, this result was obtained with direct sequencing of a PCR product. In polyploid genomes, such as mtDNA, it is not possible to detect haplotype frequencies of 17% (the frequency of the most common variant haplotype in one individual in this study) with this technique. Resorting to cloning does not help much as one would have to sequence at least 125 clones to detect, with 95% reliability, this haplotype frequency. Such large numbers are usually not sequenced after cloning and this may partly explain the low incidence with which mtDNA heteroplasmy is reported. Furthermore, mtDNA gene copy numbers per cell outnumber the nuclear gene copy numbers by a factor of at least 100–1000 (Lightowlers et al. 1997; Scheffler 2000). Thus, unless nuclear introgressed mtDNA fragments occur in large copy numbers in the nuclear genome, they will not be detected. Therefore, if the observed mtDNA variants are of nuclear origin, they have to occur in large copy numbers.

Nuclear copies of mitochondrial genes mostly behave as pseudogenes as they are in general not correctly expressed due to differences in the genetic code (Zhang and Hewitt 1996). Because they can freely accumulate mutations they are expected to eventually diverge from the original mitochondrial sequences and to incorporate nonsense mutations that would cause stop codons in coding DNA. As mentioned above, haplotype 19 contains one synonymous and one non-synonymous mutation and a one base-pair deletion leading to a stop codon. This pattern nicely fits the expectation for nuclear integrated

mtDNA sequences and thus suggests that haplotype 19 may indeed represent a nuclear pseudogene occurring in large copy numbers.

In contrast, the other two haplotypes (8, 22) have only one base pair difference from the mitochondrial consensus sequence (Table 1a). This alone does not argue against the hypothesis of nuclear integration. A low level of nucleotide divergence between the mitochondrial COI sequence and its nuclear homologue was reported from aphids (SUNNUCKS and HALES 1996). Furthermore, multiple nuclear copies of this gene were reported from aphids (SUNNUCKS and HALES 1996) and grasshoppers (BENSASSON et al. 2000), and up to several hundreds of copies of other mitochondrial sequences were found in the nuclear genome of locusts (GELLISSEN and MICHAELIS 1987; BENSASSON et al. 2000). Haplotypes 8 and 22 could thus also originate from several hundred copies of a nuclear integrated COI fragment. However, the presence of two "frequent" haplotypes (haplotypes 19 and 22) in one of the thrips individuals (SW) would then indicate two different incidences of high copy number nuclear integration of the same gene fragment. In addition, as one of the fragments has three times as many mutations as the other, one would have to postulate that the two sequences were introduced to the nucleus at two different evolutionary time points, leaving the older of them time for accumulation of more mutations. In the case of the individual SW we would have to postulate that haplotype 19 integrated into the nucleus first and that all three mutations have evolved after the integration event. Haplotype 22 should then be the younger of the nuclear-integrated sequences. The "non-synonymous" mutation would have occurred after nuclear integration because otherwise, a back-mutation would have to be postulated. The most parsimonious explanation seems thus to be that haplotype 19 represents a high-copy nuclear pseudogene, whereas haplotypes 8 and 22 may actually be heteroplasmic variants of mitochondrial DNA.

However, our results do not provide clear evidence as for the origin of the observed mitochondrial sequence variants. Further analyses using allel-specific single-cell PCR and cytological methods such as highly sensitive gold-and silverstaining are needed to allow to directly visualize and quantify the occurrence of heteroplasmy and nuclear introgression in *T. tabaci*. Nevertheless, it is important to note that due to the low frequency of the mutant haplotypes, their presence causes no problems for PCR-based molecular diagnostics of *T. tabaci*.

Acknowledgements – The participation in this project of the 97/00 biotechnology course of the University of Applied Sciences in Waedenswil is greatly appreciated. Monika Pfunder and Patrick Brunner gave valuable comments on an earlier draft of this manuscript.

REFERENCES

Bensasson, D., Zhang, D.-X. and Hewitt, G. M. 2000. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. – Mol. Biol. Evol. 17: 406–415.

Bensasson, D., Zhang, D.-X., Hartl, D. L. et al. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. – Trends Ecol. Evol. 16: 314–321.

Boyce, T. M., Zwick, M. E. and Aquadro, C. F. 1989. Mitochondrial DNA of the bark weevils: size, structure and heteroplasmy. – Genetics 123: 825–836.

Brunner, P. C., Fleming, C. and Frey, J. E. 2002. A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) using direct sequencing and a PCR-RFLP-based approach. – Agric. For. Entomol. 4: 127–136.

Dilworth, E. and Frey, J. E. 2000. A rapid method for high throughput DNA extraction from plant material for PCR amplification. – Plant Mol. Biol. Rep. 18: 61–64. See also Erratum in issue 18: 157.

De Stordeur, E. 1997. Nonrandom partition of mitochondria in heteroplasmic *Drosophila*. – Heredity 79: 615–623.

Gellissen, G. and Michaelis, G. 1987. Gene transfer: mitochondria to nucleus. – Ann. NY Acad. Sci. 503: 391–401.

Gyllensten, U. 1989. Direct sequencing of in vitro amplified DNA. – In: Ehrlich, H. A. (ed.), PCR technology – principles and applications for DNA amplification. Stockton Press, London, p. 45–60.

Honda, M., Ota, H., Kobayashi, M. et al. 2000. Phylogenetic relationships, character evolution, and biogeography of the subfamily Lygosominae (Reptilia: Scincidae) inferred from mitochondrial DNA sequences. – Mol. Phylogenet. Evol. 15: 452–461.

Hoy, M. A. 1994. Insect molecular genetics – an introduction to principles and applications. – Academic Press.

Jacobs, H. T. and Grimes, B. 1986. Complete nucleotide sequences of the nuclear pseudogenes for cytochrome oxidase subunit I and the large mitochondrial ribosomal RNA in the sea urchin *Strongylocentrotus purpuratus*. – J. Mol. Biol. 187: 509–527.

Kirby, L. T. 1992. DNA fingerprinting. – WH Freeman and Company.

Lightowlers, R. N., Chinnery, P. F., Turnbull, D. M. et al. 1997. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. – Trends Genet. 13: 450–455.

Parfait, B., Rustin, P., Munnich, A. et al. 1998. Coamplification of nuclear pseudogenes and assessment of heteroplasmy of mitochondrial DNA mutations. – Biochem. Biophys. Res. Comm. 247: 57–59.

Petri, B., Haeseler, A. V. and Pääbo, S. 1996. Extreme sequence heteroplasmy in bat mitochondrial DNA. – Biol. Chem. 377: 661–667.

Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. Molecular cloning – a laboratory manual (2nd edn). – Cold Spring Harbor Laboratory Press.

Scheffler, I. E. 2000. A century of mitochondrial research: achievements and perspectives. – Mitochondrion 1: 3–31.

Simon, C., Frati, F., Beckenbach, A. et al. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene

sequences and a compilation of conserved polymerase chain reaction primers. – Ann. Entomol. Soc. Am. 87: 651–701.

Sunnucks, P. and Hales, D. F. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). – Mol. Biol. Evol. 13: 510–524.

Tamura, K. 1992. The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA. – Mol. Biol. Evol. 9: 814–825.

Thomas, M. G., Cook, C. E., Miller, K. W. P. et al. 1998. Molecular instability in the COII-tRNALys intergenic region of the human mitochondrial genome: multiple origins of the 9-bp deletion and heteroplasmy for expanded repeats. – Philos. Trans. R. Soc. Lond. B 353: 955–965.

Zhang, D.-X. and Hewitt, G. M. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. – Trends Ecol. Evol. 11: 247–251.