Comparative evaluation of *Giardia duodenalis* sequence data

C. M. WIELINGA *and* R. C. A. THOMPSON*

WHO Collaborating Centre for the Molecular Epidemiology of Parasitic Infections and the State Agricultural Biotechnology Centre, School of Veterinary and Biomedical Sciences, Murdoch University, South Street, Murdoch, Western Australia 6150, Australia

(*Received 21 January 2007; revised 3 March and 26 April 2007; accepted 27 April 2007; first published online 19 June 2007*)

SUMMARY

A review of the *Giardia duodenalis* sequences currently available on the GenBank database was completed to compare the different genotyping loci (small subunit ribosomal DNA, glutamate dehydrogenase, triose-phosphate isomerase and beta giardin) for their ability to discern assemblage and subassemblage groups and infer phylogenetic relationships. In total, 405 *Giardia duodenalis* sequences were sorted and aligned to examine the substitutions within and between the assemblages – A and B (zoonotic), C and D (dogs), E (livestock), F (cats) and G (rodents). It was found that all of the genes could reproducibly group isolates into their assemblages and that the AI/AII subassemblage groups were robust and identifiable at all loci. However, the assemblage B subgroups were not reproducible at half of the loci (small subunit ribosomal DNA and beta giardin), not due to their conserved nature, but because there was insufficient sequence data of reference isolates available for comparison. It is anticipated that further investigation of these loci may reveal the core subgroups of this medically important and zoonotic assemblage and also those of others. The closer, more recent, phylogenetic relationships amongst the assemblages appear to be resolved; however, more sequence data from the current loci, and possibly new loci, will be required to establish the remaining relationships.

Key words: *Giardia*, genotyping, *SSU rDNA*, *gdh*, *tpi*, *β giardin*, *ef1α*, consensus sequences, phylogenetics.

INTRODUCTION

Giardia are seemingly ubiquitous intestinal parasites of vertebrates, found in all classes examined to date (Thompson *et al*. 1990; Adam, 2001). Species classification of *Giardia* has been dynamic. Originally many species were described based on host information and then these were re-classified into 3 species based on gross morphological differences – *Giardia agilis* (amphibians), *Giardia muris* (rodents and birds) and *Giardia duodenalis* (mammals, birds and rodents) (Filice, 1952). With the advent of more sophisticated ultrastructural methods of morphological characterization, additional species have been described – *Giardia psittaci* in parakeets (Erlandsen and Bemrick, 1987), *Giardia ardeae* in herons (Erlandsen *et al*. 1990) and *Giardia microti* in muskrats and voles (van Keulen *et al*. 1998); and it is expected that this trend will continue as the number of host species examined increases, particularly with respect to the recognition of previously described species (Thompson and Monis, 2004).

G. duodenalis (also referred to as *G. lamblia* and *G. intestinalis*) is the only species recovered from humans to date and hence has received the most attention. Early research on human isolates using a variety of molecular tools demonstrated consistent heterogeneity and divergence within *G. duodenalis*. Subgroups precursory to the current system were originally described by Nash – groups 1, 2 and 3 (Nash and Keister, 1985); Andrews – groups I, II, III and IV (Andrews *et al*. 1989) and Homan – Polish and Belgian (Homan *et al*. 1992). These subgroups were found to be equivalent and the nomenclature was standardized to 'Assemblages' AI and AII and BIII/BIV (Monis *et al*. 1996; Adam, 2001).

More recent research, both on isolates from a wider host range and using molecular techniques directly on host samples, has lead to the recovery and identification of more genotypes of *G. duodenalis* from a range of domestic and wild animals. Genotypes isolated include the original A and B assemblages previously detected in humans (and hence potentially zoonotic), as well as new and apparently host specific genotypes currently designated assemblage C and D, found in dogs (Hopkins *et al*. 1997; Monis *et al*. 1998); E, found in hoofed livestock (Ey *et al*. 1997); F, found in cats (Mayrhofer *et al*. 1995) and G, found in rats and mice (Monis *et al*. 1999). The assemblages of *G. duodenalis*, although apparently identical in morphology, demonstrate

* Corresponding author. Tel: +61 08 9360 2466. Fax: +61 08 9310 4144. E-mail: A.Thompson@murdoch.edu.au

 © 2007 Cambridge University Press
doi:10.1017/S0031182007003071 Printed in the United Kingdom

large genetic divergence and this has led to the proposition that *G. duodenalis* may be a species complex (Andrews *et al.* 1989; Mayrhofer *et al.* 1995; Monis *et al.* 1996). However, this remains unresolved, as to date convention has relied on morphological variation to describe different *Giardia* species and not genotypic or phenotypic variation alone.

An ongoing issue in *Giardia* research has been the question of zoonotic transmission (Thompson and Monis, 2004; Caccio *et al.* 2005; Hunter and Thompson, 2005). In order to address this, molecular epidemiological studies have involved genotyping many new isolates and so numerous methods have been described in the literature (Weiss *et al.* 1992; Ey *et al.* 1993; van Keulen *et al.* 1995; Baruch *et al.* 1996; Monis *et al.* 1996, 1999; Hopkins *et al.* 1997; Caccio *et al.* 2002). Variations between the methods lie in the gene(s) utilized (small subunit rDNA – *SSU rDNA*, variable surface protein – *vsp*, glutamate dehydrogenase – *gdh*, triose phosphate isomerase -*tpi*, elongation factor 1 alpha – *ef1α*, beta giardin – *β giardin*), the region of the gene examined (*Giardia* specific, *G. duodenalis* specific, assemblage specific) and method of analysis (mainly direct PCR for product size polymorphism, restriction fragment length polymorphisms (RFLP) or sequencing). Depending on the aim of each study, different groups have been promoting their techniques for different genotyping applications. As the amount of sequence information increases, more can be learnt about the different genes and their uses in genotyping and phylogenetic analyses.

Our aims in this review were to evaluate the value of loci currently used for the genotypic characterization of *Giardia* in terms of diagnosis, taxonomy, molecular epidemiology and phylogeny. In order to achieve this, we have collated all of the available sequence data from the GenBank database (National Center for Biotechnology Information, NCBI, www.ncbi.nlm.nih.gov) for *G. duodenalis* at the *SSU rDNA*, *gdh, tpi, ef1α* and *β giardin* loci. As researchers use this important resource to design primers, develop RFLPs and identify isolates, an overview of the currently available sequence data seemed appropriate.

