Journal of Entomology

ISSN 1812-5670

Academic Journals Inc.

www.academicjournals.com

Journal of Entomology 4 (3): 167-180, 2007
ISSN 1812-5670

Sequence Analysis of the Ribosomal DNA ITS2 Region for *Phlebotomus papatasi* (Diptera: Psychodidae)

[1]Omar Hamarsheh, [1]Wolfgang Presber, [2]Ziad Abdeen, [2]Samer Sawalha,
[3]Ahmad Al-Lahem and [1]Gabriele Schönian
[1]Institut für Mikrobiologie und Hygiene, Charité Universitätsmedizin Berlin,
Dorotheenstraße 96, 10117 Berlin, Germany
[2]Laboratory of Leishmaniasis, Faculty of Medicine,
Al-Quds University, Jerusalem, Palestine
[3]Department of Zoology, Faculty of Science, University of Aleppo, Aleppo, Syria

Abstract: We investigated the heterogeneity of the ribosomal internal transcribed spacer 2 (ITS2) within 131 individuals of *Phlebotomus papatasi* from 25 geographic localities in 10 countries by sequencing the PCR-amplified copies of this spacer. Sequence heterogeneity was observed in both sexes from all localities. There are many microsatellite sequences that show insertion-deletion polymorphisms. The length of the ITS2 ranged from 340 to 512 bp and the average GC content of the ITS2 was 29.2%. The phylogenetic relationships among the members of *P. papatasi* were inferred by neighbour-joining analysis. Although the relationships among *P. papatasi* populations could not be completely resolved, the phylogenies were in good agreement with previous enzyme electrophoresis studies.

Key words: *Phlebotomus papatasi*, ITS2, rDNA, sand fly

INTRODUCTION

Phlebotomus papatasi (Diptera: Psychodidae, Phlebotominae) is the well-known vector of *Leishmania major*, one of the causative agents of the Old World zoonotic cutaneous leishmaniasis. This sand fly species is responsible for human and animal leishmaniasis in many regions of the world (Killick-Kendrick, 1990). *P. papatasi* has a large geographical distribution; Morocco (Guernaoui *et al*., 2005), Saudi Arabia (Killick-Kendrick *et al*., 1985), Iran (Yaghoobi-Ershadi *et al*., 2005), former U.S.S.R countries (Perfile'v, 1966), Palestine and the Jordan Valley (Sawalha *et al*., 2003; Schlein *et al*., 1982). Despite this, its population structure and the evolutionary relationships among different populations are conflicting and not well understood (Esseghir *et al*., 1997, 2000). A recent study using mitochondrial *cytochrome b* DNA and microsatellite markers (Hamarsheh, unpublished data) provided, however evidence that *P. papatasi* does not represent a single genetically uniform population as previously reported.

The broad utility of ribosomal DNA (rDNA) is because the non-coding spacers evolve faster than the coding regions and because it is present as multiple tandemly repeated copies in the genome (Arnheim, 1983; Hillis and Dixon, 1991). In insects, rDNA consists of tandemly repeated transcriptional units with each unit containing the genes for the 18S, 5.8S and 28S ribosomal RNA (Gerbi, 1985; Hillis and Dixon, 1991). The internal transcribed spacers ITS1 and ITS2 are flanking the 5.8S rRNA gene and separate it from 18S and 28S rRNA genes. An external transcribed spacer at the 5' end of the 18S rRNA gene completes the transcriptional unit. The rDNA arrays vary in copy

Corresponding Author: Omar Hamarsheh, Institut Für Mikrobiologie und Hygiene,
Charité Universitätsmedizin Berlin, Dorotheenstraße 96, 10117 Berlin, Germany
Tel: +4917624089138 Fax: +4930450-524902

number; there are more than 1000 copies in several Drosophila species (Beckingham, 1982). These rDNA repeat units do not evolve independently, but at a relatively homogeneous evolutionary rate within individuals and within the species. Concerted evolution of multigene families within species has resulted in rapidly evolving spacer regions, such as the ITS2 rDNA (Dover and Coen, 1981; Arnheim, 1983; Gerbi, 1985; Tautz *et al*., 1987), including gene conversion and unequal crossing over (Dover, 1982; Seperack *et al*., 1988). While the coding regions of the rDNA tend to become highly conserved in evolution, the spacer regions appear relatively free to diverge, even in closely related organisms. The ITS sequences were used extensively for resolving the evolutionary affiliations at different taxonomic levels, including recently diverged taxa, such as sibling species (Collins and Paskewitz, 1996; Xu and Qu, 1997).

Substantial heterogeneity of rDNA spacers has been observed in different species of sand flies (Depaquit *et al*., 2002, 2000; DiMuccio *et al*., 2000) and in mosquitoes of the genera Aedes and Culex (Black *et al*., 1989; Wesson *et al*., 1992; Miller *et al*., 1996).

In the present molecular study, we present the results of the analysis of the ITS2 sequences of 131 individuals of *P. papatasi* originating from 25 different geographical localities in 10 countries and evaluate the usefulness of using ITS2 as molecular marker for *P. papatasi* sand fly.

MATERIALS AND METHODS

Sand Fly Materials

P. papatasi sand flies were caught with CDC miniature light traps or sometimes on sticky papers, or came from laboratory colonies (Table 1).

DNA Extraction

DNA was extracted from individual flies by phenol-chloroform extraction method with slight modifications. Single male flies and the thorax and anterior abdomen of the females were homogenized

Table 1: Samples, locations and collection information (for field flies) for the 25 *P. papatasi* populations included in this study, blank fields indicate that no information was available

Country	Population	Latitude	Longitude	Altitude	Year	Capture method
Syria	Aleppo	36° 12' 10N	37° 9' 31E	422	2004	Manual aspirator
	Raqah	35° 57' 0N	39° 1' 0E	225	2004	Manual aspirator
	Dayr az Zawr	35° 19' 60N	40° 8' 60E	193	2004	Manual aspirator
Turkey	Salniurfa	37° 9' 4N	38° 47' 34E	477		Colony
	Hamdun	37° 29' 44N	39° 07' 8E	644	1999	Colony
	Nizip	37° 35' 10N	38° 56' 34E	400		Colony
Cyprus	Chisochus	40° 37' 83N	46° 86' 96E	63	2004	CDC trap
	Lapta	35° 20' 15N	33° 10' 45E	190		Colony
Jordan	Jordan Valley					Colony
Israel	Jordan Valley	30° 56' 60N	35° 22' 0E	-339	2003	CDC trap
Palestine	Jericho	31° 52' 0N	35° 27' 0E	-227	2004	CDC trap
	Jordan Valley	31° 52' 0N	35° 27' 0E	-227		Colony
	Jordan Valley				2004	CDC trap
	Qabatiyah	32° 24' 31N	35° 16' 48E	284	2004	CDC trap
	Raba	32° 23' 9N	35° 22' 55E	707	2004	CDC trap
	Dayr al Gousun	32° 21' 9N	35° 4' 28E	234	2004	CDC trap
Egypt	N. Sinai	30° 49' 60N	34° 7' 0E	111		Colony
	Sinai	30° 49' 60N	34° 7' 0E	111		Colony
	Kafr ash shaykh	31° 6' 41N	30° 56' 11E	6		Colony
	Alexandria	31° 11' 53N	29° 55' 9E	4		Colony
	Al-Qalyopia	30° 11' 53N	31° 8' 3E	26		Colony
Italy	Rom	41° 53' 60N	12° 28' 60E	14		Colony
	Rocca Priora	41° 47' 60N	12° 45' 0E	686		Colony
Tunisia	Tunis	36° 48' 10N	10° 10' 47E	0	2004	CDC trap
Morocco	Marrakesh	31° 37' 60N	8° 0' 0W	450	2005	CDC trap

