Genet. Res., Camb. (1996), **67**, *pp.* 227–237 *With 6 text-figures*

Evolution of multiple families of non-LTR retrotransposons in phlebotomine sandflies

DAVID R. BOOTH[1], PAUL D. READY[1,*] AND DEBORAH F. SMITH[2]
[1] *Molecular Systematics Division, Department of Entomology, The Natural History Museum, London SW7 5BD, UK*
[2] *Department of Biochemistry, Imperial College of Science, Technology & Medicine, London SW7 2AZ, UK*

(*Received 21 April 1995 and in revised form 18 December 1995*)

Summary

In this paper we report on the diversity and distribution of a set of non-LTR retrotransposon (RTP) reverse transcriptase (RT) sequences isolated from phlebotomine sandflies, and their potential for investigating the evolutionary histories of members of this subfamily of flies (Diptera: Psychodidae, Phlebotominae). The phlebotomine RT sequence families derived from one species were as different from each other as they were from RT sequences derived from other species. When each was used to probe Southern blots of sandfly genomic DNA they hybridized only to the species of source and, usually, to others of the same subgenus, but not to DNA from other subgenera – a hybridization pattern consistent with vertical evolution. There was considerable intraspecific variation in hybridization pattern, suggesting the RTs were part of non-LTR RTPs that are (or were recently) subject to flux in genomic position and copy number. Most of the RT families detected in phlebotomines are monophyletic with respect to previously described RTs, and all are monophyletic with RTs of the F/Jockey (*Drosophila melanogaster*) type of RTP. Orthologous sequences were isolated from the closely related species *Phlebotomus perniciosus* and *P. tobbi* (subgenus *Larroussius*), and different populations of *P. perniciosus*. The level of sequence divergence among these orthologous RTs, the subgeneric distribution of each RT family, and the intraspecific variation in hybridization pattern of many of them, indicate this class of sequence will provide genetic markers at the sub-generic level.

1. Introduction

Female phlebotomine sandflies are haematophagous and, in subtropical and tropical regions, many *Phlebotomus* (Old World) and *Lutzomyia* (New World) species transmit to man arboviruses and, more importantly, the trypanosomatid protozoan *Leishmania* (Killick-Kendrick, 1990). We have isolated non-LTR retrotransposon (RTP) reverse transcriptase-like (RT) sequences from Old and New World sandflies to investigate the diversity, distribution and evolutionary history of such sequences and their value as (phylo-)genetic markers.

Non-LTR RTPs are mobile genetic elements that contain a RT gene, like all retroelements, but lack an envelope gene and long terminal repeats (LTRs). They are considered to have appeared early in the evolution of retroelements, perhaps predating the divergence of eukaryotes in which they may be ubiquitous (Doolittle *et al.* 1989; Xiong & Eickbush, 1990). Understanding the evolution of transposons is complicated by the possibility of natural, horizontal transfer between host species, but unlike many DNA elements (e.g. P, Mariner) the phylogenetic distribution of non-LTR RTPs often correlates well with that of their hosts (e.g. I elements in drosophilids, and LINES in mammals) (Bucheton *et al.* 1986; Simonelig *et al.* 1988; Hutchison *et al.* 1989; Daniels *et al.* 1990).

Elsewhere, we have reported the isolation of RT sequences of non-LTR RTPs from sandflies, by using for polymerase chain reaction amplification (PCR) 'universal primers' or 'phlebotomine specific primers' (Booth *et al.* 1994). Here, we use the data obtained by Southern hybridization and cladistic analysis to investigate the relationships and distributions of these sequences. RT sequences came from the neotropical phlebotomine *Lutzomyia longipalpis* and two taxo-

* Corresponding author: Dr P. Ready, Molecular Systematics Division, Department of Entomology, The Natural History Museum, London SW7 5BD. Phone: 0171 9389356. Fax: 0171 9388937. email: P.Ready@nhm.ac.uk.

nomic pairs of Old World leishmaniasis vectors: *Phlebotomus perniciosus* and *P. tobbi* of the subgenus *Larroussius*, and *P. papatasi* and *P. duboscqi* of the subgenus *Phlebotomus*. Individual phlebotomine species were found to harbour families of RT sequences that have diverged greatly but are clearly monophyletic with respect to all non-phlebotomine RTs except those of the F/Jockey (*Drosophila melanogaster*) type recognized by Xiong & Eickbush (1990).

2. Materials and methods

(i) *Isolation and sequencing of RT sequences*

These were obtained by PCR using pairs of degenerate primers, as follows: 'universal primers' – the amino acid motifs YRPIS{forward primer}/AYADD {reverse} (to amplify a *c.* 500 base pair (bp) DNA fragment between peptide domains 2 and 5); and 'phlebotomine specific primers' – VFSKIFE{forward}/KAFDTVC{reverse} (to amplify a *c.* 230 bp DNA fragment between peptide domains 2 and 3), and VFSKIFE{forward}/AYADD{reverse} (to amplify a *c.* 475 bp DNA fragment) (Xiong & Eickbush, 1990; Booth, 1993; Booth *et al.* 1994). Endonuclease-digested and purified PCR products were ligated into the sequencing vector M13 tg130, and *Escherichia coli* XL-1 cells (Stratagene) transformed with recombinant phage were identified by white (not blue) phenotype or by screening such libraries with a previously isolated phlebotomine RT fragment (Sambrook *et al.* 1989). Sequences hom-

