Genetic variability in cysteine protease genes of *Haemonchus contortus*

A. RUIZ[1]*, J. M. MOLINA[1], A. NJUE[2] *and* R. K. PRICHARD[2]

[1] *Parasitology Unit, Department of Animal Pathology, Veterinary Faculty, University of Las Palmas de Gran Canaria, 35416 Arucas, Las Palmas, Spain*
[2] *Institute of Parasitology, McGill University, Ste Anne-de-Bellevue, Quebec H9X 3V9, Canada*

(*Received 30 July 2003; revised 8 November 2003; accepted 8 November 2003*)

SUMMARY

To increase the existent genetic variability in cysteine proteases, a polymorphism study was performed in *Haemonchus contortus* by comparing 2 different strains of the parasite: North American (NA) and Spanish (SP) strains. For this purpose, the polymorphism of 5 previously reported genes (*AC-1*, *AC-3*, *AC-4*, *AC-5* and *GCP-7*) were analysed by PCR–SSCP and sequencing procedures. Based on the SSCP results, a total of 20 different alleles were identified for the 5 *loci* assessed. Except *locus AC-5*, all the *loci* were polymorphic. *Loci AC-1*, *AC-3*, *AC-4* and *GCP-7* showed 5, 8, 2 and 4 alleles, respectively. The allelic frequencies ranged from 0·0070 to 0·8560 and were significantly different between strains. In addition, nucleotide diversity analyses showed a significant variation within and between strains. The variations in the nucleotide sequence of the different alleles were translated in some cases into changes in the amino acid sequence. Evidence of genetic variability in cysteine proteases from two different strains of *H. contortus* for the same set of genes had not been previously reported.

Key words: *Haemonchus contortus*, cysteine proteases, genetic variation, SSCP.

INTRODUCTION

Cysteine proteases are one of the main catalytic groups of peptide hydrolases, together with serine, threonine, aspartate and metallo-proteases (McKerrow, 1989; Coombs & Mottram, 1997; Tort *et al*. 1999). Also referred to as thiol or sulfhydryl proteases (Barrett, 1994), cysteine proteases have been identified in plants (Glazer & Smith, 1971; Kumar Dubey & Jagannadham, 2003), animals (Barrett & McDonald, 1980; Bania *et al*. 2003), viruses (Bazan & Fletterick, 1988; Ziebuhr *et al*. 2003), bacteria (Morihara, 1974; Svensson *et al*. 2000) and eukaryotic microorganisms (North, 1982; Nesterenko *et al*. 1995). Cysteine proteases of parasitic organisms are divided into two main groups, referred to as clans CA and CD according to sequence similarity, possession of inserted peptide loops and biochemical specificity to small peptide substrates (Rawlings & Barrett, 1993; Barrett, 1994). The majority of parasite cysteine proteases belong to the family C1 within clan CA, and are further divided into cathepsin B and cathepsin L-like subfamilies.

Most of the human cathepsins have an acidic pH optimum which allows full activity within the lysosomal compartment (Barrett & Kirschke, 1981). In contrast, many parasitic cysteine proteases are more active at neutral or slightly alkaline pH (Eakin *et al*. 1992; Caffrey *et al*. 2001; Sajid & McKerrow, 2002). Neutral or alkaline pH optima are in accordance with the extracellular activity observed for these proteases. Roles of parasitic cysteine proteases in nutrition, tissue and cell invasion, ex/encystement, hatching and immunoevasion have been recently discussed in detail (Sajid & McKerrow, 2002).

Because of the ubiquity of cysteine proteases in both protozoan and helminth parasites, they represent attractive targets for anti-parasitic drug development. Most of this work has focused to date on the papain family of proteases (cathepsin L and B-like proteases) (Li *et al*. 1994; Du *et al*. 2002; Rosenthal *et al*. 2002). In addition, it has been extensively demonstrated that many cysteine proteases are immunogenic and this has been exploited in their use as convenient immunological diagnostic markers for infectious diseases, including infections of *Ancylostoma caninum* (see Loukas *et al*. 2000), *Fasciola hepatica* (see Neyra, Chavarry & Espinoza, 2002), *Fasciola gigantica* (see Dixit, Yadav & Sharma, 2002) and *Clonorchis sinensis* (see Na *et al*. 2002). Antibodies directed against cysteine proteases can have an inhibitory effect on their proteolytic activity. A number of encouraging studies to verify the application of an anti-cysteine protease vaccine against parasitic organisms has been carried out in *Trypanosoma congolense* (see Authie *et al*. 2001), *F. hepatica* (see Dalton *et al*. 1996), *Ostertagia ostertagi* (see Geldhof *et al*. 2002) and *H. contortus* (see Skuce *et al*.

* Corresponding author: Área de Parasitología, Departamento de Patología Animal, Facultad de Veterinaria, Trasmontaña s/n, 35416 Arucas, Las Palmas, Spain. Tel: +34 928 451113. Fax: +34 928 454341. E-mail: aruiz@dpat.ulpgc.es

Parasitology (2004), **128**, 549–559.
DOI: 10.1017/S0031182004004998 Printed in the United Kingdom

1999). These observations may explain the increasing attention addressed to cysteine proteases of parasites.

One of the best characterized families of cathepsin B-like proteases have been described in the abomasal nematode *H. contortus* (see Pratt *et al*. 1990, 1992; Rehman & Jasmer, 1998; Skuce *et al*. 1999; Jasmer, Roth & Myler, 2001). Inter- and intrageographical variation of cysteine proteases has been demonstrated among different strains of *H. contortus* in the protease profile using gelatin-containing SDS–PAGE gels (Karanu *et al*. 1993, 1997) and at the genetic level (Rehman & Jasmer, 1998; Skuce *et al*. 1999). However, to date, the magnitude of this variability has not been assessed within and among different strains of *H. contortus* from different species of host for a same set of genes under similar laboratory conditions. Using PCR coupled Single-Strand Conformation Polymorphism (SSCP) and sequencing methodologies, we have estimated the genetic polymorphism in five previously reported cysteine protease genes of *H. contortus* from sheep (North America) and goats (Spain).

MATERIALS AND METHODS

Parasites

Individual worms from 2 strains of *H. contortus* were used in this study. One strain (NA) corresponded to *H. contortus* from North America and was maintained in experimentally infected sheep at the Institute of Parasitology, McGill University. The second strain (SP) of *H. contortus* was originally isolated from naturally infected goats from the Canary Islands (Spain) and maintained experimentally in goats at the Faculty of Veterinary of the University of Las Palmas de Gran Canaria.

DNA isolation

H. contortus adult males from both the NA and the SP strains were taken from the abomasum of the corresponding single host (sheep or goats) experimentally infected with 3rd-stage larvae (L3) of the parasite. Only adult males were used to avoid the possibility of DNA contamination from the eggs or sperm present in females resulting in more than a single genotype in each sample. The worms were washed in RPMI medium (Sigma–Aldrich) at 37 °C and then frozen at −80 °C until DNA isolation was performed. The DNA was isolated from 55 individual males of the SP strain and 97 males of the NA strain. Each worm was transferred to a tube containing 200 μl of STE (0·1 M NaCl, 10 mM Tris–HCl, pH 8·0, 1 mM EDTA, pH 8·0), 0·6 M β-mercaptoethanol, 0·5% SDS and 200 μg/ml proteinase K (Sambrook, Fritsch & Maniatis, 1989) and incubated overnight at 55 °C. Two DNA extractions, one with phenol and another with phenol/chloroform, were performed. The DNA was then precipitated with 2·5 M ammonium acetate and 50% ethanol, following the addition of 10 μg of linear acrylamide as co-precipitant (Gaillard & Strauss, 1990). The DNA pellet was air-dried and redissolved in 50 μl of 1×TE (10 mM Tris–HCl, 1 mM EDTA, pH 8·0).