with a sterile glass rod in 1.5 mL microfuge tube containing 250 µL lysis buffer (0.1 M Tris-HCl pH 8, 50 mM EDTA pH 8), 2.5 µL Triton X100 and 5 µL Proteinase K. The homogenate was incubated at 60°C overnight, extracted with equilibrated phenol, phenol: chloroform: iso-amyl alcohol in 25:24:1 volume ratio (Roth, Germany), then extracted with chloroform: iso-amyl alcohol in 24:1 volume ratio. For each extraction step, tubes were mixed gently for 10 min, centrifuged using desktop centrifuge at 13000 g for 10 min, then the upper aqueous phase was transferred to a new 1.5 mL microfuge tube. DNA was precipitated with 250 µL isopropanol and 20 µL 3M sodium acetate at -20°C overnight. DNA was pelleted by centrifugation at 13000 g for 30 min, the supernatant removed and the pellet washed twice with 500 µL cold 70% ethanol. DNA pellets were dried using speed-vac centrifuge (Eppendorf, Germany) for 8 min and then dissolved in 50 µL TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8). DNA was stored at 4°C.

PCR Amplification

Amplifications of the ITS2 were done using the following conditions: an initial denaturation at 95°C for five min; followed by 35 cycles at 94°C for 30 sec, 62°C for 30 sec and 72°C for 1.5 min and a final extension at 72°C for 10 min. The PCR was performed in 50 µL reaction volume containing 50 ng genomic DNA, 10×PCR buffer (Applied Biosystems) 12 pM of each primer C1a: 5' CCTGGTTAGTTTCTTTTCCTCCGCT-3' and JTS3: 5'-CGCAGCTAACTGTGTGAAATC-3' (Depaquit *et al*., 2000), 2.5 mM of each dNTP, 1.5 mM $MgCl_2$ and 0.5 units Ampli*Taq*® DNA polymerase (Applied Biosystems). PCR amplifications were carried out in RoboCycler ® Gradient 40 (Stratagene).

SSCP Analysis and Sequencing

Each specimen was screened for polymorphisms by single strand conformation polymorphism (SSCP) analysis. Electrophoresis for SSCP was performed on gels (0.8 mm thick) which consisted of the following: 112.5 mL ddH_2O, 37.5 mL 0.5×MDE®(Cambrex, Rockland), 9 mL 10×TBE buffer, 0.6 mL 10% ammonium persulfate and 60 µL TEMED. Fifteen µL of PCR product was mixed with 2 µL of denaturing mix (1% SDS, 10 mM EDTA) and 2 µL loading mix (95% formamide, 10 mM EDTA, 0.025% bromophenol blue and 0.025% Xylene cyanol). The tube was shaken gently, heated at 98°C for 13 min in a water bath and immediately placed on ice for at least 10 min. Sixteen µL of the sample was loaded onto the SSCP gel. Electrophoresis was carried out on vertical slab gels (45×35 cm) in a cold room (4°C) using 1×TBE buffer for 16 h at 1 KV, 280 mA and 10 Watts. The gel was silver stained, fixed using 1% Nitric acid for 15 min and stained with 0.2% silver nitrate for 25 min. After washing in deionized water for 15 min the gel was placed in developer (0.28 M Na_2CO_3 and 0.1% formaldehyde) until bands became visible and dried at 70°C for 2 h using a gel drier.

Samples presented unique SSCP banding profiles were sequenced in both directions using automated sequencing techniques at sequencing facility, Department of Genetics, Humboldt University, Berlin.

Data Analysis

Sequences (481-515 bp) were manually edited and aligned using CLUSTAL W (Higgins *et al*., 1994). Gaps were treated as missing data. MEGA version 3.0 (Kumar *et al*., 2004) and DnaSP (Rozas *et al*., 2003) was used for the statistical analysis of sequences. The sequences were analyzed using neighbour-joining (NJ) (Saitou and Nei, 1987) with Kimura's two-parameter correction (Kimura, 1980), as implemented in the TREECON for Windows v. 1.3b (Van de Peer and De Wachter, 1994); gaps were excluded from the analysis. Bootstrapping (Felsenstein, 1985) was performed using NJ with 1000 replicates. The trees were rooted using the ITS2 sequence of *Phlebotomus sergenti* (GenBank: AF218323) as an out group.

RESULTS

Forty three sequences comprising 5.8S gene, the ITS2 and the 5' end of the 28S ribosomal gene, were generated for four individuals from Cyprus, one from Jordan, one from Morocco, 11 from Egypt, six from Italy, six from the Middle East, eight from Syria and six from Turkey. All sequences were determined and positively confirmed as ITS2 with a BLAST search of GenBank. Sequences are available in GenBank under accession numbers (DQ887632-DQ887674).

The size of the amplified fragments varied considerably among studied individuals; from 340 bp in Syr3T157 to 512 bp in Tuk3T14 with an average of 482.6 bp (Table 1). The sequences were aligned along with the ITS2 sequence for *P. papatasi* from Rhuhaibe, Syria which is available in the GenBank under accession number: AF218321. Much of the variation in length was due to insertion /deletion mutations and differences in the numbers of simple sequence repeats which comprised principally simple di-nucleotide repeat insertions. The overall nucleotide composition (A, T, C and G) of the sand fly populations considered here was 36.1, 34.7, 13.5 and 15.7%, respectively. Motifs containing A and T residues outnumbered most of the permutations and constituted much of the elevated AT content.

Sequence Heterogeneity

The alignment of the ITS2 sequences resulted in a total of 563 characters including gaps (Fig. 1). Of the 563 characters, 137 were variable (polymorphic), 179 were invariable (monomorphic) and 75 were parsimony informative sites. No loci diagnostic for populations were detected for ITS2 and sequences were not identical for individuals originated from the same population.

In sequencing PCR products, sequence heterogeneity was noted in most of individuals. Sequencing was repeated in certain individuals 3-4 times, but without resolving the heterogeneity. Almost all sequenced samples show different levels of sequence heterogeneity even if they are originated from the same population. It seems that there is a difficulty in reading some nucleotides in the same position which is probably due to different ITS2 copies. Observations on the sequence alignment reflected an elevated sequence variation in the 3' region compared to the 5' region.

Phylogenetic Analysis

Aligned complete sequences from each individual were analyzed phylogenetically using NJ analysis; gaps between paired alignments were treated as missing data.

A bootstrap consensus neighbour-joining (NJ) tree produced from Kimura's two-parameter is shown in Fig. 2. Most of the phylogenetically critical nodes were not well supported as can be seen by low bootstrap values (only bootstrap values above 50% were shown). Although the phylogenetical relationships were completely unresolved and did not show significant structuring, a strongly supported microclade (100%) grouped two individuals; E1T53 and E1T11, both derived from Egypt.

