33 Saha, B. *et al.* (1998) CTLA4 (CD152) modulates the Th subset response and alters the course of experimental *Leishmania major* infection. *Eur. J. Immunol.* 28, 4213–4220
34 Heinzel, F.P. and Maier, R.A.Jr. (1999) Interleukin-4-independent acceleration of cutaneous leishmaniasis in susceptible BALB/c mice following treatment with anti-CTLA-4 antibody. *Infect. Immun.* 67, 6454–6460
35 Leach, D.R. *et al.* (1996) Enhancement of antitumor immunity by CTLA-4 blockade. *Science* 271, 1734–1736
36 Murphy, M.L. *et al.* (1998) Blockade of CTLA-4 enhances host resistance to the intracellular pathogen, *Leishmania donovani*. *J. Immunol.* 161, 4153–4160
37 Murphy, M.L. *et al.* (1997) B7-2 blockade enhances T cell responses to *Leishmania donovani*. *J. Immunol.* 159, 4460–4466
38 Kaye, P.M. *et al.* (1991) Differential production of Th1- and Th2-derived cytokines does not determine the genetically controlled or vaccine-induced rate of cure in murine visceral leishmaniasis. *J. Immunol.* 146, 2763–2770
39 Wilson, M.E. *et al.* (1998) The importance of TGF-β in murine visceral leishmaniasis. *J. Immunol.* 161, 6148–6155
40 Kaye, P.M. *et al.* (1994) Deficient expression of co-stimulatory molecules on *Leishmania*-infected macrophages. *Eur. J. Immunol.* 24, 2850–2854
41 Saha, B. *et al.* (1995) Macrophage-T cell interaction in experimental visceral leishmaniasis: failure to express costimulatory molecules on *Leishmania*-infected macrophages and its implication in the suppression of cell-mediated immunity. *Eur. J. Immunol.* 25, 2492–2498
42 Gomes, N.A. *et al.* (1998) Unresponsive $CD4^+$ T lymphocytes from *Leishmania chagasi*-infected mice increase cytokine production and mediate parasite killing after blockade of B7-1/CTLA-4 molecular pathway. *J. Infect. Dis.* 178, 1847–1851
43 Elloso, M.M. and Scott, P. (1999) Expression and contribution of B7-1 (CD80) and B7-2 (CD86) in the early immune response to *Leishmania major* infection. *J. Immunol.* 162, 6708–6715
44 Karandikar, N.J. *et al.* (1998) Tissue specific up-regulation of B7-1 expression and function during the course of murine relapsing experimental autoimmune encephalomyelitis. *J. Immunol.* 161, 192–199
45 Gomes, N.A. *et al.* (2000) TGF-β mediates CTLA-4 suppression of cellular immunity in murine kalaazar. *J. Immunol.* 164, 2001–2008
46 Gomes, N.A. *et al.* (2000) Early *in vitro* priming of distinct Th cell subsets determines polarized growth of visceralizing *Leishmania* in macrophages. *Int. Immunol.* 12, 1227–1233
47 Freire-de-Lima, C.G. *et al.* (2000) Uptake of apoptotic cells drives the growth of a pathogenic trypanosome in macrophages. *Nature* 403, 199–203
48 Mukhopadhyay, R. and Madhubala, R. (1993) Effect of a bis(benzyl)polyamine analogue, and DL-alpha-difluoromethylornithine on parasite suppression and cellular polyamine levels in golden hamster during *Leishmania donovani* infection. *Pharmacol. Res.* 28, 359–365
49 Jiang, Y. *et al.* (1999) Ornithine decarboxylase gene deletion mutants of *Leishmania donovani*. *J. Biol. Chem.* 274, 3781–3788
50 Lopes, M.F. *et al.* (2000) The macrophage haunted by cell ghosts: a pathogen grows. *Immunol. Today* 21, 489–494
51 Fitzpatrick, D.R. and Bielefeldt-Ohmann, H. (1999) Transforming growth factor β in infectious disease: always there for the host and the pathogen. *Trends Microbiol.* 7, 232–236
52 Omer, F.M. *et al.* (2000) Maintaining the immunological balance in parasitic infections: a role for TGF-β ? *Parasitol. Today* 16, 18–23

Aspartyl proteinase genes from apicomplexan parasites: evidence for evolution of the gene structure

Létitia Jean, Manyuan Long, John Young, Pierre Péry and Fiona Tomley

Aspartyl proteinases are a widely distributed family of enzymes. All vertebrate aspartyl proteinases share a conserved nine-exon gene structure, but in other organisms the structure of aspartyl proteinase genes varies considerably. The exon–intron patterns generally reflect phylogeny based on amino acid sequences. However, close comparison of these gene structures reveals some striking features, such as the conservation of intron positions and intron phases between aspartyl proteinases from nematodes and apicomplexans. Here, we discuss the implications of gene structure for the possible evolution of the aspartyl proteinase family, with particular reference to the plasmepsins of *Plasmodium falciparum* and eimepsin from *Eimeria tenella*.

Eukaryotic aspartyl proteinases (APs) are thought to have evolved from a common ancestor by gene duplication and fusion. This hypothesis is supported both by the three-dimensional structure of the enzymes, which consist of two homologous lobes[1], and by the organization of vertebrate genes for APs, which have two clusters of four exons[2]. The primordial APs might have been homodimers, similar to the enzyme that is found in retroviruses[1,3]. Less is known about the genomic organization of APs in lower eukaryotes. For example, within the protozoa, genomic sequences of APs have only been analysed for the *Plasmodium falciparum* plasmepsins, which have diverse exon–intron organizations[4,5] and for an eimepsin from *Eimeria tenella* (Ref. 6).