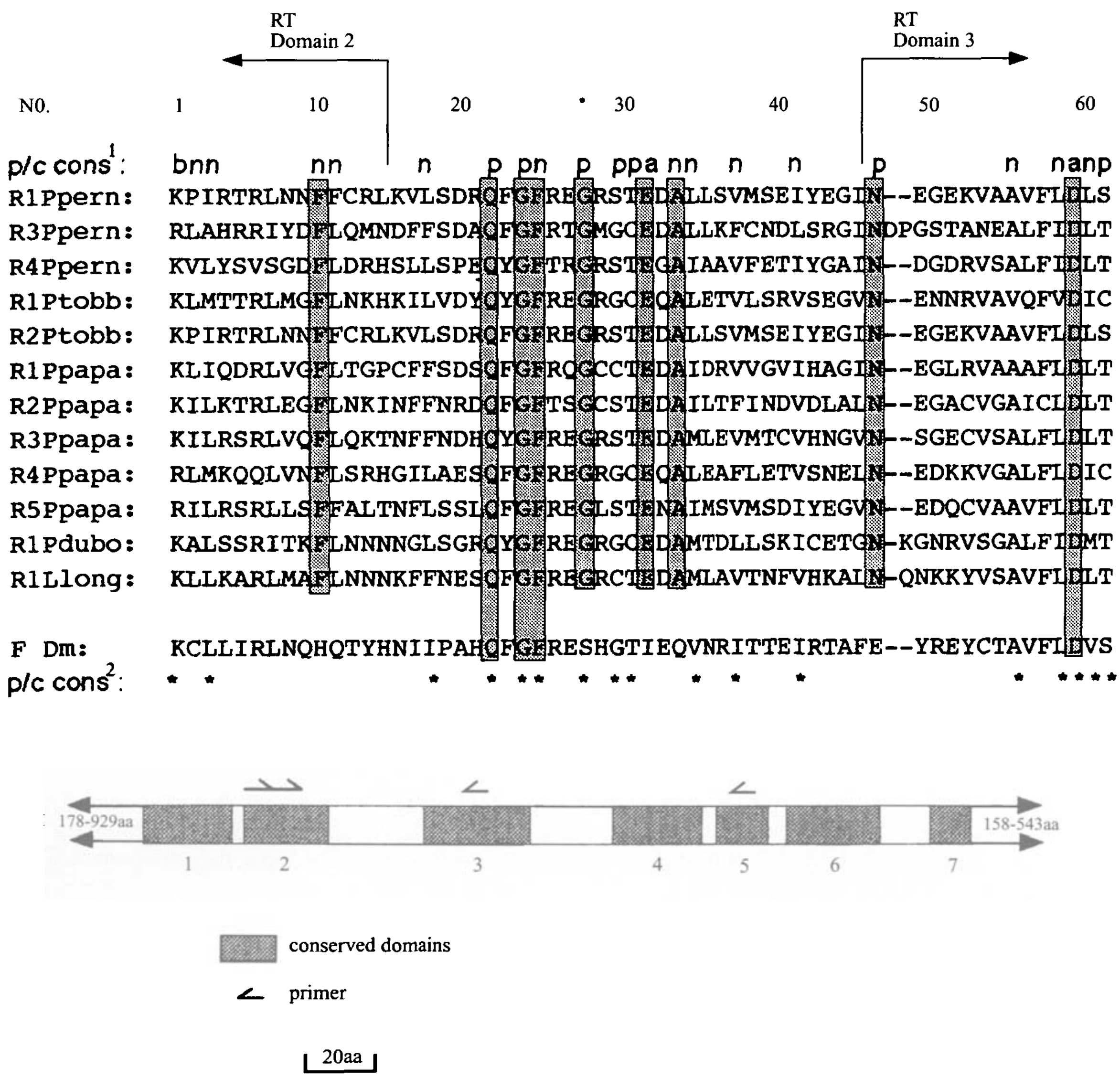

Fig. 1. Phlebotomine non-LTR retrotransposon reverse transcriptase deduced amino acid sequences. Capital letters – amino acids (universal code); p/c cons[1]: RT residues of the same physical-chemical group in all phlebotomines; p/c cons[2]: those residues also conserved in the F element of *D. melanogaster*. Amino acid residues are classified according to Lehninger (1970): n, non-polar; p, polar; a, acidic; or b, basic side-chains; the RT sequences are aligned and the domains labelled according to Xiong & Eickbush (1990). A diagram of the generalised non-LTR retrotransposon RT gene is shown at the base of the figure.

https://doi.org/10.1017/S0016672300033711 Published online by Cambridge University Press