Initially, the ability of different loci to reliably and consistently assign isolates to an assemblage/subassemblage of *G. duodenalis* was assessed. Our original focus in this area was the suitability of *SSU rDNA* for this task as this locus has some desirable traits that favour its use in genotyping, notably, a high copy number. *Giardia* has been estimated to have 60 copies of the rDNA repeat (Boothroyd *et al.* 1987; Edlind and Chakraborty, 1987), whereas the other structural, metabolic and housekeeping genes are estimated to be single or low copy number (Yee and Dennis, 1992). High copy number of a gene confers 2 immediate advantages. (i) There is a greater success rate of amplification due to increased availability of starting material. This becomes significant as researchers increasingly attempt to amplify directly from environmental samples, which are typically lower in target DNA and contain more PCR inhibitors (relative to cultured isolates). (ii) The detection of mixed templates in a sample is more likely. Since the advent of genotyping, procedures applied directly to clinical and environmental samples, mixed templates have become more apparent (Weiss *et al.* 1992; Upcroft and Upcroft, 1994; Amar *et al.* 2002; Guy *et al.* 2004) and their detection is significant. Analyses employing the high copy number *SSU rDNA* are frequently able to detect mixed templates in a single PCR (Hopkins *et al.* 1997; Berrilli *et al.* 2004) whereas this is seldom reported for the low copy genes (Lalle *et al.* 2005*b*), normally requiring multiple PCRs that produce alternate results (unpublished observations). A final reason for concentrating on the *SSU rDNA* locus, is it is the traditional gene sequence used for identification and phylogenetic analyses (Sogin *et al.* 1989; van Keulen *et al.* 1993) and valuable contributions could be made to the rDNA databases for future analyses on the systematics of *Giardia* species.

We also examine the phylogenetic trends demonstrated by the different loci used for genotyping *G. duodenalis*. The last comprehensive study investigating these relationships within *G. duodenalis* (Monis *et al.* 1999) demonstrated variable results between the loci and it was hypothesized that further sequence information may resolve this. The relationships were of interest because an understanding of the history of evolution and adaptability of the assemblages in terms of the relationships of their present hosts may provide insights into their current potential scope.

MATERIALS AND METHODS

GenBank survey

Searches were conducted of the GenBank database to gather as many sequences as possible of *Giardia duodenalis* to be organized and compared in further analyses. Of the more than 10 000 *Giardia* sequences on GenBank, approximately 90% were related to genome sequencing projects and contained predominantly uncharacterized products. Of the remainder, 405 were *G. duodenalis* sequences of interest, with 104 *SSU rDNA*, 96 *gdh*, 53 *tpi*, 141 *β giardin* and 11 *ef1α*. The details of the isolates retrieved are given in Tables 1–4, and below. In Tables 1–4, reference isolates are given in bold, asterisks mark positions outside of the alignment range, dashes either represent missing data in the isolate details or in the sequences (sequences do not extend to that base position), lower case letters were used to show small subgroups, universal code degenerate bases were

used in the consensus sequences (R=A/G, Y=C/T, K=G/T, M=A/C), samples categorized as 'A', 'B' or 'B-central' were unable to be aligned to a reference subgroup and the sample 'A other' refers to an isolate divergent from both AI and AII.

Isolate details for *SSU rDNA* are given in Table 1. Sequences that were deemed unsuitable for the analyses because they contained mixed, degenerate or unusual substitutions (that had not been reproduced) were omitted from the table and from the analyses. These sequences were as follows: ambiguous/degenerate sequences DQ118557, DQ118558 and DQ112665, mixed sequences AY775186, AY775189, AY775193 and AJ293300, hypervariable sequences AY130270, AY130272 and AY130273 and superseded 'Portland-1' sequence X05396.

Isolate details for *gdh*, *tpi* and *β giardin* are given in Tables 2–4. These tables include the intra-genotypic (intra-assemblage) substitution details for each assemblage. Some isolate sequences (listed below) were omitted from these tables (but not from the analyses) because there were no intra-genotypic variations to tabulate. For example, some assemblages were represented on the database by only one sequence, or the available sequences were identical (matching where they overlapped) or were too small to cover regions of variation within that assemblage. In addition, some isolates were deposited on the database numerous times (reference isolates) and their sequences were found to be identical and hence only the longest available sequences were tabulated. For the *gdh* locus, samples not included in Table 2 were – 'Ad-1'/L40509 ('Ad-1'/AY178735[21] was longer), 5 matching assemblage F sequences [cats 'Ad-23', 'Ad-131', 'Ad-142' and 'Ad-154' (1114–1123 bp, AF069057[7] and AY178742–44[24], Australia) and cat 'Ct1,2,3' (177 bp, AB199739[26], Japan)], 3 smaller assemblage A sequences [dog 'D3,5,8–15,17–23' (177 bp, AB199735[26], Japan), calf 'cf2' (177 bp, AB199742[26], Japan) and human 'NLH 37' (399 bp, AY826196[15], Netherlands)] and a smaller assemblage C sequence [dog 'D2' (177 bp, AB199736[26], Japan)]. For the *tpi* locus, samples not included in Table 3 were – the single assemblage F sequence [cat 'Ad-23' (479 bp, AF069558[7], Australia)] and the 2 matching assemblage G sequences [rat isolates 'Ad-157' (468 bp, AF069562[7], Australia) and '2135' (428 bp, AY228640[32], USA)]. For the *β giardin* locus, samples not included in Table 4 were – the replicates of 'Portland 1' (X14185, M36728 and X07919 in favour of longer X85958[36]), 'WB' (AY258617 in favour of longer XM763377[19]) and 'H3' (AY258616 in favour of longer DQ116605[38]) as well as the single sequences for assemblages C and F [dog 'A29' (511 bp, AY545646[28], Italy) and cat 'A101' (753 bp, AY647264[38], Italy) respectively]. There was no sequence available for assemblage G.

The references for the Accession numbers above and in Tables 1–4 (represented in superscript) were as follows. (1) Thompson *et al*. (2000), (2) Sogin *et al*. (1989), (3) van Keulen *et al*. (1995), (4) Not yet published, Xiao, S., South China Agricultural University, 2005, (5) Healey *et al*. (1990), (6) Upcroft *et al*. (1994), (7) Monis *et al*. (1999), (8) Abe *et al*. (2005*a*), (9) Abe *et al*. (2005*b*), (10) Not yet published, Abe, N., Osaka City Institute of Public Health and Environmental Sciences, 2005, (11) Not yet published, Berrilli, F., University of Rome, 2002, (12) Berrilli *et al*. (2004), (13) Trout *et al*. (2004), (14) Yong *et al*. (2000), (15) van der Giessen *et al*. (2006), (16) van Keulen *et al*. (1991), (17) Weiss *et al*. (1992), (18) Hunt *et al*. (2000), (19) McArthur *et al*. (2000), (20) Yee and Dennis (1992), (21) Ey *et al*. (1997), (22) Leonhard *et al*. (2006), (23) Monis *et al*. (1996), (24) Ey, P., University of Adelaide, 2002, submitted as a set with those from Ey *et al*. (1997), (25) Monis *et al*. (1998), (26) Itagaki *et al*. (2005), (27) Matsubayashi *et al*. (2005), (28) Robertson *et al*. (2006), (29) Mowatt *et al*. (1994), (30) Trout *et al*. (2003), (31) Baruch *et al*. (1996), (32) Sulaiman *et al*. (2003), (33) Sulaiman *et al*. (2004), (34) Not yet published, Mowatt, M., National Institute of Allergy and Infectious Diseases, 1992, (35) Not yet published, Wielinga, C., Murdoch University, 2005, (36) Holberton and Marshall (1995), (37) Not yet published, Volotao, A., Institute Ozwaldo Cruz, 2006, (38) Lalle *et al*. (2005*a*), (39) Caccio, S., Institute Superiore di Sanita, 2002 submitted as a set with those from Caccio *et al*. (2002), (40) Not yet published, Di Giovanni, G., Texas A & M University, 2005, (41) Caccio *et al*. (2002), (42) In the Press, Abe, N., Seikatsu Eisei, 2005, (43) Santin *et al*. (2003).