Evolutionary relationships in the AP gene family

Comparison of genes encoding APs from several organisms (yeast, fungi, nematodes, protozoa, plants and vertebrates) shows that, although many intron positions vary throughout the whole family, genes encoding enzymes from related classes of organisms share some structural similarities (Fig. 1). The only group with no introns in any of their AP genes are the yeasts, which have small genomes and have probably lost most of the introns that their ancestors had[7,8]. Unlike the vertebrate AP genes, which all have a nine-exon structure with perfect conservation of the exon–intron boundaries, AP genes from other groups of organisms have a variable number of exons: two to

 1471-4922/01/$ – see front matter PII: S1471-4922(01)02030-X

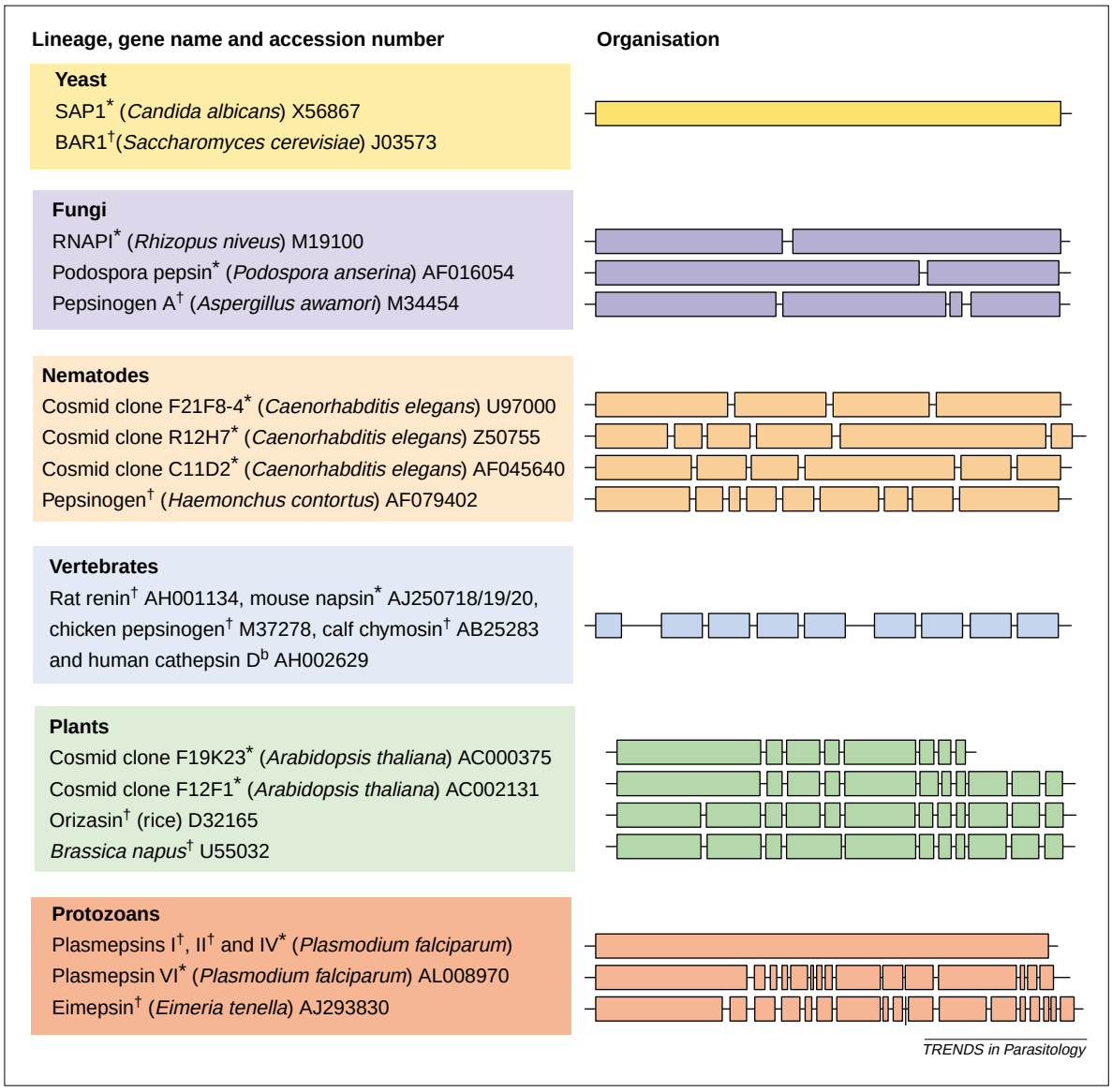


Fig.1. Comparison of the structure of the aspartyl proteinase (AP) family genes. The sequences have been aligned for maximal homology of amino acid sequences (PILEUP program in the GCG software suite)[27]. The exons are represented by the colored blocks. Sequences for plasmepsins I, II and IV were obtained from chromosome 14: locus 24. t00047, locus 24. t00046, contig>138 (www.tig.org). Symbols: *, the intron–exon boundaries were deduced from genomic sequence; †, the intron–exon boundaries were also verified experimentally .

Létitia Jean*
National Institute for Medical Research, Division of Parasitology, The Ridgeway, Mill Hill, London, UK NW7 1AA.
*e-mail: ljean@nimr.mrc.ac.uk

John Young
Fiona Tomley
Institute for Animal Health, Compton Laboratory, Newbury, UK RG20 7NN.

Manyuan Long
Dept Ecology and Evolution, The University of Chicago, 1101 East 57th Street, Chicago, IL 60637, USA.

Pierre Péry
Unité de Virologie et d'Immunologie Moléculaires, INRA, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France.

four in fungi, four to nine in nematodes, 8–12 in plants and 0–18 in protozoa.