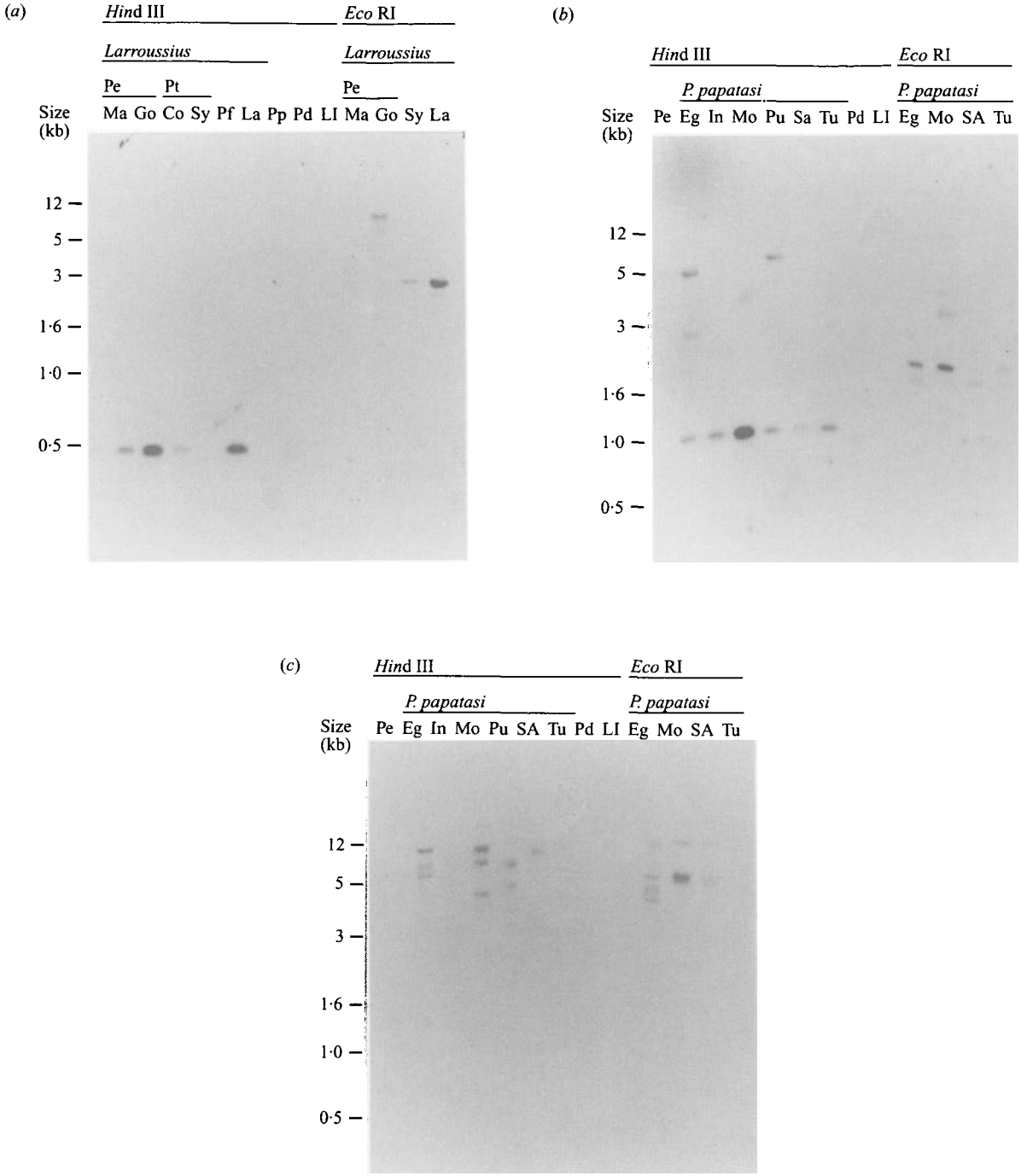

Fig. 2. (*a*) Hybridization of R4Ppern(RT) to a Southern blot of genomic DNA of various species digested with *Hind* III or *Eco*R I restriction enzymes. Tracks are *Phlebotomus* (*Larroussius*) *perniciosus* (Pe) from Madrid (Ma) and Gozo (G); *P.* (*L.*) *tobbi* (Pt) from Corfu (Co) and Syria (Sy); *P.* (*L.*) *perfiliewi* (Pf) from Corfu; *P.* (*L.*) *langeroni* (La) from Egypt; *P.* (*Phlebotomus*) *papatasi* (Pp) from Spain; *P.* (*P.*) *duboscqi* (Pd) from Senegal and *Lutzomyia* (*Lutzomyia*) *longipalpis* (Ll) from Brazil (approximately 1 μg DNA/track). (*b*) Hybridization of R4Ppapa (RT) to a Southern blot of genomic DNA of various species and strains digested with *Hind* III and *Eco*R I restriction enzymes. Tracks are *Phlebotomus* (*Larroussius*) *perniciosus* from Madrid (Pe); *P.* (*Phlebotomus*) *papatasi* from Egypt (Eg), India (In), Mojacar (Mo), Puna (Pu), Saudi Arabia (SA), and Tunisia (Tu); *P.* (*P.*) *duboscqi* from Senegal (Pd) and *Lutzomyia* (*Lutzomyia*) *longipalpis* from Brazil (Ll). (*c*) Hybridization of R5Ppapa (RT) to the same blot as (*b*). In each case, hybridization was in 50% formamide, 5 × SSC, 1% SDS, 300 μg/ml salmon sperm DNA, and 5 × Denhardt's solution, overnight at 42 °C. The filter was washed at 2 × SSC, 0·5% SDS at 65 °C and exposed to X-ray film for 5 days at −80 °C.

https://doi.org/10.1017/S0016672300033711 Published online by Cambridge University Press

Table 1. *Hybridization of phlebotomine reverse transcriptase-like sequences from various retroelements to Southern-blotted genomic DNA extracted from various species and strains*

Source of probe...	*P. perniciosus*			*P. tobbi*		*P. papatasi*					*P. duboscqi*	*L. longipalpis*
	R1	R3	R4	R1	R2	R1	R2	R3	R4	R5	R1	R1
Genomic DNA												
Genus *Phlebotomus*												
Subgenus *Larroussius*												
P. perniciosus												
Gozo	+++	++	++	−	+++	−	−			−		
Madrid	+++	+++	+	−	+++	−	−	−	−	−	−	−
P. tobbi												
Corfu	+	−	+++	+++	+					−		
Syria	++	−	+	++	++	−	−			−		−
P. perfiliewi	+	−	−	−	+	−	−			−		−
P. langeroni	−	−	+++	−	−	−	−			−		−
Subgenus *Phlebotomus*												
P. papatasi												
Egypt						+++	+	++	++	++	−	
India (C)						++	+	+	+	−	++	
India (P)						++	+	+++	++	+	++	
Spain	−	−	−	−	−	+++	++	++	+++	++	+	−
Saudi Arabia						++	++	+	+	+	++	
Tunisia						F	F	F	++	−	+	
P. duboscqi	−	−	−	−	−	F	+	−	F	−	+	−
Genus Lutzomyia												
L. longipalpis	−	−	−	−	−	−	−	−	−	−	?	+++

The more '+' signs the stronger the signal; 'F' means faint; '−' means no signal; spaces indicate not tested. The hybridization was in 50% formamide, 5 × SSC, 1% SDS, 5 × Denhardt's solution, 300 μg/ml of salmon sperm DNA overnight at 42 °C. Filters were then washed three times in 2 × SSC, 0·5% SDS at 65 °C. R2Ptobb was presumed to produce the same hybridization pattern as R1Ppern (96% similar). '?' indicates a faint smear rather than typical hybridization signal.

ologous to this RT of *P. perniciosus* (R1Ppern) were sought in other populations and species by screening PCR libraries with an RNA probe – an *in vitro* transcription product of recombinant pBS(+) RT7 (Stratagene; Booth *et al.* 1994). Hybridization was in Church buffer (Church & Gilbert, 1984): at 65 °C for targets from *P. perniciosus*, *P. tobbi*, *P. papatasi* and *P. duboscqi*, followed by washes at 65 °C in 2 × SSC, 0·1% SDS (3 × 30 min) and 0·1 × SSC, 0·1% SDS (2 × 15 min); but at 55 °C for *L. longipalpis*, with post-hybridization washes at 55 °C in 2 × SSC, 0·1% SDS (3 × 30 min) and 0·5 × SSC, 0·1% SDS (2 × 15 min). Autoradiography was carried out at −70 °C overnight, using FUJI-X film and intensifying screens. Dideoxy sequencing of single-strand DNA from the M13 tg130 recombinants was performed by standard methods using the Sequenase 2.0 kit and 6% polyacrylamide gels (USB Corp.) (Sambrook *et al.* 1989).

(ii) *Alignment and comparisons of RT sequences*

Alignment was based on that derived for RTs of non-LTR RTPs by Xiong & Eickbush (1990). Anchored by a conserved Q-GF motif in a peptide domain absent in retroviruses and LTR RTPs, it was unambiguous except for a 1–2 amino acid insertion in three sequences (Fig. 1; Booth *et al.* 1994). RT relationships were determined with Phylogenetic Analysis Using Parsimony (PAUP) software (Swofford, 1991) and Neighbor-joining algorithms (Saitou & Nei, 1987).

The GenBank accession numbers for the nucleotides sequences described here are: R1Ppern, L25911; R3Ppern, L28823; R4Ppern, L29043; R1Ptobb, L28922; R2Ptobb, L38254; R1Ppapa, L28921; R2Ppapa, L28924; R3Ppapa, L28925; R4Ppapa, L29043; R5Ppapa, L28927; R1Pdubo, L38255 and R1Llong, L28923.

(iii) *Distribution and abundance of RT sequences*

Southern hybridization was used: to identify species-specific restriction fragments (based on direct sequencing), to estimate sequence abundance and to indicate genomic distribution. Genomic DNA was extracted from sandflies of various species, restriction digested in 1 μg samples (using Boehringer Mannheim endonucleases), fractionated in agarose gels of appropriate concentrations, and Southern blotted on to GeneScreen Plus hybridization membranes (DuPont) before sequential hybridization to different radiolabelled probes and autoradiographic detection (see above; Booth *et al.* 1994).

For the PCR, sandflies were all derived from laboratory colonies established in the UK at: Imperial College at Silwood Park, the Natural History Museum, and the London School of Hygiene and Tropical Medicine. *P.* (*Larroussius*) *perniciosus* was collected in 1988 near Madrid, Spain; the Gozo strain of this species came from the island of Gozo (Malta) in 1988; *P.* (*L.*) *tobbi* came in 1990 from Syria; *P.* (*Phlebotomus*) *papatasi* came in 1990 from Mojacar, Spain; *P.* (*P.*) *duboscqi* came in 1988 from Senegal; and *Lutzomyia* (*Lutzomyia*) *longipalpis* came in 1988 from Belem, Brazil. For Southern blots of genomic DNA, additional UK laboratory populations were studied. *P. tobbi* was collected from Corfu in 1987: *P.* (*Larroussius*) *perfiliewi* was from Corfu in 1987; *P.* (*Larroussius*) *langeroni* was from Egypt in 1989; *P. papatasi* strains from Egypt and India were from colonies long-established in Cairo; whilst those from India (Pune), Saudi Arabia and Tunisia were established in the UK in 1984–6.