The *ef1α* isolates retrieved from the database included 2 '*G. lamblia*' sequences (L23957 and D14342), 'WB' (XM762925) and 9 previously presented and analysed samples ['Ad-2', 'Ad-12', 'Ad–23', 'Ad–28', 'Ad-136', 'Ad-148', 'Ad-157', 'P15' and 'BAH-12'; AF069568-75, (Monis *et al*. 1999)]. As there were so few sequences, and they had been presented elsewhere, they were not included in the current study.

All of the *G. duodenalis* genotyping sequences were derived from genomic DNA. Most of the sequences were obtained from direct sequencing (occasionally cloned) of PCR products amplified from environmental samples. The reference isolate sequences were usually derived from isolates grown in culture or passaged through suckling mice.

Alignments

Sequences gathered in the initial GenBank survey required sorting into their different genes, assemblages and subassemblages as well as alignment along the gene. The purpose of this was to establish

Table 1. Small subunit rDNA, isolate information

	Isolate	Source/Origin	Size (bp)	Accession no.[Ref]	Assemblage
COMPLETE	Cat2 BAC2	Cat/Australia	1418	AF199445[1]	AI
	Portland1	Human/USA	1453	M54878[2]	AI
	BAH40c11	Human/Australia	1418	AF199446[1]	AII
	AMC-4	Human/Netherlands	1453	U09491[3]	B
	CM	Human/USA	1452	U09492[3]	B
	BAH12c14	Human/Australia	1418	AF199447[1]	B(III)
	Dog19	Dog/Australia	1420	AF199449[1]	C
	Dog6	Dog/Australia	1420	AF199443[1]	D
	Guangzhou calf	Calf/China	1447	DQ157272[4]	E
	Goat1 BAG1	Goat/Australia	1416	AF199448[1]	E
	Cat7 BAC7	Cat/Australia	1417	AF199444[1]	F
	Rat2	Rat/Australia	1407	AF199450[1]	G
5′ END (medium)	BRIS/83/HEPU/106	Human/Australia	637	X52949[5]	A(I)
	BRIS/91/HEPU/1279	Human/Australia	495	L29192[6]	B
	BAH12	Human/Australia	455	AF113897[7]	B(III)
	Ad28	Human/Australia	422	AF113898[7]	B(IV)
	Ad136	Dog/Australia	419	AF113899[7]	C
	Ad148	Dog/Australia	466	AF113900[7]	D
	P15	Pig/Czech. R	461	AF113902[7]	E
	Ad23	Cat/Australia	412	AF113901[7]	F
	Ad157	Rat/Australia	384	AF113896[7]	G
5′ END	GH-125	Human/Japan	125	AB195219[8]	A
	GH-126	Human/Japan	125	AB195220[8]	A
	GF-1	Ferret/Japan	125	AB159796[9]	A
	GD-99H	Dog/Japan	125	AB218601[10]	A
	0711g	Water/Italy	205	AY130269[11]	A
	2811g	Water/Italy	205	AY130271[11]	A
	CGP	Water/Italy	205	AY130274[11]	A
	CGR	Water/Italy	205	AY130275[11]	A
	Nemi	Water/Italy	205	AY130276[11]	A
	0412u	Water/Italy	205	AY130277[11]	A
	1010g	Water/Italy	205	AY130278[11]	A
	1212g	Water/Italy	205	AY130279[11]	A
	1212i	Water/Italy	205	AY130280[11]	A
	1212u	Water/Italy	205	AY130281[11]	A
	dogizp5	Dog/Italy	210	AY775188[12]	A
	dogizp7	Dog/Italy	210	AY775190[12]	A
	—	Cattle/USA	292	AY655700[13]	A
	K1	Human/Korea	292	AJ278959[14]	A
	K2	Human/Korea	292	AJ293295[14]	A
	CA1	Human/China	292	AJ293296[14]	A
	CA14	Human/China	292	AJ293297[14]	A
	CA18	Human/China	292	AJ293298[14]	A
	CA13	Human/China	292	AJ293299[14]	A
	KC1	Human/Korea	292	AJ293301[14]	A
	NLH20	Human/Netherlands	302	AY826204[15]	A
	NLH45	Human/Netherlands	302	AY826205[15]	A
	NLH37	Human/Netherlands	256	AY826206[15]	A
	NLR118	Roe deer/Netherlands	301	DQ100287[15]	A
	catizp1	Cat/Italy	210	AY775201[12]	A/F
	GH-135	Human/Japan	126	AB195221[8]	B
	NLH13	Human/Netherlands	303	AY826201[15]	B
	NLH28	Human/Netherlands	281	AY826202[15]	B
	NLH25	Human/Netherlands	303	AY826203[15]	B
	NLH35	Human/Netherlands	303	AY826207[15]	B
	GD-29H	Dog/Japan	126	AB218600[10]	C
	GD-143	Dog/Japan	126	AB218603[10]	C
	dogizp1	Dog/Italy	211	AY775184[12]	C
	dogizp2	Dog/Italy	211	AY775185[12]	C
	dogizp4	Dog/Italy	211	AY775187[12]	C
	dogizp8	Dog/Italy	211	AY775191[12]	C
	dogizp9	Dog/Italy	211	AY775192[12]	C
	dogizp11	Dog/Italy	211	AY775194[12]	C
	dogizp12	Dog/Italy	211	AY775195[12]	C
	dogizp13	Dog/Italy	211	AY775196[12]	C

Table 1. *(cont.)*

	Isolate	Source/Origin	Size (bp)	Accession no.[Ref]	Assemblage
	dogizp14	Dog/Italy	211	AY775197[12]	C
	dogizp15	Dog/Italy	211	AY775198[12]	C
	dogizp17	Dog/Italy	211	AY775200[12]	C
	GD-89H	Dog/Japan	126	AB218599[10]	D
	GD-142	Dog/Japan	126	AB218602[10]	D
	dogizp16	Dog/Italy	211	AY775199[12]	D
	NLD37	Dog/Netherlands	301	AY827496[15]	D
	NLDE3	Dog/Netherlands	303	AY827497[15]	D
	CALFIZP1	Calf/Italy	280	AY297957[12]	E
	CALFIZP2	Calf/Italy	209	AY297958[12]	E
	CALFIZP3	Calf/Italy	280	AY297959[12]	E
	—	Cattle/USA	292	AY655701[13]	E
	NLS352	Sheep/Netherlands	301	AY826208[15]	E
	NLS387	Sheep/Netherlands	301	AY826209[15]	E
	NLG409	Goat/Netherlands	301	AY826210[15]	E
	Guangzhou Calf	Calf/China	334	DQ157271[4]	E
3′ END	Portland1-CCh	Human/USA	75	M73686[16]	AI
	E-2/M	Human/Egypt	183	M90524[17]	AII
	JH	Human/USA	183	M92052[17]	AII
	AB	Human/Peru	183	M92053[17]	AII
	Be-1	Beaver/Canada	183	M90523[17]	B
	E-9/M	Human/Egypt	183	M91471[17]	B
	G1M	Human/Peru	183	M91472[17]	B
	PM	Human/USA	183	M91473[17]	B
	CM	Human/USA	183	M91474[17]	B
	GS/M-H7	Human/USA	183	M91475[17]	B
	WaicalfC1H9	Calf/New Zealand	152	AF239840[18]	E
	ManacalfC13H3	Calf/New Zealand	152	AF239841[18]	E