The grouping of organisms obtained by comparing the genomic organization of the AP genes corroborates the amino acid similarity of the AP family (Fig. 2). An unrooted tree constructed by aligning the amino acid sequences of the AP family shows that the APs separate into six clusters of organisms. These clusters represent vertebrates, plants, fungi, yeast, nematodes and apicomplexan protozoa, and are the same as those defined by comparison of the genomic organization. Conservation of exon–intron boundaries occurs predominantly in the 5′ half of the AP genes studied (Fig. 3). According to the theory of AP gene duplication and fusion[1], this is consistent with modifications to the intron pattern (intron loss or addition) occurring after gene duplication and suggests that the two halves of the gene have evolved differently.

Two types of AP gene structure within a single organism

Four different aspartyl proteinases, termed plasmepsins, have been described in *P. falciparum* and several more have now been detected within the *P. falciparum* genome (see forthcoming article in *Trends in Parasitology* by Coombs *et al.*). The genes encoding *P. falciparum* plasmepsins (*Pf*PMs) I, II and IV do not have introns, whereas the gene encoding *Pf*PMIII (Ref. 4), now renamed *Pf*PMVI, is organized into 15 exons. In the related apicomplexan *E. tenella*, only one AP gene (encoding eimepsin) has been described so far and this has 18 exons[6].

The *Pf*PMVI and eimepsin genes are exceptional in comparison with other AP genes, and other apicomplexan genes in general, in that they have many introns, including some that have apparently

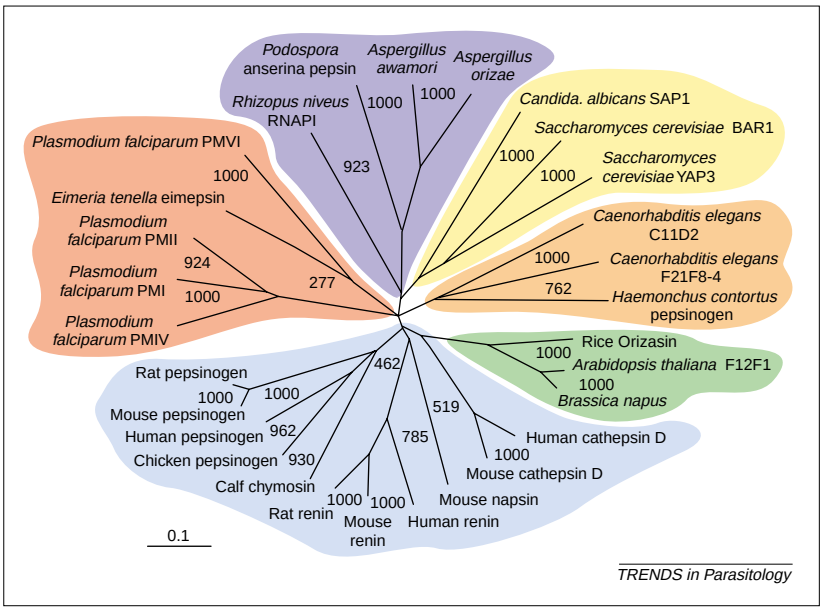


Fig. 2. Unrooted tree based on amino acid similarity of the aspartyl proteinase (AP) family. Amino acid sequences of APs with characterized genomic organizations were aligned with the program ClustalX (Ref. 25) using a neighbor-joining method. The unrooted tree was drawn using TreeView (Ref. 26). Numbers below the nodes are the number of replicates out of 1000 in which the clade was represented in a bootstrap analysis. Groups of related organisms are shown on a shaded background: protozoan parasites on red, fungi on purple, yeast on yellow, nematodes on orange, plants on green and mammals on blue. Branch lengths are scaled to average protein similarity and represent expected numbers of amino acid substitutions per site. The scale bar shows a branch length that would represent a 10% difference between the protein sequences.

evolved since the separation of *E. tenella* and *P. falciparum*. The division of the apicomplexan AP genes into two distinct groups is reinforced by phylogenetic analysis of their amino acid sequences. Thus, *Pf*PMs I, II and IV form a group that is distinct from that of *Pf*PMVI and eimepsin (Fig. 2). In a bootstrap analysis, trees generated from 1000 random subsets of the input data were used to test the reliability of this tree topology. The numbers between each node (Fig. 2) show the number of generated trees, out of 1000, in which the same group of sequences branched from that node. This bootstrap analysis supports a model in which there are two distinct apicomplexan AP groups, each of which evolved within the Apicomplexa from a different ancestral AP gene (bootstrap values 1000). However, because the bootstrap value for the node between the branches containing the two groups is very low (277), this analysis does not provide support for a common ancestor of the two groups within the Apicomplexa. In other words, the two groups are no more closely related to each other than to any of the other APs from the diverse species shown in Fig. 2.

The observation that APs in *Plasmodium* have evolved into two groups with completely dissimilar genomic organizations demands explanation. Two special aspects of apicomplexan biology offer intriguing possibilities. It is well documented that apicomplexan parasites have, during the course of their evolution, acquired a phylum-specific endosymbiotic organelle (the apicoplast), which has its own genome, and that these parasites can also be infected by viruses[9–11]. Thus, one or other of the radically different structures of AP genes might have its origin in horizontal gene transfer following either of these events.

Haemoglobin digestion is a highly specialized function shared by *Plasmodium* and other apicomplexans, but not by *Eimeria*. Both *Pf*PMI and *Pf*PMII are already known to be involved in haemoglobin digestion, whereas *Pf*PMIV and *Pf*PMVI have been described more recently and no functions have been ascribed to them. From the gene structure and phylogeny, it is tempting to speculate that *Pf*PMIV might be involved in haemoglobin digestion like *Pf*PMI and *Pf*PMII, whereas *Pf*PMVI might have a completely different function. It might be that the multi-intron genes for eimepsin and *Pf*PMVI represent the archetypal apicomplexan AP gene, with the intron-less *Pf*PMI, II and IV genes having been acquired during the evolution of *Plasmodium* under selection for efficient digestion of haemoglobin (Box 1).