3. Results

(i) *Source, abundance and distribution of RT sequences*

The following RT fragments were isolated and sequenced from *P. perniciosus* (R Ppern), *P. tobbi* (R Ptobb), *P. papatasi* (R Ppapa), *P. duboscqi* (R Pdubo) and *L. longipalpis* (R Llong) using these primer pairs: YRPIS/AYADD – R1Ppern (four clones from Madrid colony), R1Ptobb (one clone from Syrian colony), R1Ppapa (two clones from Mojacar colony), R2Ppapa (one clone from Mojacar colony), R5Ppapa (one clone from Mojacar colony), R1Pdubo (one clone), R1Llong (one clone); VFSKIFE/KAFDTVC – R1Ppern (two clones from Madrid colony), R3Ppern (one clone from Madrid colony), R3Ppapa (one clone from Mojacar colony), R4Ppapa (one clone from Mojacar colony); VFSKIFE/AYADD – R1Ppern (two clones from Madrid colony; one clone from Gozo colony), R4Ppern (one clone from Gozo colony), R2Ptobb (one clone from Syria colony). All clones had open reading frames (ORFs) in frame with the forward primer (Booth, 1993; Booth *et al.* 1994).

Hybridization of the characterized RT sequences to Southern blots of phlebotomine genomic DNA (of the five species of origin, and of *P.* (*Larroussius*) *langeroni* and *P.* (*L.*) *perfiliewi*) indicated typical characteristics of non-LTR RTPs: variable hybridization patterns among both intraspecific populations and closely related species, and absence of hybridization to DNA of phlebotomines not closely related to the species of origin, i.e. in a different subgenus (Fig. 2, Table 1; Booth, 1993; Booth *et al.* 1994). All RTs were likely to be from multicopy families, since hybridization signal varied between different strains and species (Fig. 2*a*, *b*). Small fragment sizes would be internal to the RT gene and highly likely to be conserved in closely related taxa – hence no size variation was detected. Larger fragments, in which the restriction enzymes would have cut outside the elements in the flanking DNA, were of more than one size (Fig. 2*c*). In addition, the copy number of R1Ppern has previously been determined as 32/haploid genome by hybridization to serially diluted genomic DNA (Booth, 1993). Hybridization signal for this RT with its source species was not stronger than for the other RTs. Smaller fragments of species-specific sizes were identified from the sequencing data, and these were detected on Southern blots of genomic DNA in order to check the origin of the longer RT sequences amplified (R1Ppern, R2Ptobb, R1Ppapa, R5Ppapa and R1Llong, e.g. Fig. 3)

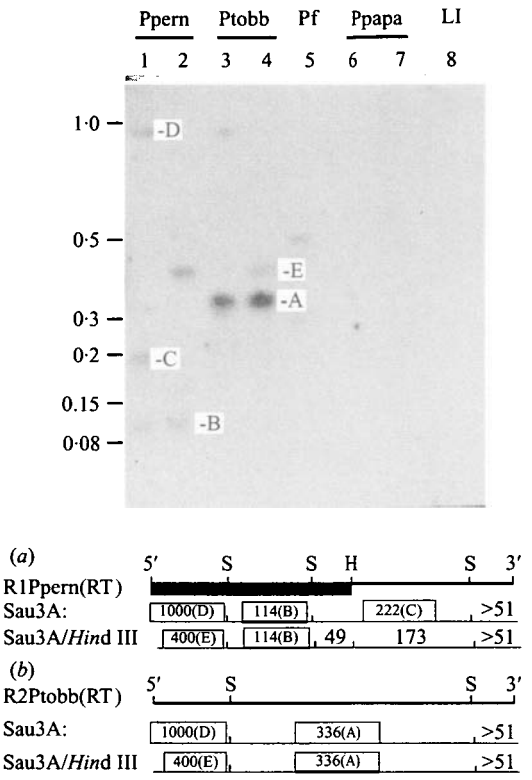

Fig. 3. Hybridization of R1Ppern (RT) to a Southern blot of genomic DNA of various species. The DNA has been digested with selected restriction enzymes. Tracks are *Phlebotomus perniciosus* digested with *Sau*3A (1), *Sau*3A and *Hind* III (2); *P. tobbi* digested with *Sau*3A (3), *Sau*3A and *Hind* III (4); *P. perfiliewi* digested with *Hae* 3 and *Hinf* 1 (5); *P. papatasi* digested with *Hae* 3 and *Hinf* I (6), with *Sau*3A (7); ad *Lutzomyia longipalpis* digested with *Sau*3A (8). Hybridization was in 50% formamide, 5 × SSC, 0·5% SDS, 300 μg/ml salmon sperm DNA, and 5 × Denhardt's solution, overnight at 42 °C. The filter was washed in 2 × SSC, 0·5% SDS at 65 °C and exposed to X-ray film for 3 days at −80 °C. (*a*) Restriction map of R1Ppern (RT) and genomic DNA of *P. perniciosus*. (*b*) Restriction map of R2Ptobb and genomic DNA of *P. tobbi*. S, *Sau*3A; H, *Hind* III; black box, probe sequence; white boxes, probable fragment detected. A *Hind* III site is 600 bp 5′ of the internal *Hind* III site (Booth *et al.* 1994). Hence the lengths of fragments D and E can be deduced.

https://doi.org/10.1017/S0016672300033711 Published online by Cambridge University Press

Table 2. % *Identity of reverse transcriptase (RT) predicted amino acid sequences from PCR products to other RT sequences between primers VFSKIFE and KAFDTVC*

Element	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. R1Ppern	*	36	44	49	99	54	43	54	43	62	39	46	36	38	18	30	16	2
2. R3Ppern		*	31	33	38	41	36	39	36	34	36	39	16	20	13	21	16	2
3. R4Ppern			*	34	44	43	39	46	36	43	39	39	23	23	15	23	16	2
4. R1Ptobb				*	51	43	38	44	57	41	43	41	25	21	16	33	25	5
5. R2Ptobb					*	56	43	54	44	62	39	48	36	38	18	30	18	2
6. R1Ppapa						*	49	52	38	49	33	51	30	30	20	33	18	3
7. R2Ppapa							*	56	38	48	33	54	26	31	20	28	15	3
8. R3Ppapa								*	39	57	43	57	34	36	20	31	18	3
9. R4Ppapa									*	38	36	46	26	18	15	28	23	7
10. R5Ppapa										*	39	43	31	30	21	30	15	2
11. R1Pdubo		Average similarity of phlebotomine									*	41	25	20	18	25	18	2
12. R1Llong		Elements with 13–18										*	33	34	18	33	23	7
13. F					28								*	31	25	18	16	2
14. I					28									*	23	25	16	2
15. R2Bm					18										*	25	15	11
16. R1Dm					29											*	31	8
17. Pt1-An					18												*	10
18. Copia					3													*

F, I, R1Dm are from one branch of non-LTR retrotransposons, R2Bm from the other; copia is an LTR retrotransposon; pt1-An is a group II intron. Numbers are percentage identical residues. (% identity is [the number of identical residues/total number of residues] × 100).

(ii) *Relationships among RT sequences*

The alignment of the predicted amino acid sequences of the DNA fragments characterized between primers VFSKIFE/KAFDTVC is shown in Fig. 1, and from this an identity matrix was derived (Table 2). Clearly, quantitatively, there is no discontinuous variation between phlebotomine and *D. melanogaster* non-LTR RT sequences, with the amino acid sequence identity showing a range of 31–62% among phlebotomine RTs, and 16–38% between phlebotomine RTs and the most similar RTs of *D. melanogaster* (from the F, I and ribosomal R1 elements).