the coverage of each gene (the amount of the gene being represented in length and in sample numbers) to determine the maximum continuous alignment length possible (to aid the resolution of relationship analyses) and to ascertain the variation in sample representation over that length (to gauge strength and accuracy of regions of the alignment). The alignments were also used to examine the inter- and intra- genotypic substitutions that determine the existing and potential groupings/subgroupings. Once these groups had been established, their continuity across the loci could also be investigated. Initially sequences were grouped into their respective loci and then each was analysed by multiple sequence alignment using either CLUSTAL W 1.83 (Thompson *et al*. 1994) or CLUSTAL X 1.81 (Thompson *et al*. 1997). Sequences were then sorted and grouped into their assemblages and subassemblages (where possible) according to the positions of their single nucleotide polymorphisms (SNPs) relative to previously characterized reference isolates. The isolates used as subassemblage reference isolates included **AI** – WB, Portland 1 and/or Ad-1, **AII** – JH, AB, KC8, Ad-2, Bris-136 and/or Ad-113, **BIII** – BAH-12 and **BIV** – Ad-7, Ad-19, Ad-28 and/or Ad-45 (Andrews *et al*. 1989; Nash, 1992; Weiss *et al*. 1992; Ey *et al*. 1992; Mayrhofer *et al*. 1995; Monis *et al*. 1996, 1999).

Once the alignments for each gene were sorted and grouped into their assemblages, they were trimmed to standardize their length for further analyses. Alignments were cut at each end at the point where the representation of any one assemblage ended. In this way the length of all of the alignments were limited by the length of the shortest (assemblage) alignment. The aim was to obtain the maximum possible length of continuous coverage of the gene by all assemblages (with at least 1 sample at any given point – but not necessarily single samples covering the whole length).

Consensus sequences

Once all of the original sequences had been aligned, the current consensus sequences could be determined for each assemblage/subassemblage per gene. Consensus sequences are defined as the common or shared sequence features for a sequence. Genetic analyses using individual sequences are often limited by the sequence's length and position along a gene. Sequences are therefore frequently omitted or cut to standardize data sets for further analyses. The advantage of using consensus sequences is increasing sequence length (to maximize the resolution of relationship analyses with increased characters) and confidence of the data set (with increasing

Table 2A. Glutamate dehydrogenase isolates, position and breakdown of intra-genotypic substitutions

	Isolate	Source/Origin	Size (bp)	Accession no.	Nucleotide position from start of gene														
Assemblage A					**237**	**246**	**603**	**621**	**699**	**753**	**807**	**831**	**861**	**867**	**870**	**894**	**902**	**1080**	**1266**
AI																			
	WB	**Human/Afghanistan**	**1350**	**XM773614**[19]	C	C	**T**	**C**	**T**	**C**	**C**	**C**	**T**	**T**	**T**	**C**	**C**	**T**	**G**
	Portland1	**Human/USA**	**1691**	**M84604**[20]	C	C	**T**	**C**	**T**	**C**	**C**	**C**	**T**	**T**	**T**	**C**	**C**	**T**	**G**
	Ad-1	**Human/Australia**	**1128**	**AY178735**[21]	C	C	**T**	**C**	**T**	**C**	**C**	**C**	**T**	**T**	**T**	**C**	**C**	**T**	**G**
	GD-99H	Dog/Japan	592	AB218607[10]	C	C	**T**	**C**	**T**	**C**	—	—	—	—	—	—	—	—	—
	GM Dog 10	Dog/Germany	431	DQ417364[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 15	Dog/Germany	432	DQ417365[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 18	Dog/Germany	432	DQ417366[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 58	Dog/Germany	420	DQ417367[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 9	Dog/Germany	424	DQ417368[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 59	Dog/Germany	424	DQ414235[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 61	Dog/Germany	424	DQ414236[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 11	Dog/Germany	432	DQ417369[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 17	Dog/Germany	432	DQ414237[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 20	Dog/Germany	432	DQ414238[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 51	Dog/Germany	432	DQ414239[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 53	Dog/Germany	432	DQ414240[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 55	Dog/Germany	432	DQ414241[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
	GM Dog 57	Dog/Germany	432	DQ414242[22]	—	C	**T**	**C**	—	—	—	—	—	—	—	—	—	—	—
AI – like																			
	GF-1	Ferret/Japan	592	AB159795[9]	C	C	**T**	T	**T**	**C**	—	—	—	—	—	—	—	—	—
A other																			
	NLR118	Roe deer/Netherlands	652	DQ100288[15]	C	C	C	**C**	C	**C**	**C**	**C**	**T**	C	C	—	—	—	—
AII – like																			
	GH-126	Human/Japan	592	AB195223[8]	**t**	**t**	C	T	**T**	T	—	—	—	—	—	—	—	—	—
	NLH20	Human/Netherlands	744	AY826194[15]	**t**	**t**	C	T	**T**	T	T	T	C	C	C	—	—	—	—
AII																			
	NLH45	Human/Netherlands	421	AY826195[15]	**t**	C	C	T	—	—	—	—	—	—	—	—	—	—	—
	GH-125	Human/Japan	592	AB195222[8]	C	C	C	T	C	T	—	—	—	—	—	—	—	—	—
	Ad-2	**Human/Australia**	**690**	**L40510**[23]	C	C	C	T	C	T	T	T	C	—	—	—	—	—	—
	Bris-136 (BRIS/83/HEPU/136)	**Human/Australia**	**1122**	**AY178737**[21]	C	C	C	T	C	T	T	T	C	C	C	T	T	C	A
	Ad-113	**Human/Australia**	**1122**	**AY178736**[24]	C	C	C	T	C	T	T	T	C	C	C	T	T	C	A
	Consensus sequence A				C	C	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**R**