Polymerase chain reaction (PCR) amplification

PCR reactions were performed in a PTC-200 Peltier Thermal Cycler (MJ Research Inc.) using reagents and *Taq* polymerase provided by Gibco BRL. Forward and reverse primers (Table 1) were based on reported sequences from *H. contortus* encoding cysteine protease genes *AC-1* (Pratt *et al*. 1990), *AC-3*, *AC-4* and *AC-5* (Pratt *et al*. 1992) and *GCP-7* (Rehman & Jasmer, 1998), and the annealing temperatures for the PCR amplifications were: 59 °C for the set of primers *AC-5*, 54 °C for the primers *AC-1* and *AC-3*, and 53 °C for the primers *AC-4* and *GCP-7*. The other conditions for the amplifications were the same for all pairs of primers: 94 °C for 2 min to denature the template DNA, followed by 35 cycles of 15 sec at 95 °C, 30 sec at the corresponding annealing temperature and 1 min at 70 °C. A final step of 2 min at 15 °C was included. For the PCR reactions ~2 ng (1–2 μl) of DNA from individual male worms were used. In addition, the reaction mixture for all pairs of primers contained 0·5 U *Taq* polymerase, 0·8 μM corresponding forward and reverse primers, 0·2 mM dNTPs, 1 mM $MgCl_2$, 10% (v/v) 10× buffer reaction and H_2O to a total volume of 25 μl. Finally, the PCR products were subjected to electrophoresis in 1·2% agarose gels and subsequent staining with ethidium bromide.

Single Strand Conformation Polymorphism (SSCP)

One μl of PCR product was mixed with 15 μl of SSCP loading buffer (95% formamide, 10 mM NaOH, 0·25% xylene cyanol, 0·25% bromophenol blue). The mixture was denatured at 95 °C for 2 min and cooled immediately on iced water before being loaded onto a non-denaturing polyacrylamide gel. Electrophoresis was performed in a Hoefer SE600 (Pharmacia Biotech, San Francisco, CA). Electrophoresis conditions were optimized for each gene (*AC-1*, *AC-3*, *AC-4*, *AC-5* and *GCP-7*) and strain (SP and NA), in order to produce unique migration and separation patterns for the single strands of the different alleles. The parameters optimized in the electrophoresis were the running time (ranging from 12 to 23 h), the power (from 75 to 110 volts) and the percentage of acrylamide/*bis*-acrylamide (from 10 to 15%). The gels were made using different percentages of a 49:1 proportion of acrylamide:*bis*-acrylamide, buffer TBE (1·11 M Tris, 1·11 M boric acid, 0·003 M EDTA, pH 8·0), 0·09% (v/v)

Table 1. Gene-specific primers and allele structure of cysteine protease genes of *Haemonchus contortus* from Spain and North America

					Allele structure†				
Gene	Primer*	Primer sequence	Allele	Accession number	5′ Exon 1	Intron 1	Exon 2	Intron 2	3′ Exon 3
AC-1	F	5′TTTCTGCCACTGACATCA3′	A_1	AF550374	1–51	51–128	128–211	211–268	268–361
	R	5′ACGGTGGGGTTGGCGCTG3′	B_1	AF550375	1–51	51–128	128–211	211–270	270–361
			C_1	AF550376	1–51	51–128	128–211	211–270	270–360
			D_1	AF550377	1–51	51–128	128–211	211–270	270–361
			E_1	AF550378	1–51	51–128	128–211	211–268	268–361
AC-3	F	5′GACATCCTGTACGCCAAC3′	A_3	AF550379	1–30	30–111	111–193	193–259	259–344
	R	5′GTTGACGCCTCTTCAGGA3′	B_3	AF550380	1–27	27–110	110–192	192–255	255–340
			C_3	AF550381	1–30	30–111	111–193	193–259	259–344
			D_3	AF550382	1–30	30–111	111–193	193–258	258–343
			E_3	AF550383	1–30	30–111	111–194	194–256	256–341
			F_3	AF550384	1–30	30–111	111–193	193–259	259–344
			G_3	AF550385	1–27	27–111	111–193	193–259	259–344
			H_3	AF550386	1–30	30–111	111–193	193–259	259–344
AC-4	F	5′ATTTTGACATGCTGCAAT3′	A_4	AF550387	1–37	37–95	95–178	178–241	241–330
	R	5′TGGAGTTGCCGCCTCTCG3′	B_4	AF550388	1–36	36–94	94–170	170–240	240–329
AC-5	F	5′TGTGGAGCACGATGTGGG3′	A_5	AF550389	1–25	25–112	112–195	195–260	260–348
	R	5′AGTGGGCGCCATTCCAAC3′							
GCP-7	F	5′GCATGCTGTGGAAAGTTC3′	A_7	AF550390	1–29	29–361	361–444	444–513	513–600
	R	5′CGGAGTGGCATAGGGATG3′	B_7	AF550391	1–29	29–350	350–433	433–531	531–622
			C_7	AF550392	1–26	26–370	370–453	453–552	552–643
			D_7	AF550393	1–26	26–350	350–433	433–517	517–608

* F, forward primer; R, reverse primer. The primers were based on reported sequences from cysteine protease genes *AC-1* (Pratt *et al*. 1990), *AC-3*, *AC-4* and *AC-5* (Pratt *et al*. 1992) and *GCP-7* (Rehman & Jasmer, 1998). The accession numbers of these 5 genes are M31112, M80388, M80386, M80385 and AF046229, respectively.

† Inferred structure of the different alleles indicating the length and the exact point of coincidence with the cDNA reported sequences. Matching was performed by standard nucleotide–nucleotide BLAST (NBC GeneBank). Two introns and 3 exons of variable length were present in all the alleles.

N,N,N,N′-tetra-methyl-ethylenediamine and 0·07% (w/v) ammonium persulphate (the solutions are expressed as final concentrations). Gels were run in 1 × TBE buffer (0·89 M Tris, 0·89 M boric acid, 0·002 M EDTA, pH 8·0) at room temperature (22–24 °C). The gels were stained with ethidium bromide, scanned using a Bio-Rad Molecular Imager® FX, and the patterns recorded with the corresponding Quantity One Software (Version 4.2.1) for subsequent analysis.

Sequence analysis

DNA fragments that displayed different electrophoretic patterns in the SSCP analyses were selected for sequencing. PCR products were purified using the Nucleospin extraction kit (Clontech) and ligated into the plasmid vector pCR®2.1 (Gibco BRL). Transformation into One Shot® competent *Escherichia coli* cells was then carried out according to the manufacturer's instructions (TA Cloning® Kit, Gibco BRL). The plasmid DNA was isolated using a Qiaprep® Spin Miniprep Kit (Qiagen) and sequencing was carried out by the ABI Big Dye cycle sequencing kit and an ABI Prism 377 automated sequencer. Before sequencing, the Miniprep products were subjected to PCR–SSCP analysis to confirm the electrophoretic patterns of inserts from recombinant clones. This analysis proved that, for every individual male used in the sequencing, the SSCP profile obtained in the first screening of the whole population (North American or Spanish populations) was identical to that observed after cloning. According to the reproducibility of the SSCP results in all the assays it is quite unlikely that the observed diversity causing the SSCP patterns was a consequence of base misincorporation during PCR.

Data analysis

Genotype frequencies for each gene were tested for Hardy–Weinberg equilibrium for an excess of homozygotes, calculating from a binomial distribution based on the observed allele frequencies (Sokal & Rohlf, 1981). Differences in allele frequencies between strains were tested for significance using a G test for heterogeneity, pooling allele classes if necessary to ensure a minimum expected number of at least 5 individuals (Sokal & Rohlf, 1981). Significance was taken at the 5% level. To further analyse the allelic frequencies, a set of intra- and inter-population genetic statistics were estimated, and corrected for small sample size (Nei, 1978) and small number of populations (Nei, 1986) using the Gene-Stat-pc 3.3 computer program (Lewis, 1994). Some of the statistics included the percentage of polymorphic *loci* (P) (95% criterion), the mean number of alleles per *locus* (A), expected heterozygosity (H_e), total genetic diversity (H_t), genetic diversity within populations (H_s), genetic diversity among populations (D_{st}) and the relative magnitude of genetic differentiation among populations ($G_{st} = D_{st}/H_t$) (Nei, 1978).