Among these insect non-LTR RTPs, 23 out of 61 positions in this RT region have amino acid residues that are conserved physico-chemically, but in only four of these positions are the residues invariant among all the sequences from phlebotomines and most from *D. melanogaster* (I, F, G, Jockey, R1 and R2 elements) and mosquitoes (Juan-A, Juan-C, Q and T1 elements). These are in the Q-GF motif, characteristic of non-LTR RTPs, and the D that is invariant in all retroelements (Fig. 1; Xiong & Eickbush, 1990; Agarwal *et al.* 1993; Besansky *et al.* 1994; Booth *et al.* 1994). Phlebotomine RTs have five other positions invariant for amino acid residues and this suggests they are monophyletic (Fig. 1). Monophyly or polyphyly could not be demonstrated using PAUP: in some trees the sequences from F and Jockey elements fell within a clade containing the phlebotomine RTs most dissimilar to R1Ppern, but it is probably impossible unambiguously to determine the relationships among sequences with this high level of divergence (Fig. 4; Swofford & Olsen, 1990).

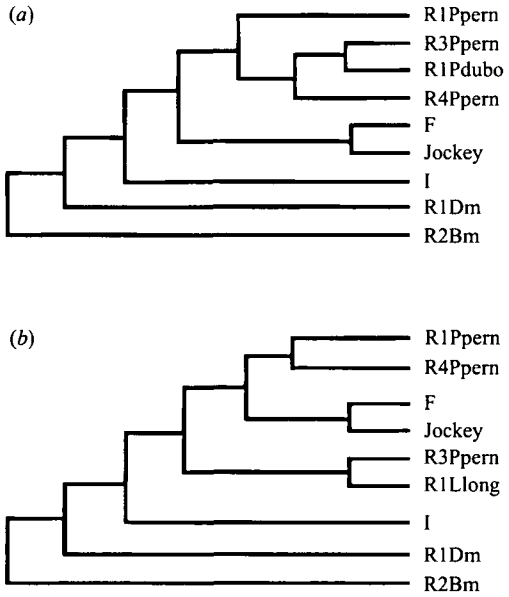

Fig. 4. Phylogenetic trees derived from RT amino acids between primers VFSKIFE and AYADD (using PAUP (Swofford, 1991)). All trees are branch and bound searches and the single most parsimonious tree is shown. (*a*) Comparison of the most divergent phlebotomine RTs with non-phlebotomine non-LTR retrotransposon RTs indicates a monophyletic clustering of phlebotomine RTs. Shortest tree found = 267, consistency index (CI) = 0·921, homoplasy index (HI) = 0·079, CI excluding uninformative characters = 0·880, HI excluding uninformative characters = 0·120, retention index (RI) = 0·604, rescaled consistency index (RC) = 0·556. (*b*) Comparing a different selection of phlebotomine RTs, a tree indicating polyphyly cf. non-phlebotomine RTs was derived. Tree length = 261, CI = 0·9001; HI = 0·100; CI excluding uninformative characters = 0·878; HI excluding uninformative characters = 0·122; RI = 0·649; RC = 0·584.

https://doi.org/10.1017/S0016672300033711 Published online by Cambridge University Press

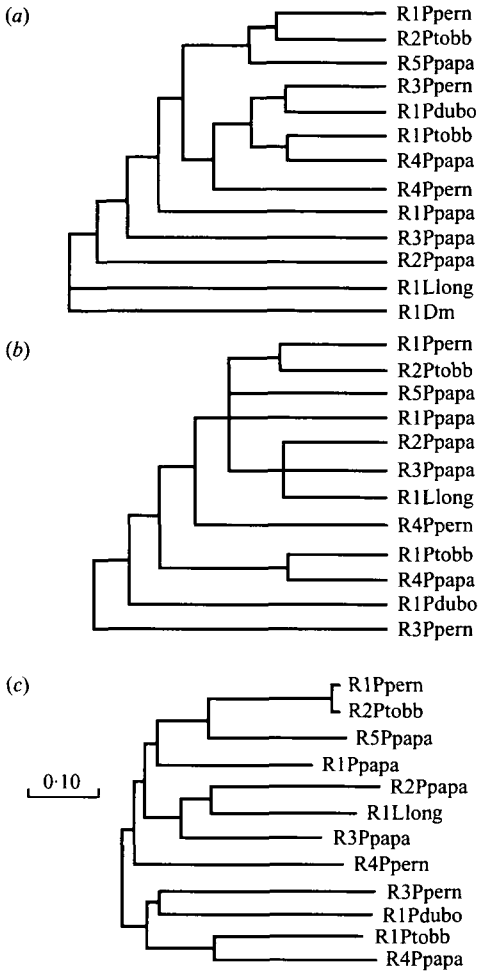

Fig. 5. Phylogenetic trees derived from RT amino acids between primers VFSKIFE and KAFDTVC determined (*a*) using PAUP (Swofford, 1991) branch and bound search for all phlebotomine RTs using R1Dm of *Drosophila melanogaster* as the outgroup, the single most parsimonious tree is shown: tree length (TL) = 271, consistency index (CI) = 0·797, homoplasy index (HI) = 0·203, CI excluding uninformative characters = 0·789, HI excluding uninformative characters = 0·211, retention index (RI) = 0·542, rescaled consistency index (RC) = 0·432. (*b*) As above, but using R3Ppern as the outgroup and excluding R1Dm. Strict consensus of the four most parsimonious trees is shown: TL = 235 CI = 0·787, HI = 0·231, CIE = 0·774, HIE = 0·226, RI = 0·541, RC = 0·426. (*c*) As for (*b*), but using the Neighbor-joining algorithm (Saitou & Nei, 1987) to determine the most parsimonious tree. The tree is unrooted. Branch lengths are proportional to the number of differences between the RTs.

Phylogenetic analyses were carried out on all 12 phlebotomine RT amino acid sequences, with PAUP using *D. melanogaster* R1 (R1Dm; Fig. 5*a*). or R3Ppern (that most dissimilar to R1Ppern; Fig. 5*b*) as an outgroup, as well as with the Neighbor-joining method of Saitou & Nei (1987) (Fig. 5*c*). Two groupings were always supported: R1Ppern with R2Ptobb (99% similar), and R1Ptobb with R4Ppapa (57% similar; Table 2), and this was also true for trees generated by PAUP using nucleotide sequences (Booth, 1993). In addition, some of the equally parsimonious trees for amino acid sequences grouped R3Ppern and R1Pdubo (36% similar; Table 2). Variable nodes for R4Ppern, R1Ppapa, R2Ppapa, R3Ppapa and R1Llong produced most of the variation among trees (Booth, 1993). The absence or presence of physico-chemical conservation of residues at any position was not shown to be associated with synapomorphies. Thus for the 32 positions in which R1Ptobb and R4Ppapa had the same amino acid, nine were invariant for the phlebotomine RTs; six had a residue shared only by these two elements (i.e. synapomorphic in character), of which four were physico-chemically conserved; and, 17 shared the same residue with other phlebotomine elements, of which seven were physico-chemically conserved in phlebotomines (Fig. 1). Similarly for R3Ppern and R1Pdubo, two characters were synapomorphic, and residues in 13 other positions were shared with other phlebotomine RTs; six of the 15 residues were conserved physico-chemically.