PCR-RFLP Assays

To determine whether multiple copies of ITS2 may vary within individuals, we subjected the PCR-amplified copies of this spacer to restriction analysis. The RFLP patterns of the amplified ITS2 obtained with the restriction enzymes *Sau961*, *AvaII and RsaI* did not exhibit any variations among the different populations, however the sum of all restricted fragments produced from all restricted enzymes exceed the maximum size of the non-restricted PCR product which probably indicates the presence of many different ITS2 copies (Fig. 3).

```
                                                                                                              1 1111111111 1111111111 1111111111
                    1 1111111112 2222222223 3333333334 4444444445 5555555556 6666666667 7777777778 8888888889 9999999990 0000000001 1111111112 2222222223
           1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
AF218321   CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
CypT20     CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTTT- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
CypT34     CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CAGAATTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
CypT48     CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
CypT6      CGCAGCTATC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAAGATCGA AT--ATTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egt5T141   CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy1T11    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy1T39    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy1T53    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy2T26    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy2T40    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTTT- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy2T54    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy3T41    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy3T69    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy4T28    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Egy4T70    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Ity1T22    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Ity1T36    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTACGGTT
Ity2T37    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACAGGAA AA--TTACTC GGACG--GAT TGC--GGTCC GCGCGGC-GT TTAAGTTTTT GTTTAAG-CT -GCAAGGACC -CGTGTGGTT
Ity3T38    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAAGGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTAGGGTT
Ity3T52    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Ity3T66    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
ILTjv      CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
JorT138    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Pal1T21    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGCCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Pal1T49    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Pal1T63    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Pal3QT160  CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Pal4RT162  CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT TTTTAAA-CT -GCATGGACC ACGTATGGTT
Syr1T15    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Syr1T29    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Syr2T58    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Syr3T157   CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC -TGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Syr3T3     CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Syr3T31    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC -TGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Syr3T45    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
SyrT17     CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GGACGCATAT TGG--GGCCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Tuk1T151   CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Tuk1T18    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTTTT GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Tuk1T32    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTTT- GAACGCATAT TGC--GGTCC -TGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Tuk1T46    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Tuk2T33    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
Tuk3T14    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
MocT144    CGCAGCTAAC TGTGTGAAAT CGTGTGAACT GCAGGACACA TGAACATCGA CA--TTTT-- GAACGCATAT TGC--GGTCC ATGCAAAAGT TTAAATTTTT GTTTAAA-CT -GCATGGACC ACGTATGGTT
```

Fig. 1: Continued

```
                    *
          1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111112 2222222222 2222222222 2222222222 2222222222 2222222222 2222222222
          3333333334 4444444445 5555555556 6666666667 7777777778 8888888889 9999999990 0000000001 1111111112 2222222223 3333333334 4444444445 5555555556
          1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
AF218321  GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T--TTTTTTA AAGCACATTT
CypT20    GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATTC CTTTGGTT-- ---------- -TAAAACCTT -GGAGCTTTG GAATTATATT TTATGCTCTT AATATATTTT T--TTTTTTA AAGCACATTT
CypT34    GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA T--TTTTTTA AAACACATTT
CypT48    GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA TATTTTTTTA AAGCACATTT
CypT6     GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA TATTTTTTTA AAGCACATTT
Egt5T141  GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
Egy1T11   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA TATTTTTTTA AAACCCATTT
Egy1T39   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTT-A AAGCACATTT
Egy1T53   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATAT- ---TTTTTTA AAACCAATTT
Egy2T26   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
Egy2T40   GTGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT ----TTTTTA AAGCACATTT
Egy2T54   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC CTTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT ----TTTTTA AAACACATTT
Egy3T41   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGGTATG GAATTATATT TCATGCTCTT AATATATATA TATTTTTTTA AAGCACATTT
Egy3T69   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT ----TTTTTA AAGCACATTT
Egy4T28   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T-TTTTTTAA AAGCATTTTT
Egy4T70   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCCCTT AATATATATT ----TTTTAA AAAGCCACAT
Ity1T22   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA TATTTTTTTA AAACACATTT
Ity1T36   GAGTGTCGTA AATTTTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA TATTTTTTTA AAGCACATTT
Ity2T37   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAACACATTT
Ity3T38   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA TATTTTTTTA AAGCACATTT
Ity3T52   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAACTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
Ity3T66   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAAAACATTT
ILTjv     GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
JorT138   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T----TTTTA AAGCACATTT
Pal1T21   GAGTGTTGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
Pal1T49   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT TTTTTT---A AAGCACATTT
Pal1T63   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTAA AACCCATTTT
Pal3QT160 GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTAA AAGCCCATTT
Pal4RT162 GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAACCAATTT
syr1T15   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATCCTCTT AATATATATT ATATTTTAAA AAGCATATTT
Syr1T29   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
Syr2T58   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTAA AACCCAATTT
Syr3T157  GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT ----TTTTTA AAGACCATTT
Syr3T3    GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
syr3T31   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA T--TTTTTTA AAGCACATTT
Syr3T45   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTT-A AAGCACATTT
SyrT17    GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA TATTTTTTTA AAACACATTT
Tuk1T151  GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT TTTTTTTTTA AAGCACATTT
Tuk1T18   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT TGGAGCTATG GAATTATATT TTATGCTCTT TAAAAATATA T--TTTTTTA AAGCCAATTT
Tuk1T32   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATA TTTTTTTTAAA AAGCACATTT
Tuk1T46   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
Tuk2T33   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTAA AAGCAAATTT
Tuk3T14   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACATT -GGAGCTATG GAATTATATT TCATGCTCTT AATATATATT T---TTTTTA AAGCACATTT
MocT144   GAGTGTCGTA AATATTAAGC AATTGAAATT GTTTTTATAC ATTTGTAT-- ---------- -AAAAACCTT -GGAACCATG GAATTAAATT TCCTGCCCTT AATATATATT T---TTTTTT TTACACCTTT
```