In order to analyse the nucleotide variability and the phylogenic relations among the different genes in both the SP and the NA strains of *H. contortus*, the sequences obtained were initially aligned using CLUSTAL W (1.81) (Higgins *et al.* 1994) and then treated with the computer program MEGA version 2.1 (Kumar *et al.* 2001). Jukes-Cantor's genetic distances (Jukes & Cantor, 1969) for each *locus* were calculated for all pairwise combinations. The mean distance within a subpopulation, the mean inter-populational distance and the mean diversity for the entire population were also determined for all *loci* using the same distance method. A dendrogram was constructed based on the matrix of the distances using the Neighbour-Joining Tree method (NJ) (Saitou & Nei, 1987). In all cases the standard error was estimated by a bootstrap procedure with a total of 500 replications. Further estimations of the variability within the two strains of *H. contortus*, including the number of polymorphic sites (S), the nucleotide diversity (π) and the heterozygosity per nucleotide site (θ) were performed using the computer program DnaSP version 3.51 (Rozas & Rozas, 1999).

RESULTS

SSCP analysis

The amplification of the *loci AC-1*, *AC-3*, *AC-4* and *AC-5* resulted in PCR products of ~350 bp. A PCR

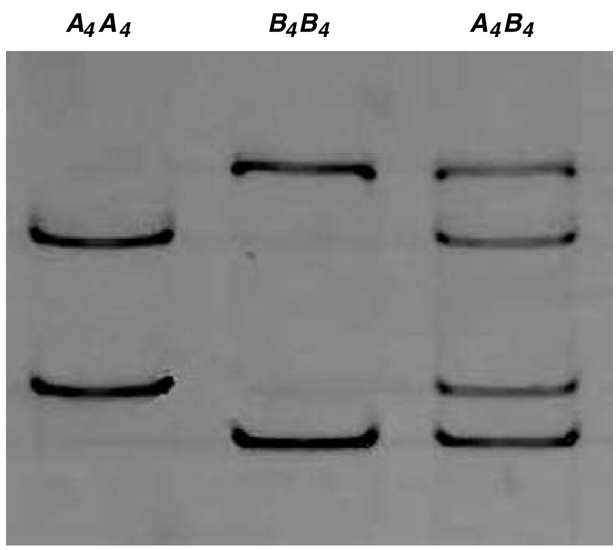

Fig. 1. SSCP profiles for the gene encoding cysteine proteases *AC-4* in two strains of *Haemonchus contortus* (Spanish and North American strains). A_4 and B_4 represent the alleles detected in the two strains of the parasite. The homozygotes A_4A_4 and B_4B_4 and the corresponding heterozygote A_4B_4 were present in both strains.

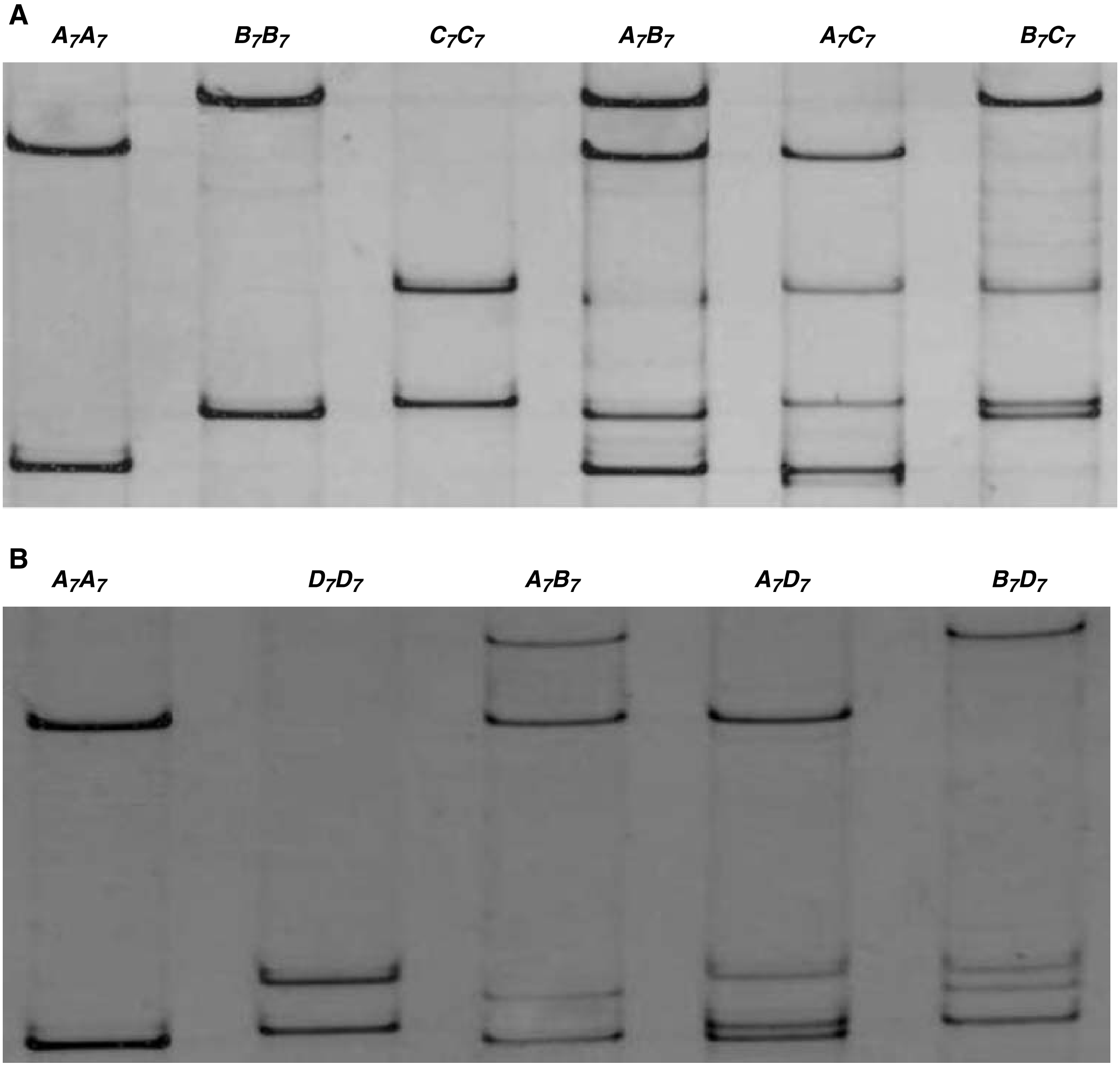

Fig. 2. SSCP profiles for the gene encoding cysteine proteases *GCP-7* in two strains of *Haemonchus contortus*. The profiles for the Spanish and the North American strains are represented separately in (A) and (B), respectively. A_7, B_7, C_7 and D_7 represent the different alleles identified in the two strains of the parasite. In the Spanish strain, 3 homozygotes (A_7A_7, B_7B_7 and C_7C_7) and the corresponding heterozygotes (A_7B_7, A_7C_7 and B_7C_7) were identified. In the NA strain the homozygotes A_7A_7 and D_7D_7 and the heterozygotes A_7B_7, A_7D_7 and B_7D_7 were detected (B).

product of ~625 bp was identified when the *locus* *GCP-7* was amplified. While no variation in size was detectable among the PCR products from individual males on agarose gels, SSCP results revealed distinct profiles among some of the samples for all the genes analysed. Examples of SSCP gels are shown in Figs 1 and 2. The frequencies of the different SSCP patterns determined for each *locus* are displayed in Table 2. Based on the SSCP results, a total of 20 different alleles were identified for the 5 *loci* assessed. Subsequent sequencing of PCR products samples for each allele confirmed that the differences in banding patterns (homozygosity and heterozygosity) detected in the SSCP results were due to nucleotide variation. The allelic frequencies for each *locus* are shown in Table 3. These frequencies did not differ significantly from the Hardy–Weinberg equilibrium for any of the *loci* assessed for both the SP and the NA strain.