(iii) *Homologues of R1Ppern*

R1Ppern (four independent clones; Madrid *P. perniciosus*), R1Ppapa and R2Ppapa were each isolated from single PCR-product libraries by selecting M13 phage of the white, recombinant phenotype. The R1Ppern RT (truncated to the first 274 bp) was then used to probe libraries of *P. tobbi* (Syria colony) *P. papatasi* (Mojacar colony), *P. duboscqi* and *L. longipalpis*, each constructed with single PCR products obtained with the same 'universal primers' YRPIS/AYADD; and the hybridization was at 55 °C (*Lutzomyia* library) and 65 °C (*Phlebotomus* libraries) and post-hybridization washes at 0·5 × SSC, 55 °C (*Lutzomyia* library) and 0·1 × SSC, 65 °C (*Phlebotomus* libraries) in order to detect RT sequences with maximum homology to R1Ppern. The RTs R1Ptobb, R5Ppapa, R1Pdubo and R1Llong were isolated and sequenced, but amino acid homologies were only 39–62%, much less than would be expected from orthologous sequences, at least for the closely related *P. tobbi*.

Primers were then designed to increase the specificity of the PCR reaction for phlebotomine RTs, particularly for R1Ppern-like sequences. The degeneracy of each of the 'phlebotomine specific primers' VFSKIFE and KAFDTVC was reduced at the 5′ ends to match R1Ppern (Booth, 1993). When the resulting PCR-product libraries were screened with an RNA probe derived from R1Ppern, the following were isolated: R1Ppern (two independent clones from Madrid colony), R1Ppern (one clone from Gozo colony), R4Ppern (one clone from Gozo colony) and R2Ptobb (one clone from Syria colony), using primers VFSKIFE/AYADD; and R1Ppern (three independent clones from Madrid colony), R3Ppern, R3Ppapa and R4Ppapa, using primers VFSKIFE/KAFDTVC. Greater specificity was not obviously achieved by

https://doi.org/10.1017/S0016672300033711 Published online by Cambridge University Press

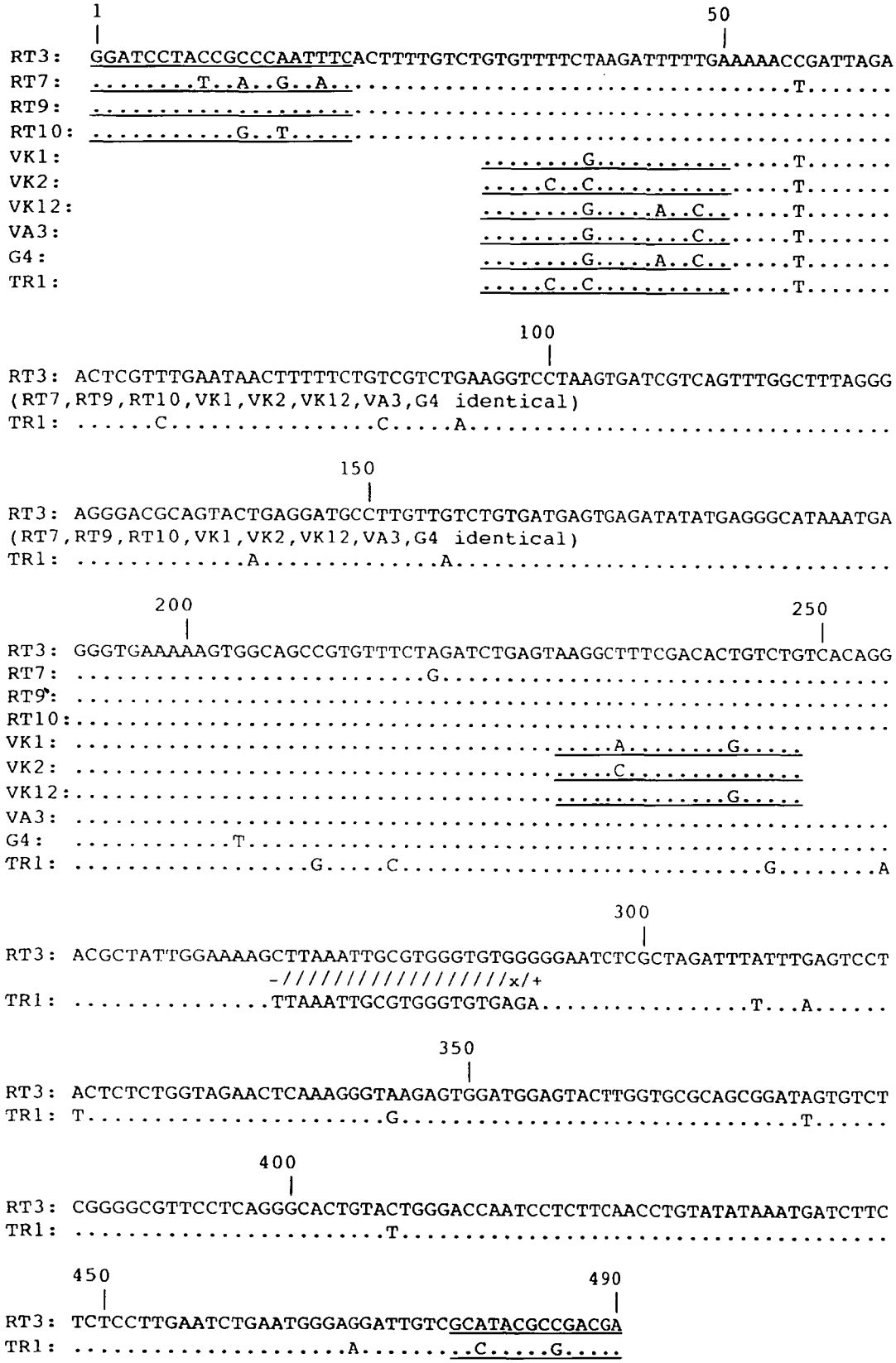

Fig. 6. Nucleotide sequences of R1Ppern clones from Madrid and Gozo populations of *P. perniciosus* and of a clone of the probable orthologous R2Ptobb derived from a Syrian *P. tobbi*. Primer sequences are underlined. RT clones were derived using the degenerate primers YRPIS-AYADD, VK clones using primers VFSKIFE-KAFDTV, the VA, G4 and TR clones using VFSKIFE-AYADD. All clones are from the Madrid strain of *P. perniciosus* except G4, which is from the Gozo strain, and TR, which is from *P. tobbi*. '.', Identical nucleotides to RT3 or VA3, except from nt 271–291, where an 'x' is a mismatch, '−' a deletion and '+' an insertion compared to VA3: this interpretation allows for three differences between the two sequences here rather than the 14 predicted from direct alignment.

https://doi.org/10.1017/S0016672300033711 Published online by Cambridge University Press

using these new primer pairs, and only the R1Ppern clones and R2Ptobb can be considered as orthologous (Fig. 1, Table 2; see Discussion). Only two nucleotide sites were polymorphic among the eight R1Ppern clones derived from Madrid *P. perniciosus* (Fig. 6). Counting from the first nucleotide of primer YRPIS (Fig. 6), five clones had a thymine (T) and three had a cytosine (C) at position 56 (giving rise to leucine or cysteine, respectively, in the deduced protein sequence); and seven clones had an adenine (A) and one a guanine (G) at position 219 (both giving leucine). The clone derived from the Gozo *P. perniciosus* had a T at position 56, an A at position 219, and a new substitution at position 204 (T for G, with no change of predicted amino acid). R2Ptobb had the nucleotides predominant in R1Ppern at positions 56, 204 and 219, but also had seven new synonymous substitutions out of 178 nucleotides between primer pair VFSKIFE/KAFDTVC. Between the primer pair VFSKIFE/AYADD, 19 differences out of 426 nucleotides were identified (Fig. 6).