Fig. 1: Continued

```
          2222222222 2222222222 2222222222 2222222223 3333333333 3333333333 3333333333 3333333333 3333333333 3333333333 3333333333 3333333333 3333333333 3333333333
          6666666667 7777777778 8888888889 9999999990 0000000001 1111111112 2222222223 3333333334 4444444445 5555555556 6666666667 7777777778 8888888889
          1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
AF218321  GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
CypT20    GAATGTACCC AA--TTG-TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAAAAA AACTT-ATTA TATTAAAAAA
CypT34    TAAAGTGCCC CCAGTG--TA AAA----TTA AAAATTTAAA AATATATATA --AAGGGGGT ATCATATTCT TTGG-GTAAA TTATTTTAAA AACACATTGT TTTACACAAA AAATT-TTTT TATTTAAAAA
CypT48    GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
CypT6     GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AGCTT-ATTA TATTAAAAAA
Egt5T141  GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAT TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
Egy1T11   TAATGGTCCC CA--AGGATA AAATA--ATA AAAATTGAAA TAAATATTAA --AAGGGGAA AATAAAATTA TTGG-ATAAA TTAATTTAAA AACCAATTGG TTTTCCAAAA AACTT-TTTT TTTTAAAAAA
Egy1T39   GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
Egy1T53   TAATGGTCCC CA--AGGTTA AAAT----TA AAAATTGAAA TATTTTTTTT --AAGGGGAT AATAAAATTA TTGG-ATAAA TTAATTTAAA AAACAATTGG TTTTCCAAAA AACTT-ATTA TTTTAAAAAA
Egy2T26   GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAT TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAAAAA AACTT-ATTA TATTAAAAAA
Egy2T40   GAATGTGCCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-GTAAT TTATCTTAAA GACATATTGT TACACAGAAA AACTT-ATTA TAATAAAAAA
Egy2T54   TAATGTGCCC CATGTG--TA AAA----TTA AAAGTTGAAA AATATATATA --AGGGGGGT ATCATATTCT TTGG-GTAAT TTATCTTAAA AACATATTGT TTTACACAAA AACTT-TTTT TATTAAAAAA
Egy3T41   GAATGTACCC AAG-TGG-TA TAAT--ATTA AAAGTTGAAA TATATATATA TAGGGGGTAT ATCATAGTCT GTGT-AAAAT ATATCATAAA CACATATTGT TATACAAAAA AACTT-TTTT ATTTTAAAAA
Egy3T69   GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCC TTGG-ATAAT TTATCTTTAA GACCTATTGT TATAAAAAAA AACTT-ATTT TTTTAATAAA
Egy4T28   GAATGCCCCC CCC-TGATAA TATAAAAAGA AAAGTAAAAA TATATGGGGG --GGGGGTTT ATCATAGTTA TTGG-ATAAT TTATCTTAAC GACATGTTGT TATACAAAAA AACTT-ATTA TATAAAAAAA
Egy4T70   ATAAGAGCGC GG--AAACTT TAAT---TGG GTTTTTAAAA AAATTTATTT AGAGGGGCGA AAAAAAATTC CCAACCCCCC AAGTGTTTAA AAAAA--TGT TAAAAAAAAA AAATTTTTTT TTTTAATATA
Ity1T22   GAATGTAACC CAATG---TA TAAT--ATTA AAAGTTGAAA TATATATATA --AAGGGGGT ATCATATTCA TTCG-AAAAT TCCAAGTTTT TATACAAAGG TATAAAAAAA AATTT-TATT TGTTTAAAAA
Ity1T36   GAATGTACCC AA--TGG-TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
Ity2T37   GATGGACCCC CC--GG--AT TAAT--ATTA AAAGTTGAAT TATTTTTTTA --AGGGGGAT ATAATAATAA TGGG-TTAAT TAATTTTAAA AACAAATTGT TATAAAAAAA AAATA-ATTA TTTTAAAAAA
Ity3T38   GAATGTCCCC GA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAAAAA AACTT-ATTA TATTAAAAAA
Ity3T52   GAATGTACCC CA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGGAT ATCATAATCA TTGG-ATAAT TTATCTTAAA AACATATTGT TATACAAAAA AACTT-ATTA TATTAAAAAA
Ity3T66   GAATGTACCA CA-TG---TA GAAT--ATTA AAAGTTGAAA TATATTTTTT --AGGGGGGT ATTATAATAA TTGG-AGAAT TTATCTTAAA AACATATTGT TGTACAAAAA AAC---AT-A TTATAAAAAA
ILTjv     GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAT TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
JorT138   GAATGTACCC CA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGGAT ATCATAATCA TTGG-ATAAT TTATCTTAAA AACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
Pal1T21   GAATGTACCC AA--TTG-GA TAATT-ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
Pal1T49   GAATGTGCCC AA--TG--TA TAAA--ATAA AAAGTTGAAA AATATATATA --AGGGGTAT ATCATATTCT TTGG-ATAAT TTATCTTAAA AACATATTGT TATACACAAA AACTT-ATTA TATTAAAAAA
Pal1T63   AATGGTCCCC A---TGG-TT TATT--TTTA AAGTTTAAAT TTTTTTTTTA --AGGGGTAA ATAA-AATTA TTGGAAAAAT TAATTTAAAA AACAAATGGT TTTCCAAAAA ACTTT-TTTT TTTTAAAAAA
Pal3QT160 AATTGTCCCC A---TGG-TT TAAA--ATTA AAATTTGAAT TATTTTTATA --AGGGGTAT ATCA-AATTC TTGGGATAAT TTATTTTAAA AACAAATTGT TATCCAAAAA ACCTT-TTTA TTTTAAAAAC
Pal4RT162 TAATGGTCCC CA--TGG-TA AAAA--ATAA AAATTTAAAA AATATTTAAA --AGGGGAAA ATAA-AATCC TTGGGAAATT TAATTTAAAA AACATATGGT TA-ACCAAAA AACTT--TTT TATTTAAAAA
syr1T15   TATTGTCCAC GA--TG--AA TA-AT-ATTA AAAGTTAAAA TATTTTTATA --AGCAGTAT ATAATAGTCG TTGA-ATAAT TTATATAAAA GACTTATTGT TATAAAAAAA AACTT-ATTA TAAAAAAAAT
Syr1T29   GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
Syr2T58   AATGGTCCCC A---AGG-TT AAAA--ATAA AAATTTAAAT ATTTTTTTTA --AGGGGATT ATAA-AATTA TTTGAATAAT TAATTTAAAA AACATTTTGT TATCCAAAAA AACTT-ATAA TTTAAAAAAA
Syr3T157  GAAAGTGCCC AC--TG--TG TAAT--ATTA AAAATTGAAA AATATATTTA --AGGGGACC ACCCCACCCC CCCGCACAAT CTATCTTATA GACATATTGT GATATACAAA AGCTT----- ----------
Syr3T3    GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAT TATATATATA --AGGGGTAT ATCATAATTA TTGG-ATAAT TTATCTTAAA AACATATTGT TATACAAAAA AACTT-ATTA TATTAAAAAA
syr3T31   GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAT TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
Syr3T45   GAATGTACTC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
SyrT17    GAGTGTGCCC CA--TGG-TA TAAT--ATAA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCT TTGT-ATAAT ATATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA T---------
Tuk1T151  GAATGTCCCC AA--TG--TA TATT--ATTA AAAGTTGAAA TATATATATG --GGGGGTAT ATCATATTCG TTGA-ATAAT TTATCTTAAA GACTTATTGT TATACAAAAA AACTT-ATTA TATAAAAAAA
Tuk1T18   GAATGTCCCC AAT-TG--GT TAAT--ATTA AAAGTTGAAA ATTTTTTTTA --AGCGGGAT ATCAAAATCA TTGG-ATAAT TTATCTTAAA AACATTTTAG TATACAAAAA AACTT-ATTA TATTAAAAAA
Tuk1T32   GAGTG-CCCC CTG-TG--TA TAAT--ATAA AAAGTTGAAA TATATATATG --GGGGGTAT ATCATAGTCT GTGG-ATAAT TTATCATAAA CACATATTGT TATACAAAAA AACTT-ATT- ATATAAAAAA
Tuk1T46   GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAGAAA AACTT-ATTA TATTAAAAAA
Tuk2T33   GAAGGTCCCC AA--GG--TT TAAT--TTTA AAAGTTAAAA TATTTTTTTA --AGGGGTAT ATCAAATTCT TTGG-TAAAT TTATTTTAAA AACAAATGGT TTTCAAAAAA AAACT-TTTT TTTTTAAAAA
Tuk3T14   GAATGTACCC AA--TG--TA TAAT--ATTA AAAGTTGAAA TATATATATA --AGGGGTAT ATCATAGTCA TTGG-ATAAT TTATCTTAAA GACATATTGT TATACAAAAA AACTT-ATTA TATTAAAAAA
MocT144   TTTTGTTCCC CC--CG--TT TAAT--ATTA AAAGTTGAAA AATTTTTTTA --AGGGGGGG ATCATCATCT TTGG-TTAAT TATTTTTCTT AACGACTTTT TTTACAACAA AACTA-CTTA TTTTAATAAA
```