To date, we have no clues to the identity of a potential source for this horizontal gene transfer. Thus, it remains possible that the intron-less AP genes are the product of a rapid phase of evolution of a common ancestral apicomplexan AP gene during selection for the efficient use of haemoglobin. However, this scenario does not explain the complete lack of introns in the newly evolved genes. Indeed, introns might have a role in accelerating evolution by allowing exon shuffling. Thus, we prefer horizontal gene transfer as the model to explain the diversity of AP gene structure in these parasites. The determination of all AP gene structures from other apicomplexan parasites that do (*Theileria*, *Babesia*) or do not (*Toxoplasma*, *Cryptosporidium*, *Eimeria*) depend on haemoglobin digestion will provide a test of this model. Current genome sequencing projects will provide some of this information rapidly.

Ancient introns in AP genes from different organisms

Closer inspection of AP exon–intron boundaries reveals that some of these are shared between lineages: between vertebrates and the free-living nematode *Caenorhabditis elegans* (Fig. 3a); between apicomplexans, plants and *C. elegans* (Fig. 3b); and between apicomplexans and the parasitic nematode *Haemonchus contortus* (Fig. 3c). Within the apicomplexan sequences, intron 1 of eimepsin and the shared intron 3 of eimepsin and *Pf*PMVI are in identical positions to introns 2 and 4 of *H. contortus* AP, and the shared intron 2 of eimepsin/intron 1 of *Pf*PMVI is identical in position to intron 2 of *C. elegans* C11D2 and is shifted by just two nucleotides from introns in plants. Similarly, intron 7 of *Pf*PMVI is shifted by only one nucleotide from intron 5 of *H. contortus*.

Using the method proposed by Iwabe *et al.*[12], we have found it to be unlikely that the positional matches of the first and third introns of eimepsin

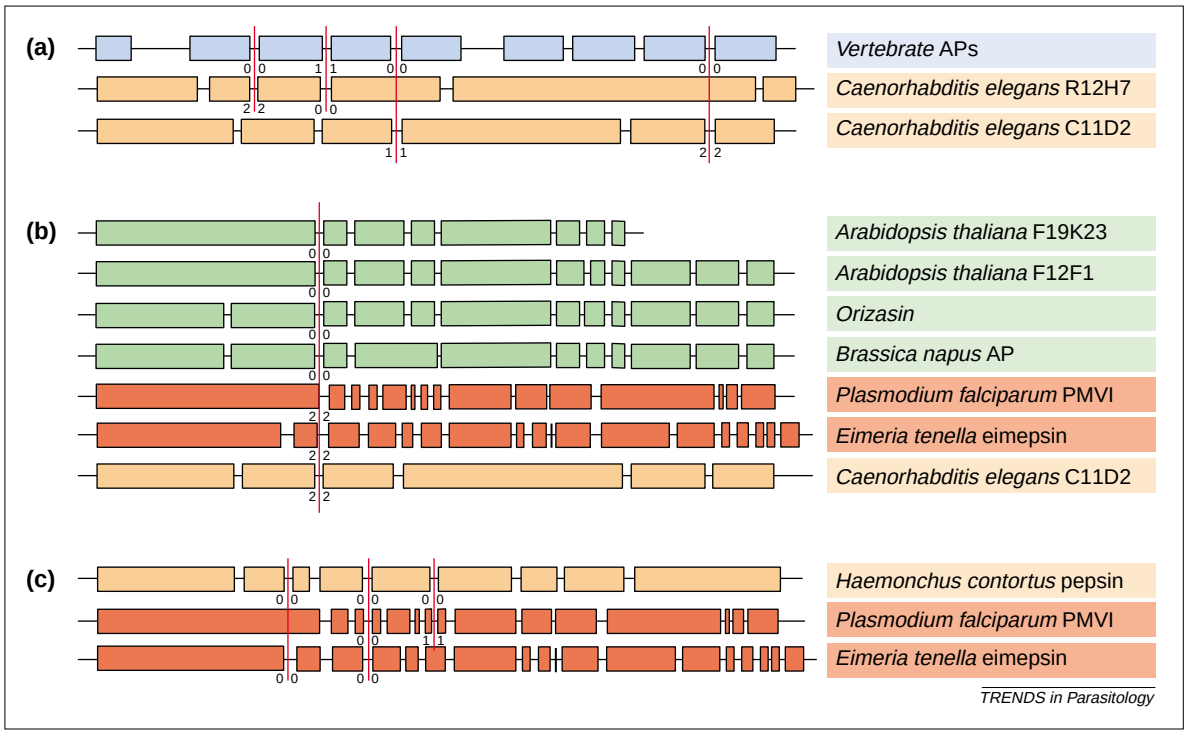


Fig. 3. Conservation of exon–intron boundaries between groups of organisms. Sequences have been aligned for maximal homology (using the PILEUP program[27]). Exons are represented by colored blocks. The intron phase is indicated below the sequences and the red lines show common introns between aspartyl proteinase (AP) genes from groups of organisms. Intron phases are defined as the relative positions of introns falling between codons (phase 0) or within a codon after the first (phase 1) or second (phase 2) nucleotides, respectively.
(a) Conserved exon–intron boundaries between the vertebrate AP and two *Caenorhabditis elegans* cosmid clones (C11D2 and R12H7).
(b) Conserved exon–intron boundaries between the apicomplexan AP genes (eimepsin and plasmepsin VI), the plant AP genes (*Arabidopsis thaliana* cosmid clones F19K23 and F12F1, orizasin, and *Brassica napus*) and *C. elegans* cosmid clone C11D2. (c) Conserved exon–intron boundaries between apicomplexan AP genes (eimepsin and plasmepsin VI) and *Haemonchus contortus* pepsinogen gene.