Five different alleles were detected within the Spanish (SP) and the North American strains (NA) for the *locus AC-1* (A_1, B_1, C_1, D_1 and E_1). The homozygotes A_1A_1 and B_1B_1, the corresponding heterozygote (A_1B_1) and 3 heterozygotes for allele A_1 (A_1C_1, A_1D_1 and A_1E_1) were detected for all males analysed. The allelic frequencies ranged from 0·0070 (NA strain E_1) to 0·8560 (SP strain A_1) and were statistically different between the two populations (Table 3). Except for the allele E_1, which was only detected in the NA strain, the other alleles were present in both strains. The most frequent SSCP patterns were A_1A_1 and A_1B_1 for the SP and NA strains, respectively (Table 2).

Table 2. Frequencies of SSCP profiles in 5 genes encoding cysteine proteases of *Haemonchus contortus* from Spain (SP) and North America (NA)

(Differences in the SSCP profiles between the strains were tested for significance using a G test, pooling SSCP migration patterns when necessary to ensure a minimum of at least five individuals. Significant differences between populations were observed at the 0·1% level ($P<0{\cdot}001$*) in genes *AC-1*, *AC-3* and *GCP-7*, while any statistical difference was detected for genes *AC-4* and *AC-5*. The most abundant SSCP profiles were A_1A_1, A_3B_3, B_4B_4, A_5A_5 and A_7A_7–A_7C_7 (Spanish strain) and A_1B_1, B_3D_3, B_4B_4, A_5A_5 and A_7A_7 (North American strain), for genes *AC-1*, *AC-3*, *AC-4*, *AC-5* and *GCP-7*, respectively.)

Loci	SSCP profiles	SP	NA
*AC-1**		N=52	N=75
	A_1A_1	0·7310	0·3070
	B_1B_1	0·0190	0·2000
	A_1B_1	0·1540	0·4400
	A_1C_1	0·0770	0·0270
	A_1D_1	0·0190	0·0130
	A_1E_1	0·0000	0·0130
*AC-3**		N=45	N=71
	A_3A_3	0·0670	0·0000
	B_3B_3	0·2220	0·0420
	C_3C_3	0·0220	0·0000
	D_3D_3	0·0440	0·1970
	E_3E_3	0·0000	0·0420
	A_3B_3	0·3560	0·0560
	A_3C_3	0·0670	0·0000
	A_3D_3	0·0000	0·0280
	A_3F_3	0·0670	0·0000
	A_3G_3	0·0440	0·0000
	B_3C_3	0·0670	0·0000
	B_3D_3	0·0220	0·2680
	B_3E_3	0·0000	0·1130
	B_3H_3	0·0220	0·0000
	D_3E_3	0·0000	0·2540
AC-4		N=37	N=87
	A_4A_4	0·0000	0·0580
	B_4B_4	0·5950	0·5630
	A_4B_4	0·4050	0·3790
AC-5		N=25	N=29
	A_5A_5	1·0000	1·0000
*GCP-7**		N=44	N=93
	A_7A_7	0·3180	0·4620
	B_7B_7	0·0910	0·0000
	C_7C_7	0·0910	0·0000
	D_7D_7	0·0000	0·0540
	A_7B_7	0·1360	0·0650
	A_7C_7	0·3180	0·0000
	A_7D_7	0·0000	0·3870
	B_7C_7	0·0450	0·0000
	B_7D_7	0·0000	0·0320

Table 3. Allelic frequencies estimated by SSCP in genes encoding cysteine proteases of *H. contortus* from Spain (SP) and North America (NA)

(Differences in the allele frequencies between the strains were tested for significance using a G test, pooling allele classes when necessary to ensure a minimum of at least five individuals. Significant differences between populations were observed at the 0·1% level ($P<0{\cdot}001$*) in genes *AC-1*, *AC-3* and *GCP-7*, while any statistical difference was detected for genes *AC-4* and *AC-5*. The most abundant alleles were A_1, B_3, B_4, A_5 and A_7 (Spanish strain) and A_1, D_3, B_4, A_5 and A_7 (North American strain) for genes *AC-1*, *AC-3*, *AC-4*, *AC-5* and *GCP-7*, respectively.)

Loci	Allele	SP	NA
*AC-1**		N=52	N=75
	A_1	0·8560	0·5530
	B_1	0·0960	0·4200
	C_1	0·0380	0·0130
	D_1	0·0100	0·0070
	E_1	0·0000	0·0070
*AC-3**		N=45	N=71
	A_3	0·3330	0·0420
	B_3	0·4560	0·2610
	C_3	0·0890	0·0000
	D_3	0·0560	0·4720
	E_3	0·0000	0·2250
	F_3	0·0330	0·0000
	G_3	0·0220	0·0000
	H_3	0·0110	0·0000
AC-4		N=37	N=87
	A_4	0·2030	0·2510
	B_4	0·7970	0·7490
AC-5		N=25	N=29
	A_5	1·0000	1·0000
*GCP-7**		N=44	N=93
	A_7	0·5450	0·6880
	B_7	0·1820	0·0480
	C_7	0·2730	0·0000
	D_7	0·0000	0·2630

A total of 8 different alleles were detected within the two populations for the *locus AC-3* (A_3, B_3, C_3, D_3, E_3, F_3, G_3 and H_3). In the SP strain, the homozygotes A_3A_3, B_3B_3, C_3C_3 and D_3D_3, the corresponding heterozygotes (A_3B_3, A_3C_3, B_3C_3 and B_3D_3) and other heterozygotes for the allele A_3 (A_3F_3 and A_3G_3) and for the allele B_3 (B_3H_3) were identified. In the NA strain, the homozygotes D_3D_3, B_3B_3 and E_3E_3, the corresponding heterozygotes (D_3B_3, D_3E_3 and B_3E_3) and other heterozygotes for the allele D_3 (A_3D_3) or B_3 (A_3B_3) were also found. The allelic frequencies ranged from 0·0110 (SP strain H_3) to 0·4720 (NA strain D_3) (Table 3) and were statistically different between the two strains. Alleles C_3, F_3, G_3 and H_3 were exclusive to the SP strain. The most frequent SSCP patterns were A_3B_3 and B_3D_3 for the SP and NA strains, respectively (Table 2).

The SSCP results for the *locus AC-4* are shown in Fig. 1. Only 2 alleles, A_4 and B_4, were detected. The homozygotes A_4A_4 and B_4B_4 and the corresponding heterozygote A_4B_4 were present in both strains. The allelic frequencies, which were higher in allele B_4, were not statistically different between the two strains (Table 3). The most frequent SSCP pattern in both strains was the heterozygote A_4B_4 (Table 2).

Only one banding pattern was detected in the SSCP analysis of the *locus AC-5* for both strains, the unique homozygote confirmed being identified as A_5.

Table 4. Estimates of the nucleotide variability within and between strains (SP and NA strains) in 5 genes encoding *Haemonchus contortus* cysteine proteases

(Number of polymorphic (segregating) sites (S), nucleotide diversity (π) and heterozygosity per nucleotide site (θ). Standard errors are given in parentheses. SP: Spanish strain. NA: North American strain.)