Fig. 1: Continued

```
                                                                                                              *
           3333333334 4444444444 4444444444 4444444444 4444444444 4444444444 4444444444 4444444444 4444444444 4444444444 4444444445 5555555555 5555555555
           9999999990 0000000001 1111111112 2222222223 3333333334 4444444445 5555555556 6666666667 7777777778 8888888889 9999999990 0000000001 1111111112
           1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
AF218321   -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
CypT20     -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
CypT34     ATTTTTAAAA A-AAGGGGAT TTTTCAAAT- --AAAAAA-- ---------- ---------- -------GTG CTACA-ATAA TTTAAAATT- TTTGCGATCT CAACTCATAC GTGACTAC-- --CCCCCGAA
CypT48     -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
CypT6      -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Egt5T141   -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATTT ATTGGGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Egy1T11    ACTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Egy1T39    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Egy1T53    ACTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Egy2T26    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Egy2T40    ACCTTTAAAA A-AATGGGGG TTTTCAAAA- --TATTAA-- ---------- ---------- ------TGTG CAACA-AAAA TTAAATATTT TTTG-GAGCC CCAATCATTC CTAACTAC-- --CCCCCCAA
Egy2T54    AATTTTAAAA A-AAGGGGGT TTTTCAAAA- --AAAAGT-- ---------- ---------- -------GTG C--CA-AAAA TTTATTATT- TTTGCGGGCC CTAACCACAC GTGACTCC-- --CCCCTGAT
Egy3T41    ATTTTTAAAA A-AGGGGGGG TTTTCAA--- --TAAAAT-- ---------- ---------- ----AGTGTG CCACA-ATAA TTTAATTTT- AGTGCGATCT CCACACATAC GTGACCCC-- --CCCCTGTA
Egy3T69    ACATTTTTTA A-AATGGGGG GATTCTG--- --TATAAT-- ---------- ---------- -------AAA CAACA-ATTT TTAATTTTTT TTTGCGA--C CAAATTATTC TCAAAAAC-- --CCCCATTC
Egy4T28    -CTTTTAAAA A-AGGGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----GATGTG CTACA-ATAA TATAATATT- GTTGCGATCT CAACTCATAC GACACCCC-- --CCCCTGAA
Egy4T70    AACGCCCCCA A-ACCGGGGC TGGCCCCTG- --CATAAA-- ---------- ---------- ---------- ----AAAAAA TTTTTAATT- TTTGGGGGGG CGA------- ---------- ----------
Ity1T22    TTCCTTACAT ACAATGGGGT TGTTCAAAA- --AAAAAA-- ---------- ---------- ------TGTG GTGCA-ATCT GTTAATGGT- ATTGCTCCCT CAAAACATAC GTTTTAG--- --CCCCTGAA
Ity1T36    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Ity2T37    -CTTTTTATA A-AATGGGGT TATTATT--- --TATTAT-- ---------- ---------- -------GTG CTACA-ACAA TTTAATATT- ATTGTTATCT CAACTCATAC TTGACTAC-- --CCCCTGAA
Ity3T38    -CTTTTAATA A-GGTGGGGT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- GTTGCGATCT CAACTCATAC GTGACTCC-- --CCCCTGTA
Ity3T52    -CTTTTAATA A-AATGGGGT TATTCAA--- --TATAAT-- ---------- ---------- -------GAG CTACA-ACAA TTAAATATT- ATTGTTATCT CAACTCATAC TTGACTAC-- --CCCCTGAA
Ity3T66    AAA-ATAATA A-AAAGGGAG TATTATTTAT --AATAAT-- ---------- ---------- ----GATGTG CTACA-ATAA AATATTATT- ---GTTATCG CAACTCCTAC TTGACTAC-- --CCCCCCCA
ILTjv      -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATTT ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
JorT138    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG GTGCA-ACAA TTTTNTATT- ATTGTTNNGA CAACTCATAC TTGACGAC-- --CCCCCGAA
Pal1T21    ACTTTTAATA A-AAGGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ------TGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Pal1T49    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----ATGGTG CAACA-ATAT TTTATTTTT- TTTGCGATCT CAACTCATAC GTGACTCC-- --CCCCTGTT
Pal1T63    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATT- TTTGGGAACT CAAATTATAC CTGAATAC-- --CCCCTGGA
Pal3QT160  -CTTTTATTA A-AATGGGAT AATTCAT--- --TATATT-- ---------- ---------- -------GTG CTCCA-AAAT TTTATTATT- ATTGGAAGGC TAACCACAAA GTGGCGAA-- --CCACCGGT
Pal4RT162  ACTTTTAAAA AAATGGGGTT TTTTCAATA- --AATTAT-- ---------- ---------- -------TGG CCAAA-ATTT TTTTATTTTT ATGGCGATCT CAATTCATAC GTGACCCC-- --CCCTTGTT
syr1T15    -TTTTTAAAA A-GGGGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTATTATT- ATTGCGACCT CAACTCATAC GTGACTAC-- --CCCCTGAT
Syr1T29    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Syr2T58    -CTTTAAAAA A-AAGGGGAT AATTCAA--- --AAAAAG-- ---------- ---------- -------GGG CAACA-AAAA TTAAATATT- TTTGGAAATC CAATCA-TAC CCGAATAC-- -CCCCCGGAA
Syr3T157   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Syr3T3     -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATTT ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
syr3T31    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Syr3T45    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
SyrT17     -CTTTTAAAA A-AAGGGGGT TATTCAA--- --TATAAT-- ---------- ---------- ----ATTGTG CTACA-ATAA TTTAATATT- TTTGCGATCT CAACTCTTAC ATGACTACTC -CCCCCTGTG
Tuk1T151   -CTTTTAATA A-GGTGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----ATTGTG CAACA-ATAA TTTATTATT- ATTGCGATCT CACCTCATAC GTGACTCC-- --CCCCTGAT
Tuk1T18    -CTTTTAATA A-AGGGGGAT TATTCAA--- --TATAAT-- ---------- ---------- -------GTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCCGAT
Tuk1T32    ACTTTTAATA A-AGGGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AGTGTG CCACA-ATTT TAAAATATT- ATTGCGATCT CAACTCACAC GTGACCCC-- --CCCGTGTT
Tuk1T46    -CTTTTAATA A-AATGGGAT TATTCAA--- --TATAAT-- ---------- ---------- ----AATGTG CTACA-ATAA TTTAATATT- ATTGCGATCT CAACTCATAC GTGACTAC-- --CCCCTGAA
Tuk2T33    ATTTTTAAAA A-AAGGGGAT TTTTCAATT- --TATTAT-- ---------- ---------- ------TGAA AAAAA-ATTT TTTATTTTT- GGGGGGGCCC CAACTCTTCC AAAACTA--- --CCCCCGAT
Tuk3T14    -CTTTTAATA A-AGTGGGAT TATTCAAAT- --TTTAATGT GCTAGGGTAT TATTACTATT TTTGAATGTG CTACA-AACA TTNAATCTT- ATTGCTAACT CAANTAATAC GTGACTAN-- --CCCCTGAA
MocT144    -CATTTTTTA A-AACGGGGG GGTTTTTTGG --TATTTT-- ---------- ---------- ---------- CTTCT-CCCT TTTTATTTT- TATGTTATGG CCACCCA-AC TCGACA---- --CCCCCCGC
```