with the second and fourth introns of *H. contortus* are the result of random insertions (probability $P = 0.023$ that the number of shared introns is ≥2). Furthermore, because the positions of intron 2 of eimepsin and intron 1 of *Pf*PMVI are identical to that of intron 2 of *C. elegans* C11D2, we have also tested how significant the three positional matches between the nematodes and apicomplexans are. Between them, the apicomplexan sequences have introns in 20 different positions, and the nematode sequences have 13. The probability that the two groups have acquired three positional matches randomly is very low (probability $P = 0.002$ that the number of shared introns is ≥3).

A further inspection of the exon nucleotide sites that flank the positionally identical introns showed that the three 5′ and 3′ exon nucleotides flanking introns 1, 2 and 3 of eimepsin are GTA | GTC, GAA | AGG and CAA | CAG (' | ' indicates the positions of these introns). By contrast, the exon nucleotides flanking introns 2 and 4 in *H. contortus* are CAC | GTT and GTT | TAT and the exon nucleotides flanking intron 2 of *C. elegans* are TAA | AAA. Thus, there are no conserved patterns for the insertion of introns, which is a minimum assumption for a proto-splice model of intron insertion[13,14]. These analyses do not support a model of random intron insertion into these positions and suggest a common ancestral relationship of these three introns that pre-dated the divergence of the various protozoan species and the nematodes. Thus, these analyses support a conclusion that the apicomplexan and nematode genes probably share ancient exon–intron structure. This suggests that the plant introns with a similar position to intron 2 of eimepsin might have originated from common ancestors and subsequently changed position by intron drift[15] or intron slippage[16], as might *H. contortus* intron 5, which has a similar position to intron 7 of *Pf*PMVI (Fig. 3b,c).

The age of introns can provide important clues to their origin. There have been several reports of ancient intron positions in, for example, the plant and vertebrate genes encoding triose-phosphate isomerase[17], and in the ancient duplicated glyceraldehyde-3-phosphate dehydrogenase genes[18–20]. The finding of ancient introns between APs from nematodes and apicomplexans extends the distribution of shared positions of nuclear introns to taxonomically more divergent organisms, protozoa, pointing to a divergence time much longer ago than that between plants and animals[21]. Finally, the divergent exon–intron structures in APs from different groups of organisms suggested a process of gene evolution involving intron insertion[13], intron loss[20] and probably intron drift[15]. Although the proportion of these events that affect the evolution of AP families is unknown, the existence of ancient

Box 1. Evolutionary model adapted from the genomic organization of the AP family

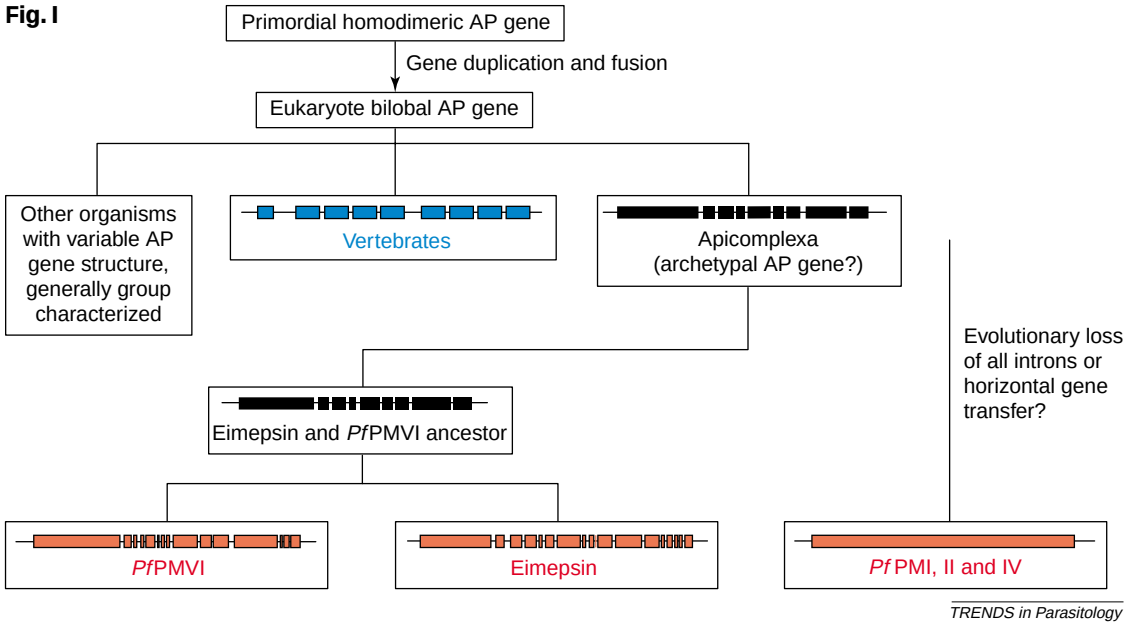
From the comparison of genomic organization, we can propose a model for the evolution of the apicomplexan aspartyl proteinases (APs) (Fig. I). It has been suggested[a] that AP genes were generated by unequal meiotic cross-over between two direct repeats of the ancestral gene. Eukaryotic APs are hypothesized to derive from this ancestral gene owing to the presence of functional conserved regions and the conservation of the two halves of the active site. Following the duplication and fusion of the ancestral AP gene, the genes in different groups of organisms have diverged.