	AC-1		*AC-3*		*AC-4*		*AC-5*		*GCP-7*	
	SP	NA	SP	NA	SP	NA	SP	NA	SP	NA
S	28	30	50	35	11	11	0	0	74	88
π	0·0129 (0·0027)	0·0232 (0·0008)	0·0431 (0·0015)	0·0328 (0·0014)	0·0110 (0·0019)	0·0127 (0·0011)	0·0000 (0·0000)	0·0000 (0·0000)	0·0555 (0·0026)	0·0417 (0·0032)
θ	0·0149 (0·0045)	0·0149 (0·0043)	0·0291 (0·0082)	0·0187 (0·0052)	0·0069 (0·0027)	0·0058 (0·0021)	0·0000 (0·0000)	0·0000 (0·0000)	0·0252 (0·0068)	0·0265 (0·0064)
Mean distance within groups	0·0134 (0·0028)	0·0239 (0·0056)	0·0451 (0·0074)	0·0342 (0·0066)	0·0112 (0·0033)	0·0130 (0·0038)	0·0000 (0·0000)	0·0000 (0·0000)	0·0595 (0·0072)	0·0446 (0·0050)
Mean distance between groups	0·0228 (0·0052)		0·0494 (0·0081)		0·0121 (0·0036)		0·0000 (0·0000)		0·0638 (0·0068)	
Mean diversity for the entire population	0·0216 (0·0050)		0·0431 (0·0069)		0·0125 (0·0036)		0·0000 (0·0000)		0·0545 (0·0056)	

Four different alleles (A_7, B_7, C_7 and D_7) were detected in the SSCP analysis of *locus GCP-7* (Fig. 2). Alleles C_7 and D_7 were exclusive to the SP and NA strains, respectively. Three homozygotes (A_7A_7, B_7B_7 and C_7C_7) and the corresponding heterozygotes (A_7B_7, A_7C_7 and B_7C_7) were identified in the SP strain (Fig. 2A). In the NA strain, the homozygotes A_7A_7 and D_7D_7 and the heterozygotes A_7B_7, A_7D_7 and B_7D_7 were found (Fig. 2B). The allelic frequencies ranged from 0·0480 (NA strain B_7) to 0·6880 (NA strain A_7) and were statistically different between the two strains (Table 3). The most frequent SSCP patterns were A_7A_7 and A_7C_7 for the SP strain, and A_7A_7 for the NA strain (Table 2).

Averaged across populations, the mean number of alleles was A = 3·2, percentage of polymorphic *loci* P = 80, and expected heterozygosity was $H_e = 0{\cdot}39$. No significant differences were found among populations for any of the polymorphic indices. Partitioning of the populations' genetic diversity showed that genetic diversity within populations, $H_s = 0{\cdot}459$, accounted for 85% of the total genetic diversity. Genetic diversity among populations, $D_s = 0{\cdot}071$, accounted for 15% of the total genetic diversity. This result was reflected in a G_{st} of 0·154 (Nei, 1978), which measures the proportion of the genetic diversity attributable to population differentiation. Based on Nei's (1978) genetic estimates a high mean identity of 0·885 was detected between the two *H. contortus* strains.

Nucleotide diversity and genetic relationships

Pairwise genetic distances (and S.E.) between alleles ranged from 0·0030 (0·0028) between alleles B_3 and A_3 to 0·1300 (0·0154) between alleles B_7 and D_7 (original data are available from authors). Other estimates of the nucleotide variability within and between strains are depicted in Table 4. Except for the *locus AC-4*, the number of polymorphic sites (S), the nucleotide diversity (π), the heterozygosity per nucleotide site (θ) and the mean distances within groups were different for each strain. The number of polymorphic sites fluctuated from 11 to 88, π from 0·0110 (0·0019) to 0·0555 (0·0026), θ from 0·0069 (0·0027) to 0·0291 (0·0082), and mean distances varied within groups from 0·0112 (0·0033) to 0·0595 (0·0072). Usually, the lowest and the highest values for all of these parameters corresponded to *loci AC-4* and *GCP-7*, respectively. Similarly, the values of the mean distances between groups and the mean diversity for the entire population were highest in the *GCP-7 locus*, followed by the *AC-3* and the *AC-1 loci*, and the lowest values corresponded to *locus AC-4*.

Nucleotide sequences for the 20 alleles have been submitted to GenBank and their accession numbers are given in Table 1. Standard nucleotide–nucleotide BLAST analysis (NBC GenBank) was performed for all the alleles of each *locus*. In all cases, a high degree of identity was found between the genomic DNA sequences of the different alleles and the cDNA published sequences: *AC-1* (M31112), *AC-3* (M80388), *AC-4* (M80386), *AC-5* (M80385) and GCP-7 (AF046229). According to the alignments, 3 exons and 2 introns were present in each of the 5 *loci* analysed in this study. A general representation of the structure of the different alleles, indicating the exact point of coincidence with the cDNA reported sequences, is detailed in Table 1. Using BLASTX, the amino acid sequences of the different alleles were also inferred. Nucleotide variability was translated into amino acid changes in some alleles with differences ranging from 1·35 (E_1 *vs.* B_1, C_1 and D_1) to 14·93% (B_7 and C_7 *vs.* A_7 and D_7).

Genetic relationships among the different *loci* were examined further by a bootstrap test of phylogeny based on the Neighbour-Joining Tree construction

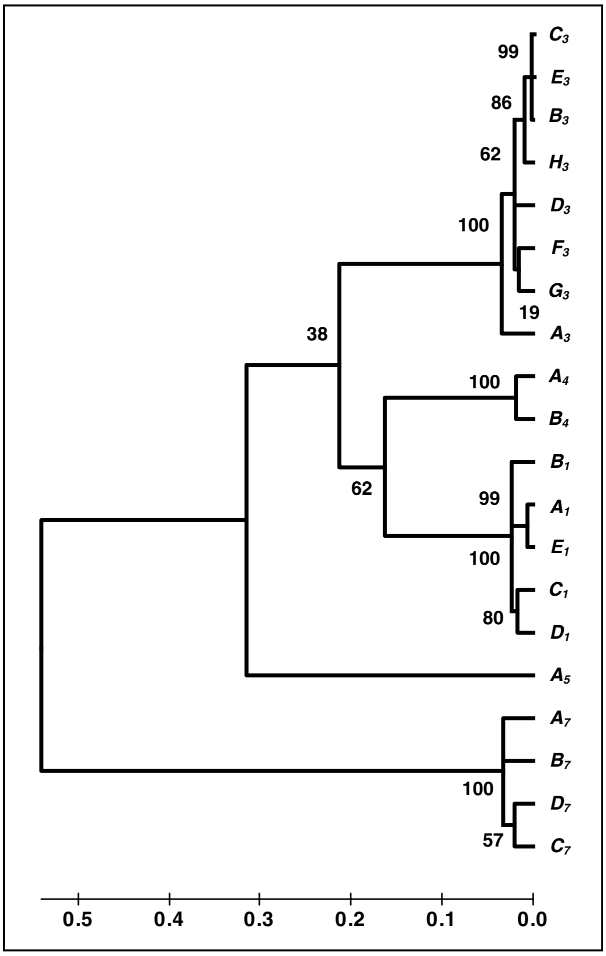

Fig. 3. Dendrogram of cysteine protease sequences of *Haemonchus contortus* from Spain and North America. The tree was obtained by 500 replicates using Jukes Cantor and Neighbour-Joining methods in MEGA 2.1 version. The numbers besides the branches are percentage support from the 500 bootstrap resamples. The scale represents the Jukes Cantor distances among the alleles of the different *loci* analysed.

method and are represented by a dendrogram (Fig. 3). The dendrogram separates the 5 *loci* into 2 primary clusters, one made up of the alleles from *locus GCP-7* and the other formed by the alleles from *loci AC 1-5*. In the second cluster, the *locus AC-5* is separated from the other 3 *loci*, two of them (*loci AC-1* and *AC-4*) being closely associated and more distant from *locus AC-3*. These results are in accordance with previous data differentiating 2 distinct clades for cathepsin B-like cysteine proteases (Pratt *et al*. 1992; Rehman & Jasmer, 1998), Clade I containing the CBL *AC-1* to *AC-5* and Clade II containing *GCP-7*.