Assemblage C				**249**	**586**	**835**	**846**								
Ad-137	Dog/Australia	1114	U60983[25]	**C**	**R**	**R**	C								
Ad-147	Dog/Australia	690	U60985[25]	**C**	**R**	**g**	C								
Ad-141	Dog/Australia	690	U60984[25]	**C**	**R**	A	C								
Ad-136	Dog/Australia	1094	U60982[25]	**C**	G	A	**t**								
GM Dog 16	Dog/Germany	428	DQ417370[22]	T	G	—	—								
GM Dog 44	Dog/Germany	428	DQ414243[22]	T	G	—	—								
GM Dog 50	Dog/Germany	428	DQ414244[22]	T	G	—	—								
GM Dog 54	Dog/Germany	428	DQ414245[22]	T	G	—	—								
Consensus sequence C				**Y**	G	**R**	C								
Assemblage D				**222**	**414**	**603**	**615**	**624**	**1122**						
Ad-148	Dog/Australia	1120	U60986[25]	C	T	**t**	**g**	G	**Y**						
GD-142	Dog/Japan	592	AB218606[10]	C	T	**t**	**g**	**a**	—						
NLE3	Dog/Netherlands	440	AY827498[15]	—	T	C	A	G	—						
GM Dog 60	Dog/Germany	424	DQ417372[22]	—	T	C	A	G	—						
GM Dog 19	Dog/Germany	424	DQ417371[22]	—	**c**	C	A	G	—						
NLD37	Dog/Netherlands	356	AY827499[15]	**t**	**c**	—	—	—	—						
D1,4,6,7,16,18,21,23,24	Dog/Japan	177	AB199737[26]	—	**c**	—	—	—	—						
Consensus sequence D				**Y**	**Y**	**Y**	**R**	G	**Y**						
Assemblage E				**258**	**405**	**546**	**582**	**651**	**705**	**723**	**896**				
—	Hoofed livestock/Mixed	608	U47632[21]	**g**	G	C	**R**	G	**t**	**K**	—				
P-15	Pig/Czech Republic	1114	AY178741[21]	A	G	C	**g**	**a**	C	**g**	**A**				
Ad-133	Calf/Australia	1115	AY178740[21]	A	G	**Y**	A	G	C	T	G				
GC-155	Calf/Japan	592	AB182127[27]	A	G	**t**	A	G	C	T	—				
NLG409	Goat/Netherlands	531	AY826198[15]	A	G	C	A	G	—	—	—				
NLS387	Sheep/Netherlands	430	AY826200[15]	A	G	C	—	—	—	—	—				
NLS352	Sheep/Netherlands	428	AY826199[15]	A	G	C	—	—	—	—	—				
Cf1,3	Calf/Japan	177	AB199740[26]	—	G	—	—	—	—	—	—				
Cf4,5	Calf/Japan	177	AB199741[26]	—	**a**	—	—	—	—	—	—				
Consensus sequence E				A	G	C	**R**	G	C	**K**	**R**				
Assemblage G				**201**	**447**	**621**	**762**	**1000**	**1047**	**1056**	**1068**	**1107**	**1269**		
Ad-167	Rat/Australia	1117	AY178746[24]	**c**	**t**	C	**a**	A	**C**	**c**	**t**	**t**	**t**		
Ad-155	Rat/Australia	1117	AY178745[24]	T	C	**t**	G	A	**C**	T	C	C	C		
Ad-157	Rat/Australia	1085	AF069058[7]	—	C	C	G	A	**Y**	T	C	C	C		
Ad-171	Rat/Australia	1114	AY178747[24]	**N**	C	C	G	**g**	T	T	C	C	C		
Ad-170	Mouse/Australia	1111	AY178748[24]	T	C	C	G	**g**	T	T	C	C	C		
Consensus sequence G				**Y**	C	C	G	**R**	**Y**	T	C	C	C		

Table 2B. Glutamate dehydrogenase isolates, position and breakdown of intra-genotypic substitutions – Assemblage B

Isolate	Source/Origin	Size (bp)	Accession no.	Nucleotide position from start of gene																														
				219	**297**	**309**	**357**	**360**	**429**	**447**	**519**	**540**	**561**	**570**	**576**	**582**	**597**	**612**	**690**	**699**	**705**	**723**	**756**	**786**	**807**	**825**	**921**	**969**	**1077**	**1143**	**1251**	**1254**
BIII																																		
gd-ber7	Human/Norway	425	DQ090538[28]	—	**t**	**C**	**t**	**a**	**T**	**T**	C	**C**	**C**	C	G	**a**	C	**G**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gd-ber6	Human/Norway	425	DQ090537[28]	—	**t**	**C**	C	**a**	**T**	**T**	C	**C**	**C**	C	G	**a**	C	**G**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gd-ber5	Human/Norway	427	DQ090536[28]	—	**t**	**C**	**t**	G	**T**	**T**	C	**C**	**C**	C	G	**a**	C	**G**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gd-ber4	Human/Norway	428	DQ090535[28]	—	**t**	**C**	C	**a**	**T**	**T**	C	**C**	**C**	C	G	G	C	**G**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
BAH-12	**Human/Australia**	**592**	**AF069059[7]**	**T**	C	**C**	**t**	G	**T**	**T**	C	**C**	**C**	C	G	G	C	**G**	G	**t**	T	**C**	T	C	—	—	—	—	—	—	—	—
GH-135	Human/Japan	592	AB195224[8]	**T**	C	**C**	**t**	G	**T**	**T**	**t**	**C**	**C**	C	G	G	C	**G**	G	C	T	**C**	T	C	—	—	—	—	—	—	—	—
FCQ21	Human/Mexico	1108	AY178756[24]	**T**	C	**Y**	**y**	G	**T**	**T**	**y**	**C**	**C**	C	G	G	C	**G**	G	C	T	**C**	T	C	**g**	**a**	**g**	**t**	**t**	**t**	**c**	**y**
BIII-like																																		
gd-ber1	Human/Norway	423	DQ090532[28]	—	C	**C**	**t**	G	**T**	**T**	C	T	**C**	C	G	G	C	**G**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gd-ber9	Human/Norway	424	DQ090540[28]	—	**t**	**C**	C	G	**T**	**T**	**t**	**C**	**C**	C	G	G	**t**	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gd-ber10	Human/Norway	424	DQ090541[28]	—	**t**	**C**	C	G	**T**	C	**t**	**C**	**C**	C	G	G	**t**	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—
gd-ber2	Human/Norway	427	DQ090533[28]	—	C	T	C	G	**T**	**T**	C	T	**C**	C	G	G	C	**G**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
B-central																																		
NLH25	Human/Netherlands	773	AY826193[15]	**T**	**t**	**C**	**t**	G	C	C	C	T	**C**	C	G	G	C	**G**	G	C	T	**T**	T	C	T	G	—	—	—	—	—	—
GH-158	Human/Japan	592	AB188825[8]	**T**	C	**C**	C	G	C	C	C	T	**C**	**t**	G	G	C	A	G	**t**	T	**C**	T	C	—	—	—	—	—	—	—	—
gd-ber3	Human/Norway	426	DQ090534[28]	—	C	T	C	G	C	C	C	**C**	**C**	C	G	G	C	**G**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
BIV-like																																		
gd-ber8	Human/Norway	424	DQ090539[28]	—	C	T	C	G	C	C	C	T	**C**	C	G	G	C	**G**	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
NLH28	Human/Netherlands	760	AY826192[15]	C	C	T	C	G	C	C	C	T	**C**	C	G	G	C	**G**	G	C	T	**T**	T	C	T	G	—	—	—	—	—	—	—
Ad-85	Human/Australia	1127	AY178755[24]	C	C	T	C	G	C	C	C	T	T	C	**a**	G	C	A	G	C	T	**C**	T	C	**g**	**a**	**g**	**t**	**t**	**t**	**c**	**t**
Ad-158	Ape (marmoset)/ Australia	1115	AY178753[24]	C	C	T	C	G	C	C	C	T	**C**	C	G	G	**t**	A	G	C	T	**T**	**c**	C	—	—	—	—	—	—	—	—
GH-156	Human/Japan	592	AB182126[27]	C	C	T	C	G	C	C	C	T	**C**	**t**	G	G	**t**	A	**a**	C	**c**	**T**	**c**	C	—	—	—	—	—	—	—	—
NLH35	Human/Netherlands	740	AY826197[15]	C	C	T	C	G	C	C	C	T	**C**	**t**	G	G	**t**	A	**a**	C	**c**	**T**	**c**	C	—	—	—	—	—	—	—	—
Ad-156	Ape (marmoset)/ Australia	1121	AY178752[24]	C	C	T	C	G	C	C	C	T	**C**	**t**	G	G	C	A	G	C	T	**T**	T	C	T	G	A	C	C	C	T	C
Ad-82	Human/Australia	1121	AY178754[24]	C	C	T	**y**	G	C	C	C	T	Y	C	**r**	G	C	A	G	C	T	**Y**	T	C	T	G	**g**	C	C	C	**y**	C
BIV																																		
CZ:D47	Dog/Czech Rep.	1119	AY178749[24]	C	C	T	C	G	C	C	C	T	T	C	G	G	C	A	G	C	T	**T**	T	**t**	T	G	A	C	C	C	T	C
Vanc/89/UBC /059 (WOOF)	Dog/Canada	1121	AY178750[24]	C	C	T	C	G	C	C	C	T	T	C	G	G	C	A	G	C	T	**T**	T	**t**	T	G	A	C	C	C	T	C
CH-105	Chinchilla/Czech Rep.	1121	AY178751[24]	C	C	T	C	G	C	C	C	T	T	C	G	G	C	A	G	C	T	**T**	T	**t**	T	G	A	C	C	C	T	C
Ad-45	**Human/Australia**	**1110**	**AY178739[24]**	C	C	T	**t**	G	C	C	C	T	T	C	G	G	C	A	G	C	T	**T**	T	C	T	G	A	C	C	C	T	**t**
Ad-28	**Human/Australia**	**1123**	**AY178738[24]**	C	C	T	**t**	G	C	C	C	T	T	C	G	G	C	A	G	C	T	**T**	T	C	T	G	A	C	C	C	T	C
NLH13	Human/Netherlands	759	AY826191[15]	C	C	T	**t**	G	C	C	C	T	T	C	G	G	C	A	G	C	T	**T**	T	C	T	G	—	—	—	—	—	—
Ad-7	**Human/Australia**	**690**	**L40508[23]**	C	C	T	**t**	G	C	C	C	T	T	C	G	G	C	A	G	C	T	**T**	T	C	T	G	A	C	C	C	T	C
M1,2,3	Monkey/Japan	177	AB199738[26]	—	—	T	**t**	G	C	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Consensus sequence B				**Y**	C	**Y**	**Y**	G	**Y**	**Y**	C	**Y**	**Y**	C	G	G	C	**R**	G	C	T	**Y**	T	C	T	G	**R**	C	C	C	T	C