In this model, the ancestral gene of eimepsin and *Plasmodium falciparum* plasmepsin VI (*Pf*PMVI) is assumed to be the archetypal AP gene of the apicomplexan parasites, with eight conserved introns. During the course of evolution, the genes of eimepsin and *Pf*PMVI have evolved by the insertion or removal of introns to give rise to their existing organization in 18 and 15 exons, respectively. The differences between the organization of the *Pf*PMI, II and IV genes, and the eimepsin and *Pf*PMVI genes, and the fact that they do not possess any introns, suggest that their genes probably arose from an horizontal gene transfer and were selected for the specific function of haemoglobin digestion.

Reference

a Holm, I. *et al.* (1984) Evolution of aspartyl proteases by gene duplication: the mouse renin gene is organized in two homologous clusters of four exons. *EMBO J.* 3, 557–562

Fig. I



intron positions identified in this study does not support an extreme model that proposes that the origin and evolution of all the AP introns is relatively recent.

Evolutionary relatedness of *Pf* PMVI and eimepsin genes

Eimepsin and *Pf*PMVI genes have many exons, making them the most complex of the AP family. Comparison and alignment of their exon–intron structures shows that they share an exon-grouping pattern that creates nine-exon blocks with coincident boundaries (e.g. exons 1 and 2 of eimepsin correspond precisely to exon 1 of *Pf*PMVI, and so they form block I) (Fig. 4). Compared with other APs, eimepsin has a unique C-terminal extension sequence that is encoded by exon 18 and has not be taken into account in the comparisons. The fact that this extended sequence coincides with a splice junction might be related to a particular functional difference of eimepsin, in accordance with the finding by Craik *et al.*[22] that new functional domains in protein families might be associated with a splice site.

The similar complexity of the genomic organization of eimepsin and *Pf*PMVI, the common exon–intron boundaries, and the nine-exon-block structure all suggest that eimepsin and *Pf*PMVI are derived from a recent common ancestor. The statistical tests have shown that the shared introns between these two proteins are most probably ancient and present in the common ancestor before divergence of these two parasites. Unfortunately, comparison of the nucleotide sequences of *Pf*PMVI and eimepsin introns is not informative because of the bias introduced by the AT richness of the *Plasmodium* genome[23]. However, it is possible to analyse the intron phases at the exon-block boundaries and these are all perfectly conserved