Fig. 1: Continued

```
          5555555555 5555555555 5555555555 5555555555 555
          2222222223 3333333334 4444444445 5555555556 666
          1234567890 1234567890 1234567890 1234567890 123
AF218321  TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC AGG
CypT20    TTTAAGCATA ATAATAAGC- GGAGGAAAAA AA-ACTAACC CGG
CypT34    ATTAAACACA TTAAAAAAC- GGGGGAAAAA AAAACTACCC CGG
CypT48    TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC AGG
CypT6     TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC CGG
Egt5T141  TTTAACCATA TAAAAAACCC GGAGGAAAAA AA-ACTACCC AGG
Egy1T11   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC CGG
Egy1T39   TTTAAGCCTA TTAATAAGC- GGAGGAAAAG AA-ACTAACC CGG
Egy1T53   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC AGG
Egy2T26   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC GGG
Egy2T40   TTTTTGCCTC CCAATAACCC GAGGGAAAAA AACAATTGCC AGG
Egy2T54   TTAAAGCATT TTTATACGC- GGAGGAAAAG AA-ACTAACC AGG
Egy3T41   TTTACACATA TAAATCCGC- GGAAGAAAAA AA-ACTAACC AGG
Egy3T69   AAAATGAATG CTAATTC-CA CATATTTAAA AAAAATACCC AGG
Egy4T28   TTTAATCATA ATAAGAAGC- GGAGGAAAAA AA-CCCCCCC AGG
Egy4T70   ---------- ---------- ---------- ---------- ---
Ity1T22   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC AGG
Ity1T36   TTTAAGCATA TTAATAAGC- GGGGGAAAAG AA-ACTAACC AGG
Ity2T37   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AAAAAAACCC GGG
Ity3T38   TTTAAGCATA TTAATAAGC- GGAGGAAAAA AA-ACTCACC AGG
Ity3T52   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-AAAAACC AGG
Ity3T66   TTTAAGCATA ATAATAATC- AGAGGAGAAG AAAAAAACCC AGA
ILTjv     TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC AGG
JorT138   GTTTTTCATC TTATTAAGC- GGAGGAGGAG AA-AATACTC AGG
Pal1T21   TTTAAGCATA TTAATAAGC- GGGGGAAAAA AA-ACTAACC AGG
Pal1T49   TTTAAACATA TTAAAAAGC- GGAGGAAAAA AA-ACTCCCC GGG
Pal1T63   TTTAAACCTT TTAATTAA-C CGAGGAAAAA AAACCT-ACC CGG
Pal3QT160 TTTTAACCTT TTTATAAAAC CGCGGAAAAA AAAACACACC CGG
Pal4RT162 TTCAACCTTA TTAACCCGC- -GGAGAAAAA AAAACTACCC GGG
syr1T15   TTTAATCATA TTAATAAGC- GGAGGAAAAA AA-ACTAACG NGG
Syr1T29   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC GGG
Syr2T58   TTTAACCCAT TTAAAAAA-G GAGGGAAAAA AAATCT-ACC AGG
Syr3T157  ---------- ---------- ---------- ---------- ---
Syr3T3    TTTAAGCATA TTAATAAGC- GGAGGAAAAA AA-ACTAACC AGG
syr3T31   TTTAAGCATA TTAATAAGC- GGAAGGAAAA AA-ACTAACC AGG
Syr3T45   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC AGG
SyrT17    TTTAAGCATA TAAATAAGC- GGAGGAAAAA AA-ACTAACC AGG
Tuk1T151  TTTAACCTTA TTAATCGGC- GGAGAAAAAA AA-ACTAACC AGG
Tuk1T18   TTTTAG---T TTTATAAGC- GGGAGAAAAG A----TAACC AGG
Tuk1T32   TTTAAGCATA TTAATAAGC- GGGAGAAAAA AA-ACTAACC GGG
Tuk1T46   TTTAAGCATA TTAATAAGC- GGAGGAAAAA AA-ACTACCC GGG
Tuk2T33   TTTTGGCCCG TTAATAACC- GGAGGAAAAA AA-ATTCACC AGG
Tuk3T14   TTTAAGCATA TTAATAAGC- GGAGGAAAAG AA-ACTAACC AGG
MocT144   GTTTAGCATA ATCCTAAGC- GGAGGAAAAG AA-ACTAACC AGG
```

Fig. 1: Sequence alignment of the ITS2 and flanking 5.8S and 28S coding regions of the rDNA of 43 *P. papatasi* individuals. Start and end of the presumptive ITS2 region are marked by asterisks. Insertions or deletions (indels) in alignment are denoted by a dash (-). Abbrevaitions are the same as in Table 1