sample contributions). The disadvantage of this method lies in the potential over-simplification of the sequences.

To determine each consensus sequence, all of the variable sites were analysed to establish either the major nucleotide represented or the appropriate degenerate base (in universal code) to represent the composition of the original sequences. The details of the criteria used to determine these consensus sequences are given below and the breakdown of each of these intra-genotypic (intra-assemblage) substitution sites for *gdh*, *tpi* and *β giardin* are given in Tables 2 to 4. Since there were so few intra-genotypic substitutions at the *SSU rDNA* locus they were not tabulated and instead were included in the consensus sequence figure. All intra-genotypic substitutions for the *SSU rDNA* were noted for the small data sets (those with few available sequences), whereas those from the larger sets were only noted if there was more than 1 sample with the same substitution. The completed consensus sequences are given in Figs 1 and 2.

Once the variable sites were identified, the more significant substitutions were tabulated. For assemblages with numerous sequences available (A and B in *gdh* and *tpi* and A, B and E in *β giardin*) only intra-genotypic substitutions with a representation greater than one were tabulated; for all other assemblages, all of the intra-genotypic substitutions were tabulated. This distinction was made because in the larger data sets isolated substitutions were deemed less significant, whereas in the smaller sample sets their potential significance was unknown. Variable sites at the extreme ends of a sequence (within 20 bp) were not included owing to known potential problems in the sequencing and interpretation of these regions and hence the reduced certainty that they were valid substitutions. The consensus nucleotides – per variable site – were subsequently determined by the majority. Degenerate bases were used at either known variable sites for the subassemblages (AI/AII and BIII/BIV) or sites with greater than 25% deviation amongst the samples (within an assemblage). Degenerate bases in the original sequences were not included unless there were no other sequences available for that assemblage in that region.

Substitution analyses

The aligned consensus sequences were then examined at the inter-genotypic level for substitutions between the assemblages. Similarities and differences between the sequences determine their grouping into assemblages and subassemblages and the relatedness inferred in phylogenetic analyses. Analyses of the substitution patterns of a gene and comparison between the genes, provides information about the suitability of the gene or gene regions for use in different genotyping applications.

As all of the loci being analysed had at least 1 complete sequence available on the database from the *Giardia* genome project (McArthur *et al*. 2000) and most *Giardia* genes do not contain introns, it was possible to determine the amino acid codon frame of each of the consensus sequence alignments from the start codon of that gene. Sequences were therefore aligned into their (amino acid) codon frame and substitutions at a particular nucleotide position were noted as those departing from the majority of other assemblages. Total rates of substitution were noted (substitutions per nucleotide) as well as the types of substitutions – expressed (non-synonymous), silent (synonymous) and unique (those not shared by another assemblage – unique to that assemblage).

Phylogenetic trees

In an effort to learn more about the relationships amongst the assemblages of *G. duodenalis*, phylogenetic trees were constructed using the current consensus sequences. Although trends in relationships can be seen in aligned sequences, evaluating the sum of these patterns maybe difficult, whereas phylogenetic trees can clearly demonstrate the likely associations and interactions. Sequences were aligned with CLUSTAL X 1.81 and viewed with Treeview 68K (Page, 1996). Sequences were then analysed further in MEGA 3.1 (Kumar *et al*. 2004). Pairwise distances and Neighbour-Joining phylogenies (1000 estimations) for the nucleotide sequences were calculated using each of the available models (p-distance, Jukes-Cantor, Kimura 2-parameter, Tajima-Nei, Tamura 3-parameter and Tamura-Nei). Nucleotide frequencies were calculated and transition/transversion ratios examined. Among-site variation was also investigated using the different default options for gamma distribution parameters in MEGA (0·25, 0·5, 1·0 and 2·0) as well as comparison of pairwise distances and phylogenies (using multiple distance estimation models) of the different codon positions (1st, 2nd and 3rd). Amino acid sequence phylogenies were also examined with multiple models [p-distance, Poisson-correction, equal input model, Dayhoff (PAM matrix) and Jones-Taylor-Thornton (JTT matrix)]. Trees were constructed from the nucleotide and amino acid aligned consensus sequences for individual and concatenated genes.