DISCUSSION

The present study demonstrates genetic variation in cysteine protease genes between Spanish and North American strains of *H. contortus* by SSCP analysis and subsequent sequencing. Although genetic variation has already been reported in cathepsin B-like cysteine proteases (CBL) of *H. contortus* (see Cox *et al*. 1990; Pratt *et al*. 1990, 1992; Rehman & Jasmer, 1998; Skuce *et al*. 1999; Jasmer *et al*. 2001), the magnitude of this genetic diversity had not been measured previously within and among different strains of *H. contortus* for a same set of genes. Allelic variability for other *H. contortus* genes has also been investigated, which make this parasite an extremely diverse nematode at the genetic level. Nucleotide diversity has been demonstrated in the genes encoding 2 β-tubulins of the parasite (Kwa *et al*. 1993; Beech, Prichard & Scott, 1994), several P-glycoproteins (Pgps) (Blackhall *et al*. 1998*a*; Sangster *et al*. 1999), 2 glutamate-gated chloride channel subunits (GluCla and GluCl β subunits), an *N*-acetylcholine receptor and a phosphoenolpyruvate carboxykinase (Blackhall *et al*. 1998*b*). Genetic diversity has been proved as well in tandem-repeat-type galectins (Greenhalgh, Beckham & Newton, 1999), the transposable element Tc1 and transposon integration (Hoekstra *et al*. 1999, 2000) and in microsatellite analyses (Hoekstra *et al*. 1997; Otsen *et al*. 2000).

SSCP analysis and direct sequencing have been used in the characterization of DNA polymorphism in the 432-bp core region of the cruzipain gene, which encodes the active site of cathepsin L-like cysteine protease (De Leon *et al*. 1998). However, the usefulness of SSCP has not been previously demonstrated for detecting polymorphism in cathepsin B-like cysteine protease genes. SSCP banding patterns obtained for each of the 5 *loci* assessed in this study were all readily characterized once optimal conditions for the method were determined. Alleles which differed by 2 bases (A_1 *vs*. E_1) could be distinguished. The variation in SSCP patterns among the 20 alleles reflected the sequence variability estimated by Jukes-Cantor pairwise distances. This evidence, together with the reproducibility of the SSCP results, indicates that PCR-linked SSCP provides a reliable method for displaying sequence variation in cysteine protease genes of *H. contortus*.

According to the polymorphic indices (A, P and H_e) and the genetic variability statistics (H_t, H_s, D_{st} and G_{st}) based on the SSCP results, cysteine protease genes from *H. contortus* showed a considerable degree of polymorphism. Except for *locus AC-5*, all the *loci* were polymorphic, with a total of 20 alleles and a number of alleles per *locus* ranging from 2 to 8. *Locus AC-3* was the most polymorphic, followed by *locus GCP-7*, then *locus AC-1* and finally *locus AC-4*. The degree of allelic variation detected by SSCP agree with estimates of the nucleotide variability for all *loci*, except for *locus GCP-7* in which the number of polymorphic sites (S), the nucleotide diversity (π) and the heterozygosity per nucleotide site (θ) had the highest values in both the Spanish (SP) and the North American strains. The discordance between the SSCP results and sequencing analyses for *locus*

GCP-7 suggests that the number of regions of base pairing, rather than the primary structure of the molecule, is the main factor that determines the spatial conformation of the single-stranded DNA, since point mutations can be detectable by SSCP for fragments of >600 bp (Kukita *et al*. 1997).

The genetic diversity within populations (H_s) found in this study accounted for 85% of the total genetic diversity, while the genetic diversity between populations (D_s) accounted for 15% of the total genetic diversity. This is in accordance with Nei's (1978) $G_{st}=0{\cdot}154$ which measures the proportion of genetic diversity attributable to population differentiation. In agreement with these data, studies in 4 species of trichostrongylid nematodes, including *H. contortus*, indicated that 96–99% of nucleotide diversity is found within populations (Blouin *et al*. 1992).

A number of interesting reports have demonstrated the human influence on the ecology, geographical distribution and genetic diversity among different organisms, including plants (*Asimina tribola*) (Huang, Layne & Kubisiak, 2000), animals (*Ehhydra lutris*) (Larson *et al*. 2002) and parasites including protozoa such as *Plasmodium falciparum* and diverse nematodes (Blouin *et al*. 1992, 1995; Read & Taylor, 2001; Wootton *et al*. 2002). International commerce in livestock animals between Europe and North America could have led to gene flow among livestock species and then among the parasites they harbour. Nevertheless, important differences were found to occur between the SP and NA strains of *H. contortus* at some of the *loci*. Allelic frequencies based on SSCP results were significantly different between the two populations for each of the *loci AC-1*, *AC-3* and *GCP-7* (G test), with some alleles being detected exclusively in the SP strain (F_3, G_3, H_3 and C_7) or the NA strain (E_1 and D_7). Accordingly, the estimation of the nucleotide variability between strains indicated a mean distance between groups ranging from 0·0121 (0·0036) (*locus AC-4*) to 0·0638 (0·0068) (*locus GCP-7*). Although inter-geographical variability may be a consequence of a moderate geographical isolation, the influence of host species (goats and sheep for the SP and the NA strains, respectively) on genetic divergence should also be taken into account, as discussed by Zhu *et al*. (2000) who studied the mitochondrial DNA polymorphism within and among species of *Capillaria sensu lato* from Australian marsupials and rodents. While these authors did not find significant variation in SSCP profiles within morphospecies within a particular host species, significant variation occurred between morphospecies originating from different host species.

Even though many investigations indicate that genetic variation in parasitic nematodes is an issue of considerable practical and theoretical significance, either from a morphological (Mendoza-Leon, Luis & Martinez, 2001), biological (Watkins & Fernando, 1984), therapeutic (Hejmadi *et al*. 2000) and immunological (Goyal & Wakelin, 1993) point of view, the extent to which genetically determined variation affects the host–parasite interaction is still poorly understood (Wakelin, Farias & Bradley, 2002). Much more work is needed to elucidate the functional implications of the genetic variability in cysteine protease genes from *H. contortus* measured in this and previous studies.

We cordially thank Dr Juan Manuel Afonso (University of Las Palmas de Gran Canaria, Spain) for critical reading and invaluable help with the manuscript. We also thank all the staff of the Dr Prichard's Lab and Dr Beech's Lab at the Parasitology Institute (McGill University, Canada). This study has been supported by the Canary Government, by the Spanish Ministry of Science and Technology, and by the FEDER Funds (Project AGL2002-03528).

REFERENCES

AUTHIE, E., BOULANGE, A., MUTETI, D., LALMANACH, G., GAUTHIER, F. & MUSOKE, A. J. (2001). Immunisation of cattle with cysteine proteases of *Trypanosoma congolense*: targetting the disease rather than the parasite. *International Journal for Parasitology* **31**, 1429–1433.

BANIA, J., GATTI, E., LELOUARD, H., DAVID, A., CAPPELLO, F., WEBER, E., CAMOSSETO, V. & PIERRE, P. (2003). Human cathepsin S, but not cathepsin L, degrades efficiently MHC class II-associated invariant chain in nonprofessional APCs. *Proceedings of the National Academy of Sciences, USA* **100**, 6664–6669.

BARRETT, A. J. (1994). Classification of peptidases. *Methods in Enzymology* **224**, 1–15.

BARRETT, A. J. & KIRSCHKE, H. (1981). Cathepsin B, Cathepsin H, and cathepsin L. *Methods in Enzymology* **80**, 535–561.