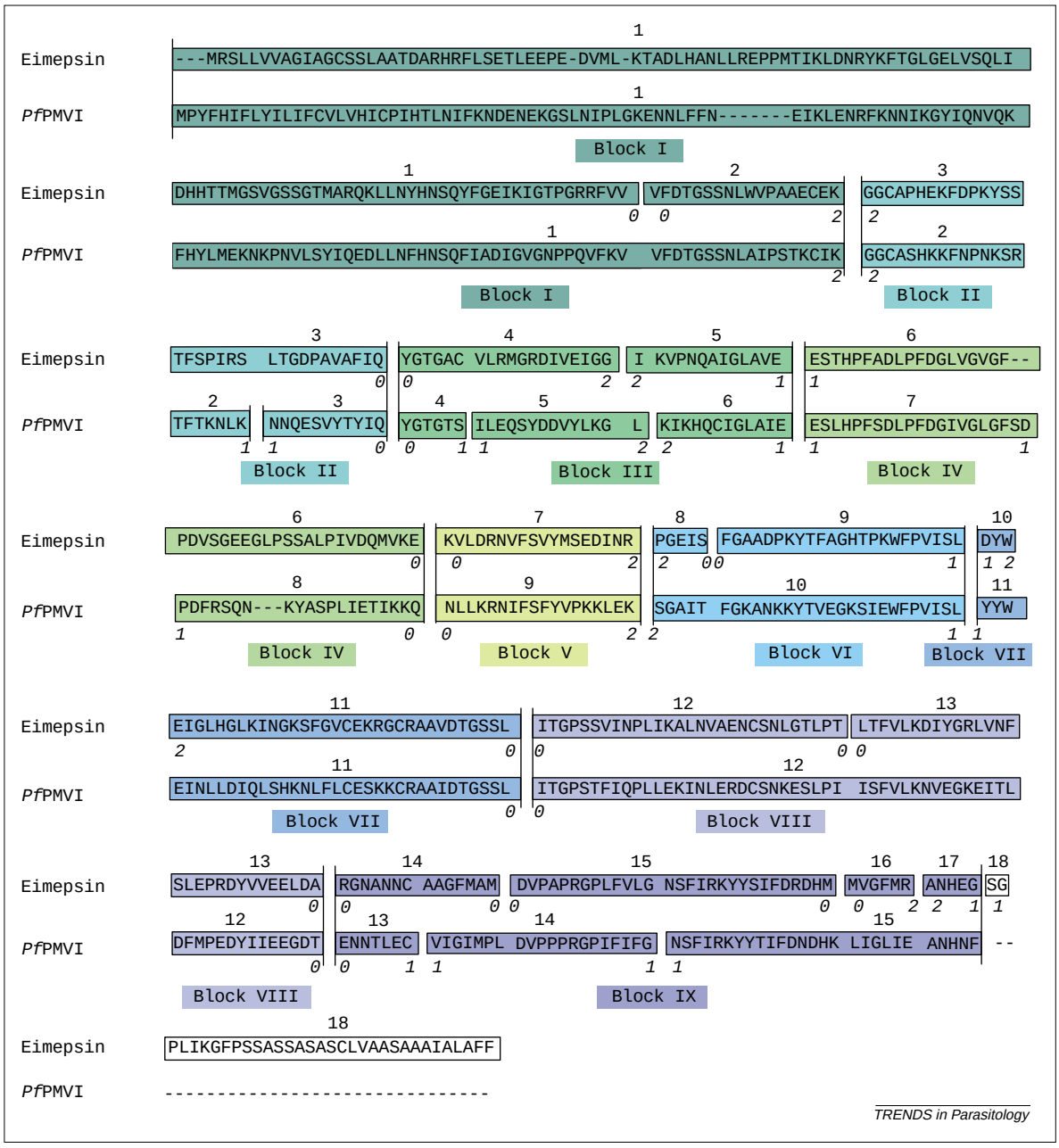


Fig. 4. Comparison of eimepsin and *Plasmodium falciparum* plasmepsin VI (*Pf*PMVI) exon–intron structures. Eimepsin and *Pf*PMVI amino acid sequences have been aligned for maximal homology (using the PILEUP program[27]) and the sequence corresponding to an exon for each protein separated and boxed. Exon numbering is shown above the sequences, exon-block numbering is shown below the sequences and the intron phase is shown in italics below the sequences. Gaps were introduced to separate the exons and to conserve the alignment when an exon in one sequence did not correspond to an exon in the other sequence.

(Fig. 4). Because there is conservation of both intron position and the phase with which the reading frame is cut[24], it is justifiable to hypothesize that the eimepsin and *Pf*PMVI exon-block-bounding introns derive from a common ancestral gene. The extra introns found at different positions in these two genes probably represent divergent evolution (intron loss or addition) after the acquisition of the nine-exon-block structure.

Comparing apicomplexan AP genes with vertebrate AP genes

The exon grouping between eimepsin and *Pf*PMVI reveals a nine-exon-block pattern, as though this grouping is designed to mimic the nine-exon structure of the vertebrate AP genes. However, a comparison of the eimepsin and *Pf*PMVI genes with vertebrate AP genes (e.g. the human cathepsin-D gene) shows that there is no conservation of boundaries between exons or exon blocks, suggesting that these genes are evolutionarily distant (Fig. 5). In addition, positioning the exon–intron boundaries on the three-dimensional structure of a typical single-chain AP (e.g. pepsin) did not show any relationship between the nine exon blocks of *Pf*PMVI and eimepsin, and their structure and function. For example, the boundaries were not positioned at external loops or between important functional domains. Thus, the nine-exon-block