The amino acid alignments were included to investigate the effects of the non-synonymous substitutions alone. Since the non-synonymous substitutions have greater selection pressure (occurring less often within a gene than the synonymous ones) the shared non-synonymous substitutions are less likely to occur randomly (2 assemblages developing the same non-synonymous substitution independently) and are most likely to be a result of their

Table 3. Triose phosphate isomerase isolates, position and breakdown of intra-genotypic substitutions

	Isolate	Source/Origin	Size (bp)	Accession no.	Nucleotide position from start of gene										
Assemblage A					**129**	**399**	**567**	**675**							
AI															
	Ad-1	**Human/Australia**	**466**	**AF069556[7]**	**T**	**C**	—	—							
	WB	**Human/Afghanistan**	**1583**	**L02120[29]**	**T**	**C**	C	T							
	—	White tail deer/USA	508	AY302562[30]	**T**	**C**	—	—							
	—	Cattle/USA	512	AY655704[13]	**T**	**C**	—	—							
AII															
	JH	**Human/USA**	**1112**	**U57897[31]**	C	T	A	C							
	Ad-2	**Human/Australia**	**479**	**AF069557[7]**	C	T	—	—							
	2907	Human/Peru	467	AY228647[32]	C	T	—	—							
	1503	Water/USA	532	AY368157[33]	C	T	—	—							
	3906	Water/USA	532	AY368158[33]	C	T	—	—							
	4220	Water/USA	532	AY368159[33]	C	T	—	—							
	4218	Water/USA	532	AY368160[33]	C	T	—	—							
	4230	Water/USA	532	AY368161[33]	C	T	—	—							
	Consensus sequence A				**Y**	**Y**	*	*							
Assemblage B					**39**	**45**	**91**	**162**	**165**	**168**	**210**	**216**	**297**	**429**	**483**
BIII															
	BAH-12	**Human/Australia**	**456**	**AF069561[7]**	—	—	**C**	G	**C**	**C**	**G**	C	A	G	A
	2887	Human/Peru	468	AY228631[32]	—	—	**C**	G	**C**	**C**	**G**	C	A	G	A
	2434	Water/USA	533	AY368165[33]	**G**	T	**C**	G	**C**	**C**	**G**	C	A	G	A
	1794	Water/USA	532	AY368164[33]	**G**	T	**C**	G	**C**	**C**	**G**	C	A	G	A
	2924	Human/Peru	532	AY228628[32]	**G**	T	**C**	G	**C**	**C**	**G**	C	A	G	A
	2582	Human/India	449	AY228629[32]	—	—	**C**	**a**	**C**	**C**	**G**	C	A	G	A
	2506	Human/Peru	468	AY228630[32]	—	—	**C**	**a**	**C**	**C**	**G**	C	A	G	A
	2436	Water/USA	532	AY368163[33]	**G**	**c**	**C**	G	**C**	**C**	**G**	**t**	A	G	**g**
BIII-like															
	3920	Water/USA	532	AY368166[33]	**G**	T	**C**	G	T	**C**	**G**	C	A	G	**g**
	2877	Human/Peru	468	AY228633[32]	—	—	T	**a**	**C**	**C**	**G**	C	A	G	A
	2902	Human/Peru	469	AY228632[32]	—	—	T	G	**C**	**C**	**G**	C	**g**	G	A
	2623	Water/USA	532	AY368162[33]	**G**	**c**	T	G	**C**	T	**G**	C	**g**	G	A

BIV-like															
	2901	Human/Peru	468	AY228635[32]	—	—	T	G	T	T	**G**	C	A	G	A
	2900	Human/Peru	468	AY228634[32]	—	—	T	G	T	T	**G**	C	A	G	A
	7327	Water/USA	532	AY368168[33]	A	T	T	G	T	T	**G**	C	A	G	A
	7115	Water/USA	532	AY368167[33]	A	T	T	G	T	T	**G**	C	A	G	A
BIV															
	3565	Muskrat/USA	469	AY228637[32]	—	—	T	G	T	T	A	**t**	A	G	A
	3470	Muskrat/USA	468	AY228636[32]	—	—	T	G	T	T	A	C	A	G	A
	3577	Muskrat/USA	448	AY228638[32]	—	—	T	G	T	T	A	C	A	G	A
	1758	Rabbit/China	468	AY228639[32]	—	—	T	G	T	T	A	C	A	**a**	A
	5409	Water/USA	532	AY368171[33]	A	T	T	G	T	T	A	C	A	**a**	A
	Ad-19	**Human/Australia**	**479**	**AF069560[7]**	A	T	T	G	T	T	A	C	A	**a**	A
	GS/M	Human/USA	1701	L02116[34]	A	T	T	G	T	T	A	C	A	G	A
	2476	Water/USA	532	AY368169[33]	A	T	T	G	T	T	A	C	A	G	A
	2100	Water/USA	532	AY368170[33]	A	T	T	G	T	T	A	C	A	G	A
	Consensus sequence B				*	*	**Y**	G	**Y**	**Y**	**R**	C	A	G	A
Assemblage C					**150**	**330**	**383**	**393**							
	2665	Dog/USA	468	AY228644[32]	**t**	C	**t**	**c**							
	2674	Dog/USA	468	AY228643[32]	G	C	**t**	**c**							
	2669	Dog/USA	468	AY228642[32]	G	C	C	A							
	2643	Dog/USA	532	AY228641[32]	G	**t**	C	A							
	Ad-136	Dog/Australia	479	AF069563[7]	G	**t**	C	A							
	Consensus sequence C				G	**Y**	**Y**	**M**							
Assemblage D					**333**										
	GM Dog 19	Dog/Germany	530	DQ220289[35]	**T**										
	GM Dog 60	Dog/Germany	530	DQ246216[35]	C										
	Consensus sequence D				**Y**										
Assemblage E					**72**	**93**	**109**	**326**	**362**	**363**	**471**	**489**			
	P-15	Pig/Czech Republic	479	AF069559[7]	**c**	C	**A**	A	**g**	**g**	**g**	**g**			
	15	Cattle/USA	457	AY228646[32]	—	C	**A**	A	A	T	A	A			
	—	Cattle/USA	512	AY655706[13]	T	C	**A**	A	A	T	A	A			
	Guangzhou calf	Dairy calf/China	688	DQ157270[4]	T	C	G	**g**	A	T	A	A			
	109	Cattle/USA	457	AY228645[32]	—	**t**	G	A	A	T	A	A			
	—	Cattle/USA	512	AY655705[13]	T	**t**	G	A	A	T	A	A			
	Consensus sequence E				T	**Y**	**R**	A	A	T	A	A			

Table 4. Beta giardin isolates, position and breakdown of intra-genotypic substitutions