4. Discussion

(i) *Diversity and targeted isolation of RT sequences*

Based on their low amino acid sequence homology (31–62%), 11 of the 12 RT fragments isolated each belongs to a different family of non-LTR RTPs. Three distinctive RTs were isolated from four independent PCR products of *P. perniciosus*, two from two independent PCR products of *P. tobbi*, five from three independent PCR products of *P. papatasi*, and one each from single PCR products of *P. duboscqi* and *L. longipalpis* (Fig. 1; Table 2). Such diversity is typical of non-LTR RTPs, e.g. the findings for *D. melanogaster*, culicine and anopheline mosquitoes and other higher eukaryotes (Jacubczak *et al.* 1991; Agarwal *et al.* 1993; Besansky *et al.* 1994).

We conclude that our PCR primers were 'generic' for RTs of sandfly non-LTR RTPs of the F/Jockey (*D. melanogaster*) type – all the phlebotomine RTs were monophyletic with the RTs of this type of element (Fig. 4; Xiong & Eickbush, 1990; Booth, 1993). The RT sequences isolated from any one species were not more closely related to each other than they were to those from other species. No increased specificity in amplification of RT families was achieved using primers designed to be more specific for R1Ppern. The RTs isolated in this manner were as varied as before. The divergence of most of these sequences is too great to be explained in terms of PCR amplification errors or by a high frequency of pseudogenes, although these are likely explanations for some of the variation among the nine isolates of R1Ppern. Many non-LTR RTP elements are pseudogenes, inactivated by 5′ truncation during reverse transcription and, perhaps, by a high nucleotide substitution rate resulting from reverse transcription errors, which for some reverse transcriptases can be 10^4 times greater than those for DNA polymerases (Gojobori & Yokoyama, 1985; Bucheton *et al.* 1986; Simonelig *et al.* 1988).

Orthologous elements are those which were present in a common ancestor, and diverged from the time of species divergence. The presence of putative orthologues in species of the same subgenus was indicated by hybridization (Table 1). In species related to *P. perniciosus*, in the subgenus *Larroussius*, RTs possibly orthologous to R1Ppern were indicated for *P. perfiliewi* and *P. tobbi*, but not *P. langeroni* (Table 1). The probable orthologous RT from *P. tobbi* has been isolated (R2Ptobb, Figs. 1 and 6). Fortunately, it has two restriction enzyme differences from R1Ppern. Digestion of genomic DNA from *P. tobbi* and *P. perniciosus* with these enzymes yields the expected fragment sizes for R1Ppern and R2Ptobb respectively, using R1Ppern as a probe (Fig. 3). This confirms that R2Ptobb was isolated from *P. tobbi*, and that the *P. tobbi* R1Ppern orthologue is consistent with the R2Ptobb sequence.

Similarly, orthologous RTs could be isolated from the other phlebotomine RTs, except for those that hybridized only to the species of origin (R3Ppern, R1Ptobb, R3Ppapa, R5Ppapa and R1Llong; Table 1). It should be possible to design PCR primers specific for each of the RT families described and, indeed, this has been achieved for R1Ppapa (J. R. Sackin & P. D. Ready, unpublished observations).

Several factors mask identification of orthologous RTs in more divergent taxa. Within each species, RTs from multiple families of non-LTR RTPs may have been amplified non-preferentially by the 'universal' and 'phlebotomine specific' PCR primers; turnover of non-LTR RTPs within evolutionary lines may have removed orthologous sequences from some species (perhaps increasing in likelihood with time of divergence); and relatively rapid sequence divergence could have disguised RT relationships (e.g. Daniels *et al.* 1990).

(ii) *Phylogenetic value of RT sequences*

To be useful phylogenetically, sequence variation should be such that: (1) sequence alignment is straightforward; (2) orthologous sequences are identifiable over the range of species targeted; (3) horizontal or reticulate evolution has minimal influence; (4) rates of change are comparable in different species lineages; and, for closely-related species, (5) sequence divergence is relatively rapid (Syvanen, 1987; Hillis & Moritz, 1990). For phlebotomine RT sequences, alignment is straightforward, and orthologous sequences can be targeted using specific primers/probes. However, on present evidence, these RTs have a species distribution that makes them appropriate only for phylogenetic analysis within subgeneric clades, and also for the identification of such monophyletic groups by absence/presence of hybridization with a

panel of RT probes (Table 1). We are currently investigating the value of RFLP 'finger prints' for identifying infraspecific populations; the variable hybridization patterns on Southern blots of genomic DNA suggest that many of the RTs are from transposons that are (or were recently) subject to flux in genomic position and copy number, e.g. Booth *et al.* (1994, Fig. 3*b*).

RTs chosen for both tests must display vertical, not horizontal, evolution. The hybridization data for R1Ppern and R2Ptobb probes (Table 1) suggest an order of relatedness of *P. perniciosus*, *P. tobbi* (until recently treated as a subspecies of *P. perniciosus*), *P. perfiliewi* and *P. langeroni*, which agrees with the phylogeny based on an intuitive appraisal of morphological characters (Lewis, 1982) but not with the phenetic analysis of Rispail (1990). Only the hybridization pattern of R4Ppern was consistent, in part, with horizontal transfer: hybridization signals were more intense for the more centrally located Mediterranean populations of *P. perniciosus* (Gozo island, near Malta), *P. tobbi* (Corfu island, Greece) and *P. langeroni* (Egypt) than they were for the peripheral populations of *P. perniciosus* (Madrid) and *P. tobbi* (Syria), and the restriction fragment size was conserved across species, as would be expected with a sequence whose divergence time was less than the host species divergence time (Fig. 2).

Sequences that could be considered as orthologous were obtained only for one RT family (R1Ppern/R2Ptobb), which was isolated from the Madrid and Gozo populations of *P. perniciosus* as well as from the Syrian population of *P. tobbi* (Table 2, Fig. 6). The island of Gozo is thought to have been surrounded by sea for more than 5 million years (my), similar to the expected time of divergence of *P. perniciosus* (found in the western Mediterranean Basin) and the parapatric *P. tobbi* (with a more easterly range) (Attenborough, 1987; Marchais, 1992). The R1Ppern-like sequences from Madrid *P. perniciosus* and Syrian *P. tobbi* differed in 19 out of 426 nucleotide positions (or 7 out of 178 between primers VFSKIFE and KAFDTVC; Fig. 6), but there was only one non-polymorphic difference between the Madrid and Gozo populations of *P. perniciosus*. This inconsistency would be explained if *P. perniciosus* had arrived relatively recently on Gozo, perhaps being transported by boat or assisted by winds. The fragment of RT sequenced contains some of the most conserved domains in non-LTR RTPs (Xiong & Eickbush, 1990) and, therefore, the failure to detect RT sequences of greater similarity across subgenera strongly suggests that the families of sequences we isolated either have a fast, or non-uniform, turnover rate (of excision and/or nucleotide substitution), perhaps accelerating inactivation, and/or they invaded different sandfly lineages relatively recently.