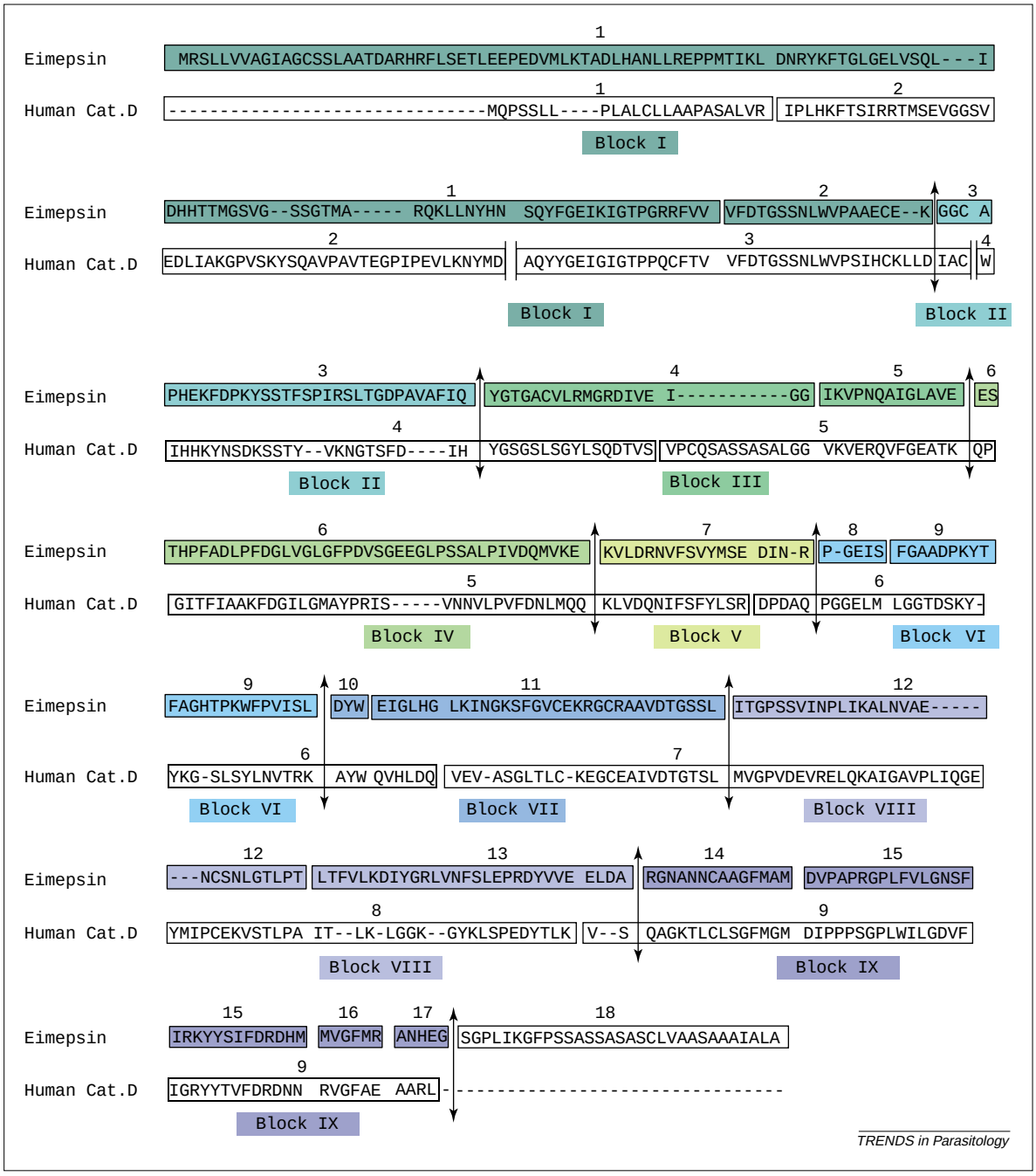


Fig. 5. Comparison of eimepsin and human cathepsin (Cat.) D exon–intron structures. Eimepsin and human cathepsin D amino acid sequences have been aligned for maximal homology (using the PILEUP program[27]) and the sequence corresponding to an exon for each protein separated and boxed. Exon numbering is shown above the sequences and exon-block numbering is shown below the sequences. Gaps were introduced to separate the exons and to conserve the alignment when an exon in one sequence did not correspond to an exon in the other sequence. The double-headed arrows represent the position of the nine exon blocks described in Fig. 4.

pattern between eimepsin and *Pf*PMVI genes does not mimic the nine-exon structure of the vertebrate AP genes and is coincidental.

Conclusions

Analysis of AP genomic sequences from diverse organisms shows that genes cluster into separate groups, owing to differences in the number of exons and the positions of exon–intron boundaries. However, some exon–intron boundaries are conserved between more than two phylogenetic lineages, suggesting that some AP introns are ancient. Interestingly, an exon-grouping pattern that creates nine exon blocks is conserved between eimepsin and *Pf*PMVI, suggesting that these two genes derive from a common ancestor. It is most likely that eimepsin and *Pf*PMVI represent archetypal APs of Apicomplexa and that other, intronless, *Pf*PM genes such as *Pf*PMI, II and IV have been acquired and retained in the *Plasmodium* genome for the specific purpose of haemoglobin digestion. The determination of further apicomplexan AP genomic sequences will provide clearer insights into this aspect of AP biology.

Acknowledgements
We would like to thank J. Kaufman for helpful discussions. L.J. was a recipient of an European Community fellowship FAIR-BM-974829. M.L. was supported by the US National Science Foundation and a Packard Fellowship. This work was carried out at the Institute for Animal Health, Compton, Newbury, UK RG20 7NN.

References

1 Tang, J. *et al.* (1978) Structural evidence for gene duplication in evolution of acid proteases. *Nature* 271, 618–621
2 Holm, I. *et al.* (1984) Evolution of aspartyl proteases by gene duplication: the mouse renin gene is organized in two homologous clusters of four exons. *EMBO J.* 3, 557–562
3 Tang, J. and Lin, X. (1994) Engineering aspartic proteases to probe structure and function relationships. *Curr. Biol.* 5, 422–427
4 Humphreys, M.J. *et al.* (1999) The aspartic proteinase from the rodent *Plasmodium berghei* as a potential model for plasmepsins from the human malaria parasite, *Plasmodium falciparum*. *FEBS Lett.* 463, 43–48
5 Bowman, S. *et al.* (1999) The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* 400, 532–538
6 Jean, L. *et al.* (2001) Genomic organisation and developmentally regulated expression of an apicomplexan aspartyl proteinase. *Gene* 262, 129–136
7 Fink, G.R. (1987) Pseudogenes in yeast? *Cell* 49, 5–6
8 Patthy, L. (1999) Genome evolution and the evolution of exon-shuffling – a review. *Gene* 238, 103–114
9 Lee, S. *et al.* (1996) dsRNA associated with virus-like particles in *Eimeria* spp. of the domestic fowl. *Parasitol. Res.* 82, 518–523
10 Dunn, P.P. *et al.* (1998) *Eimeria tenella*: two species of extrachromosomal DNA revealed by pulse-field gel electrophoresis. *Parasitol. Res.* 84, 272–275
11 Waller, R.F. *et al.* (1998) Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA.* 95, 12352–12357
12 Iwabe, N. *et al.* (1990) Compartmentalized isozyme genes and the origin of introns. *J. Mol. Evol.* 31, 205–210
13 Logsdon, J.M. Jr (1998) The recent origins of spliceosomal introns revisited. *Curr. Opin. Genet. Dev.* 8, 637–648
14 Long, M. and Rosenberg, C. (2000) Testing the 'proto-splice sites' model of intron origin: evidence from analysis of intron phase correlations. *Mol. Biol. Evol.* 17, 1789–1796
15 Rogozin, I.B. *et al.* (2000) Intron sliding in conserved gene families. *Trends Genet.* 16, 430–432
16 Martinez, P. *et al.* (1989) Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. *J. Mol. Biol.* 208, 551–565
17 Marchionni, M. and Gilbert, W. (1986) The triosephosphate isomerase gene from maize: introns antedate the plant–animal divergence. *Cell* 46, 133–141
18 Shih, M.C. *et al.* (1988) Intron existence predated the divergence of eukaryotes and prokaryotes. *Science* 242, 1164–1166
19 Kersanach, R. *et al.* (1994) Five identical intron positions in ancient duplicated genes of eubacterial origin. *Nature* 367, 387–389
20 Long, M. and de Souza, S. J. (1998) Intron–exon structures: from molecular to population biology. *Adv. Genome Biol.* 5A, 143–178
21 Tudge, C. (2000) *The Variety of Life*, Oxford University Press
22 Craik, C.S. *et al.* (1983) Splice junctions: association with variation in protein structure. *Science* 220, 1125–1129
23 Vinkenoog, R. *et al.* (1995) Comparison of introns in a *cdc2*-homologous gene within a number of *Plasmodium* species. *Mol. Biochem. Parasitol.* 71, 233–241
24 Patthy, L. (1987) Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett.* 214, 1–7
25 Thompson, J.D. *et al.* (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24, 4876–4882
26 Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12, 357–358
27 Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U. S. A.* 85, 2444–2448