BARRETT, A. J. & McDONALD, J. K. (1980). *Mammalian Proteases: a Glossary and Bibliography, Vol. 1: Endopeptidases*. Academic Press, London.

BAZAN, J. F. & FLETTERICK, J. R. (1988). Viral cysteine proteases are homologous to the trysin-like family of serine proteases: structural and functional implications. *Proceedings of the National Academy of Sciences, USA* **85**, 7872–7876.

BEECH, R. N., PRICHARD, R. K. & SCOTT, M. E. (1994). Genetic variability of the beta-tubulin genes in benzimidazole-susceptible and -resistant strains of *Haemonchus contortus*. *Genetics* **138**, 103–110.

BLACKHALL, W. J., LIU, H. Y., XU, M., PRICHARD, R. K. & BEECH, R. N. (1998*a*). Selection at a P-glycoprotein gene in ivermectin- and moxidectin-selected strains of *Haemonchus contortus*. *Molecular and Biochemical Parasitology* **95**, 193–201.

BLACKHALL, W. J., POULIOT, J. F., PRICHARD, R. K. & BEECH, R. N. (1998*b*). *Haemonchus contortus*: selection at a glutamate-gated chloride channel gene in ivermectin- and moxidectin-selected strains. *Experimental Parasitology* **90**, 42–48.

BLOUIN, M. S., DAME, J. B., TARRANT, C. A. & COURTNEY, C. H. (1992). Unusual population genetics of a parasitic nematode: mtDNA variation within and among populations. *Evolution* **46**, 470–476.

BLOUIN, M. S., YOWELL, C. A., COURTNEY, C. H. & DAME, J. B. (1995). Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* **141**, 1007–1014.

CAFFREY, C. R., HANSELL, E., LUCAS, K. D., BRINEN, L. S., ALVAREZ HERNANDEZ, A., CHENG, J., GWALTNEY, S. L. 2nd, ROUSH, W. R., STIERHOF, Y. D., BOGYO, M., STEVERDING, D. & MCKERROW, J. H. (2001). Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of *Trypanosoma brucei rhodesiense*. *Molecular and Biochemical Parasitology* **118**, 61–73.

COOMBS, G. H. & MOTTRAM, J. C. (1997). Parasite proteases and amino acid metabolism: possibilities for chemotherapeutic exploitation. *Parasitology* (Suppl.) **114**, S61–S80.

COX, G. N., PRATT, D., HAGEMAN, R. & BOISVENUE, R. J. (1990). Molecular cloning and primary sequence of a cysteine protease expressed by *Haemonchus contortus* adult worms. *Molecular and Biochemical Parasitology* **41**, 25–34.

DALTON, J. P., MCGONIGLE, S., ROLPH, T. P. & ANDREWS, S. J. (1996). Induction of protective immunity in cattle against infection with *Fasciola hepatica* by vaccination with cathepsin L proteases and with hemoglobin. *Infection and Immunity* **64**, 5066–5074.

DE LEON, M. P., YANAGI, T., KIKUCHI, M., MU, J., AYAU, O., MATTA, V., PAZ, M., JUÁREZ, S., KANBARA, H., TADA, I. & HIRAYAMA, K. (1998). Characterization of *Trypanosoma cruzi* by DNA polymorphism of the cruzipain gene detected by single-stranded DNA conformation polymorphism (SSCP) and direct sequencing. *International Journal for Parasitology* **28**, 1867–1874.

DIXIT, A. K., YADAV, S. C. & SHARMA, R. L. (2002). 28 kDa *Fasciola gigantica* cysteine protease in the diagnosis of prepatent ovine fasciolosis. *Veterinary Parasitology* **109**, 233–234.

DU, X., GUO, C., HANSELL, E., DOYLE, P. S., CAFFREY, C. R., HOLLER, T. P., MCKERROW, J. H. & COHEN, F. E. (2002). Synthesis and structure-activity relationship study of potent trypanocidal thiosemicarbazone inhibitors of the trypanosomal cysteine protease cruzain. *Journal of Medical Chemistry* **45**, 2695–2707.

EAKIN, A. E., MILLS, A. A., HARTH, G., MCKERROW, J. H. & CRAIK, C. S. (1992). The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. *The Journal of Biological Chemistry* **267**, 7411–7420.

GAILLARD, C. & STRAUSS, F. (1990). Ethanol precipitation of DNA with linear polyacrylamide as carrier. *Nucleic Acids Research* **18**, 378.

GELDHOF, P., CLAEREBOUT, E., KNOX, D., VERCAUTEREN, I., LOOSZOVA, A. & VERCRUYSSE, J. (2002). Vaccination of calves against *Ostertagia ostertagi* with cysteine protease enriched protein fractions. *Parasite Immunology* **24**, 263–270.

GLAZER, A. N. & SMITH, E. L. (1971). Papain and other plant sulfhydryl proteolytic enzymes. In *The Enzymes, 3rd Edn* (ed. Boyer, P. D.). Academic Press, New York.

GOYAL, P. K. & WAKELIN, D. (1993). Influence of variation in host strain and parasite isolate on inflammatory and antibody responses to *Trichinella spiralis*. *Parasitology* **106**, 371–378.

GREENHALGH, C. J., BECKHAM, S. A. & NEWTON, S. E. (1999). Galectins from sheep gastrointestinal nematode parasites are highly conserved. *Molecular and Biochemical Parasitology* **98**, 285–289.

HEJMADI, M. V., JAGANNATHAN, S., DELANY, N. S., COLES, G. C. & WOLSTENHOLME, A. J. (2000). L-glutamate binding sites of parasitic nematodes: an association with ivermectin resistance? *Parasitology* **120**, 535–545.

HIGGINS, D., THOMPSON, J., GIBSON, T., THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

HOEKSTRA, R., CRIADO-FORNELIO, A., FAKKELDIJ, J., BERGMAN, J. & ROOS, M. H. (1997). Microsatellites of the parasitic nematode *Haemonchus contortus*: polymorphism and linkage with a direct repeat. *Molecular and Biochemical Parasitology* **89**, 97–107.

HOEKSTRA, R., OTSEN, M., LENSTRA, J. A. & ROOS, M. H. (1999). Characterisation of a polymorphic Tc1-like transposable element of the parasitic nematode *Haemonchus contortus*. *Molecular and Biochemical Parasitology* **102**, 157–166.

HOEKSTRA, R., OTSEN, M., TIBBEN, J., LENSTRA, J. A. & ROOS, M. H. (2000). Transposon associated markers for the parasitic nematode *Haemonchus contortus*. *Molecular and Biochemical Parasitology* **105**, 127–135.

HUANG, H., LAYNE, D. R. & KUBISIAK, T. L. (2000). RAPD inheritance and diversity in Pawpaw (*Asimina triloba*). *Journal of the American Society for Horticultural Science* **125**, 454–459.

JASMER, D. P., ROTH, J. & MYLER, P. J. (2001). Cathepsin B-like cysteine proteases and *Caenorhabditis elegans* homologues dominate gene products expressed in adult *Haemonchus contortus* intestine. *Molecular and Biochemical Parasitology* **116**, 159–169.

JUKES, T. H. & CANTOR, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism* (ed. Munro, H. N.), pp. 21–231. Academic Press, New York.

KARANU, F. N., RURANGIRWA, F. R., MCGUIRE, T. C. & JASMER, D. P. (1993). *Haemonchus contortus*: identification of proteases with diverse characteristics in adult worm excretory–secretory products. *Experimental Parasitology* **77**, 362–371.

KARANU, F. N., RURANGIRWA, D. P., MCGUIRE, T. C. & JASMER, D. P. (1997). *Haemonchus contortus*: inter- and intrageographic isolate heterogeneity of proteases in adult worm excretory–secretory products. *Experimental Parasitology* **86**, 89–91.