The RT sequence (R1Llong) isolated from the American phlebotomine *L. longipalpis* is phylogenetically most similar to R2Ppapa (54% amino acid similarity) and R3Ppapa (57% amino acid similarity) from the Old World *P. papatasi* (Fig. 5*b*, *c*; Table 2). The Old and New World phlebotomines are believed to have diverged 80–120 my ago, before the end of the Cretaceous period; species of the same genus (extant, or known from fossils of less than 120 my) have not been found in both hemispheres (Hennig, 1972; Lewis, 1982). This could indicate that many of the phlebotomine RT families arose from a common ancestor before isolation of the sandfly faunas by the Atlantic Ocean, because from the Old World species investigated only two RT families were clearly closer than the New World–Old World pairs mentioned above (these are R1Ptobb and R4Ppapa, which showed 57% amino acid similarity and were placed in the same clade (Fig. 5, Table 2)). This conclusion assumes that, for different RTPs and sandfly hosts, rates of divergence are similar through time for the RT domains sequenced. There is some evidence for this in mammals and *Drosophila* (Hutchison *et al.* 1989; Booth, 1993). Nevertheless, within different lineages of phlebotomine RTs, these domains might well have diverged by 38–69% in less than 80–120 my and then approached saturation for nucleotide substitutions because of functional constraints. The hosts of all the phlebotomines in this study are mammals, and it is an interesting possibility that many non-LTR RTP lineages could have arisen contemporaneously with the diversification of *Phlebotomus* and *Lutzomyia* species following the rapid radiation of mammals at the beginning of the Tertiary era (65 my ago).

We thank Jo Testa and Penny Allen for excellent technical assistance in the Natural History Museum (NHM), Bob and Mireille Killick-Kendrick (Biology Department, Imperial College at Silwood Park), Richard Lane (London School of Hygiene and Tropical Medicine) and the US Naval Research Unit (Cairo) for sandfly specimens. We also thank David Williams for assistance with phylogenetic analysis. D.R.B. was in receipt of an NHM studentship, and this work formed part of his PhD thesis for the University of London.

References

Agarwal, M., Bensaadi, N., Salvado, J.-C., Campbell, K. & Mouches, C. (1993). Characterization and genetic organization of full-length copies of a LINE retroposon family dispersed in the genome of *Culex pipiens* mosquitoes. *Insect Biochemical and Molecular Biology* **23**, 621–629.

Attenborough, D. (1987). *The First Eden: The Mediterranean World and Man*. London: Collins.

Besansky, N. J., Bedell, J. A. & Mukabayire, O. (1994). Q: a new retrotransposon from the mosquito *Anopheles gambiae*. *Insect Molecular Biology* **3**, 49–56.

Booth, D. R. (1993). Transposon sequences and evolution in phlebotomine sandflies. PhD thesis, University of London.

Booth, D. R., Ready, P. D. & Smith, D. F. (1994). Isolation of non-LTR retrotransposon reverse transcriptase-like sequences from phlebotomine sandflies. *Insect Molecular Biology* **3**, 89–96.

Bucheton, A., Simonelig, M., Vaury, C. & Crozatier, M. (1986). Sequences similar to the I transposable element

involved in I-R hybrid dysgenesis in *D. melanogaster* occur in other *Drosophila* species. *Nature* **322**, 650–652.
Church, G. M. & Gilbert, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Sciences, USA* **81**, 1991–1995.
Daniels, S. B., Chovnick, A. & Boussy, I. A. (1990). Distribution of *hobo* transposable elements in the genus *Drosophila*. *Molecular Biology and Evolution* **7**, 589–606.
Doolittle, R. F., Feng, D., Johnson, M. S. & McLure, M. A. (1989). Origins and evolutionary relationships of retroviruses. *Quarterly Review of Biology* **64**, 1–30.
Gojobori, T. & Yokoyama, S. (1985). Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. *Proceedings of the National Academy of Sciences, USA* **82**, 4198–4201.
Hennig, W. (1972). Insektfossilien aus der unteren Kreide. *Stuttgarter Beiträge zur Naturkunde* **241**, 1–67.
Hillis, D. M. & Moritz, C. (1990). *Molecular Systematics*. Sunderland, MA: Sinauer Associates.
Hutchison, C. A., Hardies, S. C., Loeb, D. D., Shehee, W. R. & Edgell, M. H. (1989). LINES and related retroposons: long interspersed repeated sequences in the eucaryotic genome. In *Mobile DNA* (Ed. D. Berg and M. Howe). Washington, DC: American Society of Microbiology.
Jacubczak, J. L., Burke, W. D. & Eickbush, T. H. (1991). Retrotransposable elements R1 and R2 interrupt the rDNA genes of most insects. *Proceedings of the National Academy of Sciences, USA* **88**, 3295–3299.
Killick-Kendrick, R. (1990). Phlebotomine vectors of the leishmaniases: a review. *Medical and Veterinary Entomology* **4**, 1–24.
Lehninger, A. L. (1970). *Biochemistry*. New York: Worth Publishers.
Lewis, D. J. (1982). A taxonomic review of the genus *Phlebotomus* (Diptera: Psychodidae). *Bulletin of the British Museum of Natural History* (*Entomology*) **45**, 121–209.
Marchais, R. (1992). Spéciation et vicariance chez les *Larroussius* du groupe *perniciosus* (Diptera: Psychodidae). PhD thesis, University of Reims.
Rispail, P. (1990). Approche phénétique et cladistique du genre *Phlebotomus* Rondani & Berte, 1840 (Diptera: Psychodidae). PhD thesis, University of Montpellier.
Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.
Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory.
Simonelig, M., Bazin, C., Pelisson, A. & Bucheton, A. (1988). Transposable and nontransposable elements similar to the I factor involved in inducer-reactive (IR) hybrid dysgenesis in *Drosophila melanogaster* coexist in various *Drosophila* species. *Proceedings of the National Academy of Sciences, USA* **85**, 1141–1145.
Swofford, D. L. (1991). PAUP: Phylogenetic analysis using parsimony, version 3.0s. Compute program distributed by the Illinois Natural History Survey, Champaign, Illinois.
Swofford, D. & Olsen, G. J. (1990). Phylogeny reconstruction. In *Molecular Systematics* (Ed. D. M. Hillis and C. Moritz). Sunderland, MA: Sinauer Associates.
Syvanen, M. (1987). Molecular clocks and evolutionary relationships: possible distortions due to horizontal gene flow. *Journal of Molecular Evolution* **26**, 16–23.
Xiong, Y. & Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO Journal* **9**, 3353–3362.