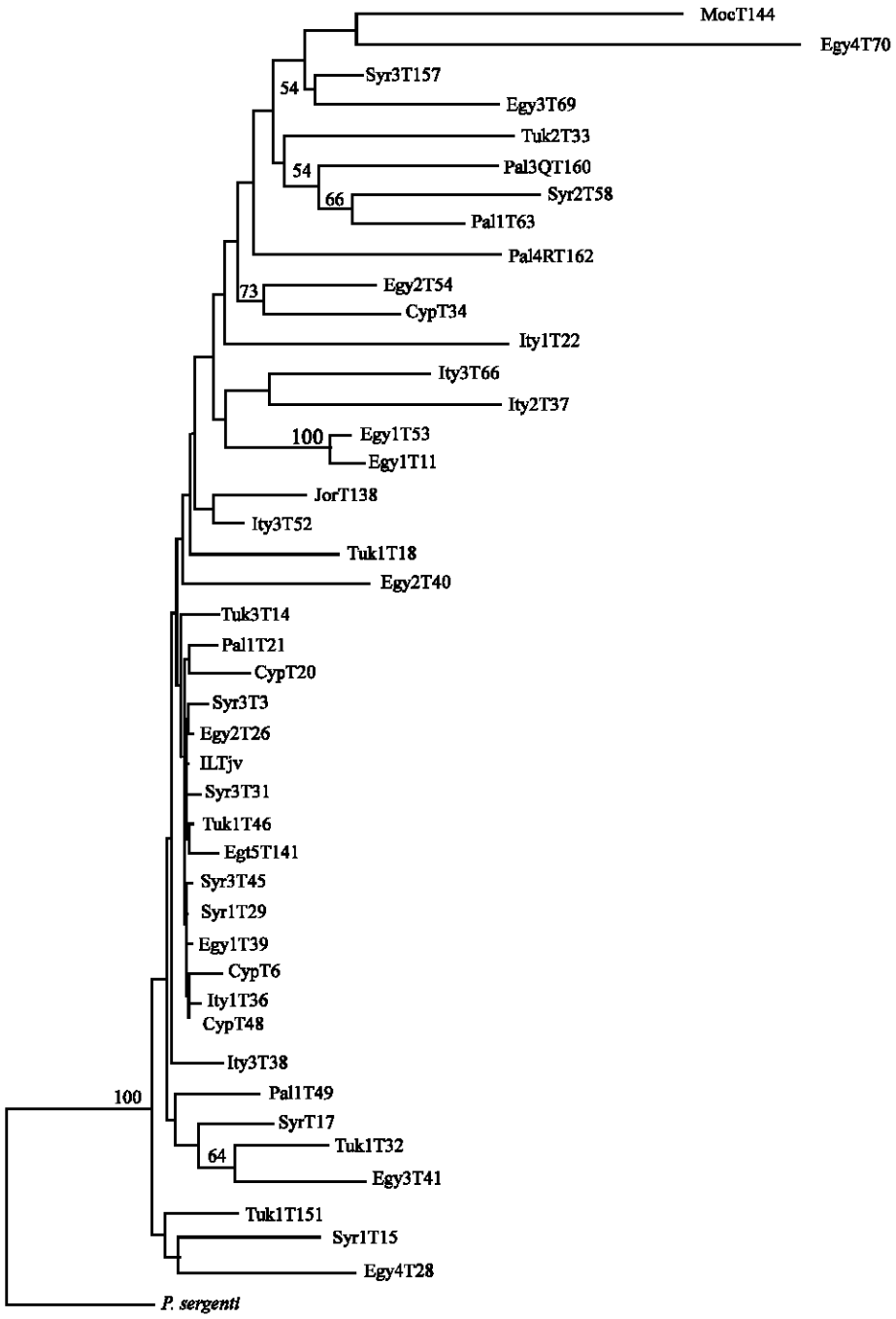

Fig. 2: A neighbour-joining bootstrap tree based on rDNA ITS2 sequence data for 43 samples of *P. papatasi* from 25 different populations originating from 10 countries. The tree was rooted with *P. sergenti* ITS2 sequence (GenBank: AF462332.1). Numbers show bootstrap confidence levels above 50% (from 1000 replications) above the branches tested. Sample codes are the same as in Table 2

Table 2: Length and GC content of ITS2 for the 43 samples sequenced in this study

Sample code	Length (bp)	GC (%)	Accession No.
AF218321	488	28.5	
CypT20	490	28.6	DQ887632
CypT34	491	28.7	DQ887633
CypT48	490	28.4	DQ887634
CypT6	490	28.6	DQ887635
Egt5T141	486	28.8	DQ887636
Egy1T11	493	28.0	DQ887637
Egy1T39	483	29.2	DQ887638
Egy1T53	484	27.9	DQ887639
Egy2T26	487	28.5	DQ887640
Egy2T40	490	30.8	DQ887641
Egy2T54	484	31.0	DQ887642
Egy3T41	495	30.3	DQ887643
Egy3T69	484	27.7	DQ887644
Egy4T28	494	32.0	DQ887645
Egy4T70	419	32.9	DQ887646
ILTjv	485	28.7	DQ887653
Ity1T22	492	29.7	DQ887647
Ity1T36	491	28.9	DQ887648
Ity2T37	483	30.2	DQ887649
Ity3T38	490	29.8	DQ887650
Ity3T52	484	27.9	DQ887651
Ity3T66	485	28.1	DQ887652
JorT138	480	29.6	DQ887654
MocT144	481	33.5	DQ887674
Pal1T21	488	28.9	DQ887655
Pal1T49	487	29.0	DQ887656
Pal1T63	484	27.7	DQ887657
Pal3QT160	486	29.6	DQ887658
Pal4RT162	486	29.2	DQ887659
Syr1T15	486	27.3	DQ887660
Syr1T29	484	28.9	DQ887661
Syr2T58	484	27.9	DQ887662
Syr3T157	340	33.6	DQ887663
Syr3T3	485	27.6	DQ887664
Syr3T31	487	28.3	DQ887665
Syr3T45	483	28.5	DQ887666
SyrT17	485	30.1	DQ887667
Tuk1T151	490	29.4	DQ887668
Tuk1T18	483	28.7	DQ887669
Tuk1T32	490	31.0	DQ887670
Tuk1T46	487	28.7	DQ887671
Tuk2T33	487	29.1	DQ887672
Tuk3T14	512	28.3	DQ887673
Average	482.6	29.2	

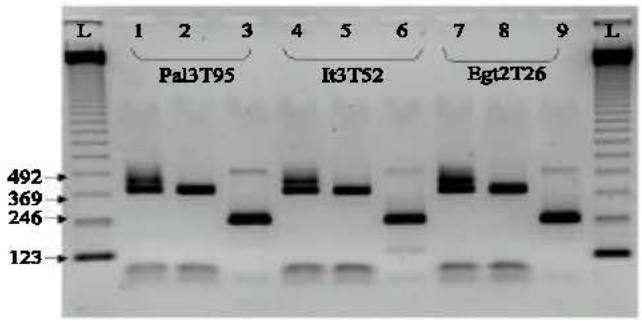

Fig. 3: Polymerase chain reaction restriction fragment length polymorphism patterns for the internal transcribed spacer 2 (ITS2) of three *P. papatasi* individuals; Pal3T95, It3T52 and Egt2T26 after digestion with Sau961(lane1-3), *AvaII* (lane 4-6)and *RsaI* (lane 7-9) restriction enzymes

DISCUSSION

The high level of genetic diversity in the ITS2 region of the *P. papatasi* individuals studied here denotes intraspecific variation and even intra-individual variation in the multiple copies of the rDNA subunits. The GC content (29.2%) of the ITS2 sequences obtained fell below the range of the anopheline species (Beebe and Cooper, 2000). The ITS2 region in *P. papatasi* appears to display several AT dinucleotide microsatellite repeats which in part contributes to the polymorphic nature of the ITS2. This could explain the bias toward A and T which was consistent with the base composition of ITS2 sequences of other sand fly species such as *Phlebotomus sergenti*, *Phlebotomus similis* (Depaquite, 2002) and *Lutzomyia shannoni* (Miller *et al*., 1997).

It seems that variants within individual sand flies often differed as much as those between flies from different localities. This is actually the reason for poor geographical structuring in the NJ tree and also contradicts with our previous studies on *P. papatasi* based on *cytochrome b* gene sequences (data not shown). This is not surprising however, since *cytochrome b* gene is maternally inherited and not recombinant while ITS2 is free to diverge.

At the genomic level, the effectiveness of molecular drive at homogenizing a multigene family depends on the rates of unequal crossing over and gene conversion and is negatively correlated with both mutation rate and number of unlinked loci of the multigene family (Perelson and Bell, 1977). Intra-individual sequence heterogeneity indicated that the process of concerted evolution had not homogenized the ITS2 sequence in some individuals. Perelson and Bell (1977) noted that unequal crossing over increases heterogeneity among copies of a multigene family if mutation rates are high relative to homogenization rates.

Extensive intragenomic variation can confound phylogenetic and population studies based on ribosomal spacers, however it is necessary to first determine the extent of homogenization of a multigene family at several hierarchical levels (within individual, population and species) before employing them in phylogenetic studies (Williams *et al*., 1988; Sanderson and Doyle, 1992).

The high polymorphism of ribosomal ITS2 in *P. papatasi* is mainly due to insertion\deletion mutations and changes in short tandem repeat sequences. Increases or decreases in the size of ITS2 may have been caused by replication slippage. In this study ITS2 sequences were shown to be too variable for phylogenetic analysis so that genetic relationships between different populations could not completely resolved, this actually makes the use of the current ITS2 sequences of little value for phylogenetic interpretation.

As a conclusion, the ITS2 does not represent a suitable marker for inferring phylogenetic and population genetic relationships across *P. papatasi* sand flies and that other genetic markers are needed for this purpose.

ACKNOWLEDGMENTS

We are grateful to Hala Kassem, Shaden Kamhawi, Michele Maroli, Karim Oun, Alon Warburg, Peter Volf, Vladimir Ivovic, Dr. Souad Guernaoui and Asli Belen for providing sand fly samples. Thanks go as well to Deutscher Akademischer Austauch Dienst (DAAD) for providing a Ph.D scholarship to Omar Hamarsheh.

REFERENCES

Arnheim, N., 1983. Concerted Evolution of Multigene Families. Evolution of Genes and Proteins Nei, M. and R.K. Koehn (Eds.), Sinauer Associates, Sunderland, MA., pp: 38-61.