KUKITA, Y., TAHIRA, T., SOMMER, S. S. & HAYASHI, K. (1997). SSCP analysis of long DNA fragments in low pH gel. *Human Mutation* **10**, 400–407.

KUMAR DUBEY, V. & JAGANNADHAM, M. V. (2003). Procerain, a stable cysteine protease from the latex of *Calotropis procera*. *Phytochemistry* **62**, 1057–1071.

KUMAR, S., TAMURA, K., JAKOBSEN, I. B. & NEI, M. (2001). MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* **17**, 1244–1245.

KWA, M. S., KOOYMAN, F. N., BOERSEMA, J. H. & ROOS, M. H. (1993). Effect of selection for benzimidazole resistance in *Haemonchus contortus* on beta-tubulin isotype 1 and isotype 2 genes. *Biochemical and Biophysical Research Communications* **191**, 413–419.

LARSON, S., JAMESON, R., ETNIER, M., FLEMING, M. & BENTZEN, P. (2002). Loss of genetic diversity in sea otters (*Enhydra lutris*) associated with the fur trade of the 18th and 19th centuries. *Molecular Ecology* **11**, 1899–1903.

LEWIS, P. O. (1994). GeneStat-PC, v. 3.3. N. C. State University of Raleigh, North Carolina.

LI, Z., CHEN, X., DAVIDSON, E., ZWANG, O., MENDIS, C., RING, C. S., ROUSH, W. R., FEGLEY, G., LI, R., ROSENTHAL, P. J. *et al*. (1994). Anti-malarial drug development using models of enzyme structure. *Chemistry and Biology* **1**, 31–37.

LOUKAS, A., DOWD, A. J., PROCIV, P. & BRINDLEY, P. J. (2000). Purification of a diagnostic, secreted cysteine protease-like protein from the hookworm *Ancylostoma caninum*. *Parasitology International* **49**, 327–333.

McKERROW, J. H. (1989). Parasite proteases. *Experimental Parasitology* **68**, 111–115.

MENDOZA-LEON, A., LUIS, L. & MARTINEZ, C. (2001). The beta-tubulin gene region as a molecular marker to distinguish *Leishmania* parasites. *Methods in Molecular Biology* **79**, 61–83.

MORIHARA, K. (1974). Comparative specificity of microbial proteases. *Advances in Enzymology* **41**, 179–243.

NA, B. K., LEE, H. J., CHO, S. H., LEE, H. W., CHO, J. H., KHO, W. G., LEE, J. S., LEE, J. S., SONG, K. J., PARK, P. H., SONG, C. Y. & KIM, T. S. (2002). Expression of cysteine protease of *Clonorchis sinensis* and its use in serodiagnosis of clonorchiasis. *Journal of Parasitology* **88**, 1000–1006.

NEI, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590.

NEI, M. (1986). Definition and estimation of fixation indices. *Evolution* **40**, 643–645.

NESTERENKO, M. V., TILLEY, M. & UPTON, S. J. (1995). A metallo-dependent cysteine protease of *Cryptosporidium parvum* associated with the surface of sporozoites. *Microbios* **83**, 77–88.

NEYRA, V., CHAVARRY, E. & ESPINOZA, J. R. (2002). Cysteine proteases Fas1 and Fas2 are diagnostic markers for *Fasciola hepatica* infection in alpacas (*Lama pacos*). *Veterinary Parasitology* **105**, 21–32.

NORTH, M. J. (1982). Comparative biochemistry of the proteases of eukaryotic microorganisms. *Microbiological Research* **46**, 308–340.

OTSEN, M., PLAS, M. E., LENSTRA, J. A., ROOS, M. H. & HOEKSTRA, R. (2000). Microsatellite diversity of isolates of the parasitic nematode *Haemonchus contortus*. *Molecular and Biochemical Parasitology* **110**, 69–77.

PRATT, D., ARMES, L. G., HAGEMAN, R., REYNOLDS, V., BOISVENUE, R. J. & COX, G. N. (1992). Cloning and sequence comparisons of four distinct cysteine proteases expressed by *Haemonchus contortus* adult worms. *Molecular and Biochemical Parasitology* **51**, 209–218.

PRATT, D., COX, G. N., MILHAUSEN, M. J. & BOISVENUE, R. J. (1990). A developmentally regulated cysteine protease gene family in *Haemonchus contortus*. *Molecular and Biochemical Parasitology* **43**, 181–192.

RAWLINGS, N. D. & BARRETT, A. J. (1993). Evolutionary families of peptidases. *The Biochemical Journal* **290**, 205–218.

READ, A. F. & TAYLOR, L. H. (2001). The ecology of genetically diverse infections. *Science* **292**, 1099–1102.

REHMAN, A. & JASMER, D. P. (1998). A tissue specific approach for analysis of membrane and secreted protein antigens from *Haemonchus contortus* gut and its application to diverse nematode species. *Molecular and Biochemical Parasitology* **97**, 55–68.

ROSENTHAL, P. J., SIJWALI, P. S., SINGH, A. & SHENAI, B. R. (2002). Cysteine proteases of malaria parasites: targets for chemotherapy. *Current Pharmaceutical Design* **8**, 1659–1672.

ROZAS, J. & ROZAS, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.

SAITOU, N. & NEI, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

SAJID, M. & McKERROW, J. H. (2002). Cysteine proteases of parasitic organisms. *Molecular and Biochemical Parasitology* **120**, 1–21.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

SANGSTER, N. C., BANNAN, S. C., WEISS, A. S., NULF, S. C., KLEIN, R. D. & GEARY, T. G. (1999). *Haemonchus contortus*: sequence heterogeneity of internucleotide binding domains from P-glycoproteins. *Experimental Parasitology* **91**, 250–257.

SKUCE, P. J., REDMOND, D. L., LIDDELL, S., STEWART, E. M., NEWLANDS, G. F. J., REDMOND, D. L., SKUCE, P. J., KNOX, D. P. & SMITH, W. D. (1999). Molecular cloning and characterization of gut-derived cysteine proteases associated with a host protective extract from *Haemonchus contortus*. *Parasitology* **119**, 405–412.

SOKAL, R. R. & ROHLF, J. F. (1981). *Biometry*. Freeman & Co., New York.

SVENSSON, M. D., SCARAMUZZINO, D. A., SJOBRING, U., OLSEN, A., FRANK, C. & BESSEN, D. E. (2000). Role for a secreted cysteine protease in the establishment of host tissue tropism by group A streptococci. *Molecular Microbiology* **38**, 242–253.

TORT, J., BRINDLEY, P. J., KNOX, D., WOLFE, K. H. & DALTON, J. P. (1999). Proteases and associated genes of parasitic helminths. *Advances in Parasitology* **43**, 161–266.

WAKELIN, D., FARIAS, S. E. & BRADLEY, J. E. (2002). Variation and immunity to intestinal worms. *Parasitology* **125** (Suppl.), S39–S50.

WATKINS, A. R. & FERNANDO, M. A. (1984). Arrested development of the rabbit stomach worm *Obeliscoides cuniculi*: manipulations of the ability to arrest through processes of selection. *International Journal for Parasitology* **14**, 559–570.

WOOTTON, J. C., FENG, X., FERDIG, M. T., COOPER, R. A., MU, J., BARUCH, D. I., MAGILL, A. J. & SU, X. Z. (2002). Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature, London* **418**, 320–323.

ZHU, X., SPRATT, D. M., BEVERIDGE, I., HAYCOCK, P. & GASSER, R. B. (2000). Mitochondrial DNA polymorphism within and among species of *Capillaria* sensu lato from Australian marsupials and rodents. *International Journal for Parasitology* **30**, 933–938.

ZIEBUHR, J., BAYER, S., COWLEY, J. A. & GORBALENYA, A. E. (2003). The 3C-like protease of an invertebrate nidovirus links coronavirus and potyvirus homologs. *Journal of Virology* **77**, 1415–1426.