Beckingham, K., 1982. Insect rDNA. The Cell Nucleus. Busch, H. and L. Rothblum (Eds.), Academic Press, New York, pp: 205-263.

Beebe, N.W. and R.D. Cooper, 2000. Systematics of malaria vectors with particular reference to the Anopheles punctulatus group. Intl. J. Parasitol., 30: 1-17.

Black, W.C.I.V., D.K. McLain and K.S. Rai, 1989. Patterns of variation in the rDNA cistron within and among world populations of a mosquito, *Aedes albopictus* (Skuse). Genetics, 121: 539-550.

Collins, F.H. and S.M. Paskewitz, 1996. A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic Anopheles species. Insect Mol. Biol., 5: 1-9.

Depaquit, J., H. Ferte´, N. Le´ger, R. Killick-Kendrick, J.A. Rioux, M. Killick-Kendrick, H.A. Hanafi, and S. Gobert, 2000. Molecular systematics of the Phlebotomine sandflies of the subgenus *Paraphlebotomus* (Diptera, Psychodidae, Phlebotomus) based on ITS2 rDNA sequences. Hypotheses of dispersion and speciation. Insect Mol. Biol., 9: 293-300.

Depaquit, J., H. Ferte´, N. Le´ ger, F. Lefranc, C. Alves Pires, H. Hanafi, M. Maroli, F. Morillas-Marquez, J.A. Rioux, M. Svobodova and P. Volf, 2002. ITS 2 sequences heterogeneity in *Phlebotomus sergenti* and *Phlebotomus similis* (Diptera, Psychodidae): Possible consequences in their ability to transmit *Leishmania tropica*. Intl. J. Parasitol., 32: 1123-1131.

Dover, G. and E. Coen, 1981. Springcleaning ribosomal DNA: A model for multigene evolution? Nature, 290: 731-732.

Dover, G., 1982. Molecular drive: A cohesive mode of species evolution. Nature, 299: 111-117.

Di Muccio, T., M. Marinucci, L. Frusteri, M. Maroli, B. Pesson and M. Gramiccia, 2000. Phylogenetic analysis of *Phlebotomus* species belonging to the subgenus Larroussius (Diptera, Psychodidae) by ITS2 rDNA sequences. Insect Biochem. Mol. Biol., 30: 387-393.

Esseghir, S., P.D. Ready, R. Killick-Kendrick and R. Ben-Ismail, 1997. Mitochondrial haplotypes and phylogeography of *Phlebotomus* vectors of *Leishmania major*. Insect Mol. Biol., 6: 211-225.

Esseghir, S., P.D. Ready and R. Ben-Ismail, 2000. Speciation of *Phlebotomus* sand flies of the subgenus Larroussius coincided with the late Miocene-Pliocene aridification of the Mediterranean subregion. Biol. J. Linn. Soc., 70: 189-219.

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution, 39: 783-791.

Gerbi, S.A., 1985. Evolution of ribosomal DNA. In: Molecular Evolutionary Genetics. MacIntyre, R.J. (Ed.), Plenum Press, New York, pp: 419-517.

Guernaoui, S., A. Boumezzough, B. Pesson and G. Pichon, 2005. Entomological investigations in Chichaoua: An emerging epidemic focus of cutaneous leishmaniasis in Morocco. Med. Vet. Entomol., 42: 697-701.

Higgins, D., J. Thompson, T. Gibson, J.D. Thompson, D.G. Higgins and T.J. Gibson, 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22: 4673-4680.

Hillis, D.M. and M.T. Dixon, 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. Quart. Rev. Biol., 66: 411-453.

Killick-Kendrick, R., A.J. Leaney, W. Peters, J.A. Rioux and R.S. Bray, 1985. Zoonotic cutaneous leishmaniasis in Saudi Arabia: The incrimination of *Phlebotomus papatasi* as the vector in the Al-Hassa oasis. Trans. R. Soc. Trop. Med. Hyg., 79: 252-255.

Killick-Kendrick, R., 1990. Phlebotomine vectors of the leishmaniases: A review. Med. Vet. Entomol., 4: 1-24.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol., 6: 111-120.

Kumar, S., K. Tamura and M. Nei, 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform., 5: 150-163.

Miller, B.R., M.B. Crabtree and H.M. Savage, 1996. Phylogeny of fourteen *Culex mosquito* species, including the *Culex pipiens* complex, inferred from the internal transcribe spacers of ribosomal DNA. Insect Mol. Biol., 5: 93-107.

Miller, B.R., M.B. Crabtree and H.M. Savage, 1997. Phylogenetic relationships of the *Culicomorpha inferred* from 18S and 5.8S ribosomal DNA sequences (Diptera: Nematocera). Insect. Mol. Biol., 6: 105-114.

Perelson, A.S. and G.I. Bell, 1977. Mathematical models for the evolution of multigene families by unequal crossing over. Nature, 265: 304-310.

Perfil'ev, P.P., 1966. Fauna of the U.S.S.R. Diptera. Academy of Sciences of U.S.S.R. Zoological Institute. New Series 93 III. (Translated from Russian by Israel Program for Scientific Translations, Jerusalem, 1968.

Rozas, J., J.C. Sánchez-DelBarrio, X. Messeguer and R. Rozas, 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics, 19: 2496-2497.

Sanderson, M.J. and J.F. Doyle, 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: Concerted evolution, homoplasy and confidence. Syst. Biol., 41: 4-17.

Saitou, N. and M. Nei, 1987. The neighbour-joining method, a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., 4: 406-425.

Sawalha, S.S., M.S. Shtayeh, H.M. Khanfar, A. Warburg and Z.A. Abdeen, 2003. Phlebotomine sand flies (Diptera: Psychodidae) of the Palestinian West Bank: Potential vectors of leishmaniasis. J. Med. Entomol., 40: 321-328.

Schlein, Y., A. Warburg, L.F. Schnur and A.E. Gunders, 1982. Leishmaniasis in the Jordan Valley, II. Sandflies and transmission in the central endemic area. Trans. R. Soc. Trop. Med. Hyg., 76: 582-586.

Seperack, P., M. Slatkin and N. Arnheim, 1988. Linkage disequilibrium in human ribosomal genes: Implications for multigene family evolution. Genetics, 119: 943-949.

Tautz, D., C. Tautz, D. Webb and G.A. Dover, 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species. J. Mol. Biol., 195: 525-542.

Van de Peer, Y. and R. De Wachter, 1994. TREECON for Windows: A software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput. Applied Biosci., 40: 569-570.

Wesson, D.M., C.H. Porter and F.H. Collins, 1992. Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). Mol. Phylogen. Evol., 1: 253-269.

Williams, S.M., R.W. DeBry and J.L. Feder, 1988. A commentary on the use of ribosomal DNA in systematic studies. Syst. Zool., 37: 60-62.

Xu, J.N. and F.Y. Qu, 1997. Ribosomal DNA difference between species A and D of the Anopheles dirus complex of mosquitoes from China. Med. Vet. Ent., 11: 134-138.

Yaghoobi-Ershadi, M.R., A.A. Akhavan, A.R. Zahraei-Ramazani, A.R. Jalali-Zand and N. Piazak, 2005. Bionomics of *Phlebotomus papatasi* (Diptera: Psychodidae) in an endemic focus of zoonotic cutaneous leishmaniasis in central Iran. J. Vect. Ecol., 30: 115-118.