Isolate	Source/Origin	Size (bp)	Accession no.	Nucleotide position from start of gene							
Assemblage A				**450**	**460**	**468**	**606**	**729**			
AI											
Portland 1	**Human/UK**	**1622**	**X85958**[36]	C	C	T	**C**	**A**			
WB C6	**Human/Afghanistan**	**819**	**XM763377**[19]	C	C	T	**C**	**A**			
1–4, 6, 9 & 10C	Dog/Brazil	330	DQ466724-30[37]	C	C	T	**C**	**A**			
2G	Cat/Brazil	330	DQ466731[37]	C	C	T	**C**	—			
1H-53H,55H-56H,58H-62H	Human/Brazil	288–323	DQ466732-84, 86-87,89-93[37]	—	—	—	**C**	**A**			
GD-99H	Dog/Japan	472	AB218605[10]	**t**	C	T	**C**	G			
GF-1	Ferret/Japan	472	AB159797[8]	**t**	C	T	**C**	G			
—	Cattle/USA	659	AY655702[13]	**t**	C	T	**C**	G			
GD37	Human/Italy	511	AY545644[38]	C	C	T	**C**	—			
A14	Dog/Italy	511	AY545649[38]	C	C	T	**C**	—			
A44	Calf/Italy	511	AY545642[38]	C	C	T	**C**	—			
—	White tail deer/USA	500	AY302561[30]	C	C	**c**	**C**	—			
A											
GD83	Human/Italy	465	AY545643[38]	C	C	T	—	—			
AII											
ISSGF7		753	AY072724[39]	C	**t**	**c**	**T**	G			
STS-U		722	DQ090542[28]	C	**t**	**c**	**T**	G			
CBHRG9	Water/Mexico	677	DQ116612[40]	C	**t**	**c**	**T**	G			
CBHRG6	Water/Mexico	684	DQ116609[40]	C	C	T	**T**	G			
CBHRG7	Water/Mexico	646	DQ116610[40]	C	C	T	**T**	G			
CBHRG16	Water/Mexico	699	DQ116617[40]	C	C	T	**T**	G			
CBHRG18	Water/Mexico	687	DQ116619[40]	C	C	T	**T**	G			
CBHRG8	Water/Mexico	552	DQ116611[40]	C	C	T	**T**	—			
CBHRG17	Water/Mexico	639	DQ116618[40]	C	C	T	**T**	—			
GD115	Human/Italy	511	AY545645[38]	C	C	T	**T**	—			
54H	Human/Brazil	311	DQ466785[37]	—	—	—	**T**	G			
57H	Human/Brazil	288	DQ466788[37]	—	—	—	**T**	G			
KC8	**Human/Israel**	753	**AY072723**[41]	C	C	T	**T**	G			
Consensus sequence A				C	C	T	**Y**	*			
Assemblage D				**172**	**204**	**210**	**246**	**247**	**252**	**327**	**615**
—	Dog/USA	558	AY370531[43]	**g**	**W**	**g**	A	**g**	C	**c**	**g**
A27	Dog/Italy	511	AY545648[38]	A	**g**	T	**g**	A	C	A	A
—	Coyote/USA	433	AY370530[43]	—	A	T	A	A	**t**	A	—
A21	Dog/Italy	753	AY545647[38]	A	A	T	A	A	C	A	A
GD-142	Dog/Japan	472	AB218604[10]	—	—	—	—	—	—	A	A
Consensus sequence D				**R**	A	T	A	A	C	A	A

Assemblage B				**105**	**210**	**228**	**354**	**369**	**378**	**438**	**564**	**648**
BAH8	Human/Australia	753	AY072727[41]	C	C	A	**C**	C	C	C	T	G
LD18	Human/Belgium	753	AY072726[41]	C	C	A	**C**	C	C	C	T	G
BG-Ber3	Human/Norway	667	DQ090524[28]	**t**	C	A	**C**	C	C	C	T	G
BG-Ber4	Human/Norway	676	DQ090525[28]	**t**	C	A	**C**	C	**t**	C	T	G
BG-Ber9	Human/Norway	711	DQ090530[28]	**t**	C	A	**C**	C	**t**	C	T	G
BG-Ber7	Human/Norway	679	DQ090528[28]	**t**	C	A	**C**	C	C	C	T	**a**
BG-Ber8	Human/Norway	676	DQ090529[28]	**t**	C	A	**C**	C	C	C	**c**	**a**
BG-Ber5	Human/Norway	725	DQ090526[28]	C	C	A	**C**	C	C	C	**c**	**a**
BG-Ber10	Human/Norway	717	DQ090531[28]	C	**t**	A	**C**	C	C	C	T	**a**
Nij5	Human/Netherlands	753	AY072725[41]	C	**t**	A	T	C	C	**t**	T	**a**
GH-156	Human/Japan	472	AB182124[27]	—	—	—	T	**t**	C	**t**	T	G
H3	Human	693	DQ116605[38]	C	**t**	A	T	**t**	C	C	T	G
A88	Calf/Italy	511	AY647266[38]	—	**t**	A	**C**	C	C	C	T	G
ISSGF4		688	AY072728[39]	C	C	**g**	T	C	C	C	**c**	G
BG-Ber1	Human/Norway	721	DQ090522[28]	C	C	**g**	**C**	C	C	C	T	G
BG-Ber6	Human/Norway	721	DQ090527[28]	C	C	**g**	T	C	C	C	T	G
BG-Ber2	Human/Norway	720	DQ090523[28]	C	C	A	T	C	C	C	T	G
A82	Calf/Italy	511	AY647265[38]	—	C	A	T	C	C	C	T	G
GH-158	Human/Japan	472	AB188826[42]	—	—	—	T	C	C	C	T	G
Consensus sequence B				*	C	A	**Y**	C	C	C	T	**R**

Assemblage E				**201**	**306**	**408**	**549**	**660**	**684**	**690**	**714**
CBHRG5	Sheep/Mexico	699	DQ116608[40]	**C**	A	C	T	C	G	T	T
CBHRG21	Sheep/Mexico	701	DQ116621[40]	**C**	A	C	Ty	C	G	T	T
CBHRG1	Sheep/Mexico	677	DQ116604[40]	**C**	A	C	T	C	**a**	T	T
CBHRG3	Sheep/Mexico	671	DQ116606[40]	**C**	A	C	T	C	**a**	T	T
CBHRG4	Sheep/Mexico	702	DQ116607[40]	**C**	A	C	T	C	**a**	T	T
CBHRG11	Sheep/Mexico	606	DQ116614[40]	**C**	A	C	T	C	**a**	T	—
CBHRG19	Sheep/Mexico	710	DQ116620[40]	**C**	A	C	T	C	**a**	T	T
CBHRG25	Sheep/Mexico	680	DQ116625[40]	**C**	A	C	T	C	**a**	T	T
CBHRG24	Sheep/Mexico	711	DQ116624[40]	T	A	C	T	C	**a**	T	T
A46	Cattle/Italy	458	AY545650[38]	**C**	**g**	C	**c**	—	—	—	—
CBHRG10	Sheep/Mexico	661	DQ116613[40]	T	**g**	C	**c**	**t**	G	**c**	T
CBHRG13	Sheep/Mexico	658	DQ116616[40]	T	**g**	C	**c**	**t**	G	**c**	T
CBHRG22	Sheep/Mexico	718	DQ116622[40]	T	**g**	C	**c**	**t**	G	**c**	T
CBHRG23	Sheep/Mexico	718	DQ116623[40]	T	**g**	C	**c**	**t**	G	**c**	T
A98	Calf/Italy	511	AY653159[38]	T	A	**t**	T	—	—	—	—
GC-155	Cattle/Japan	472	AB182125[27]	—	A	**t**	T	C	G	T	**c**
—	Cattle/USA	629	AY655703[13]	T	A	**t**	T	C	G	T	**c**
P-15	Pig/Czech Republic	753	AY072729[41]	T	A	C	T	C	G	T	**c**
CBHRG12	Sheep/Mexico	666	DQ116615[40]	T	A	C	T	C	G	T	T
Consensus sequence E				**Y**	**R**	C	**Y**	*	*	*	*

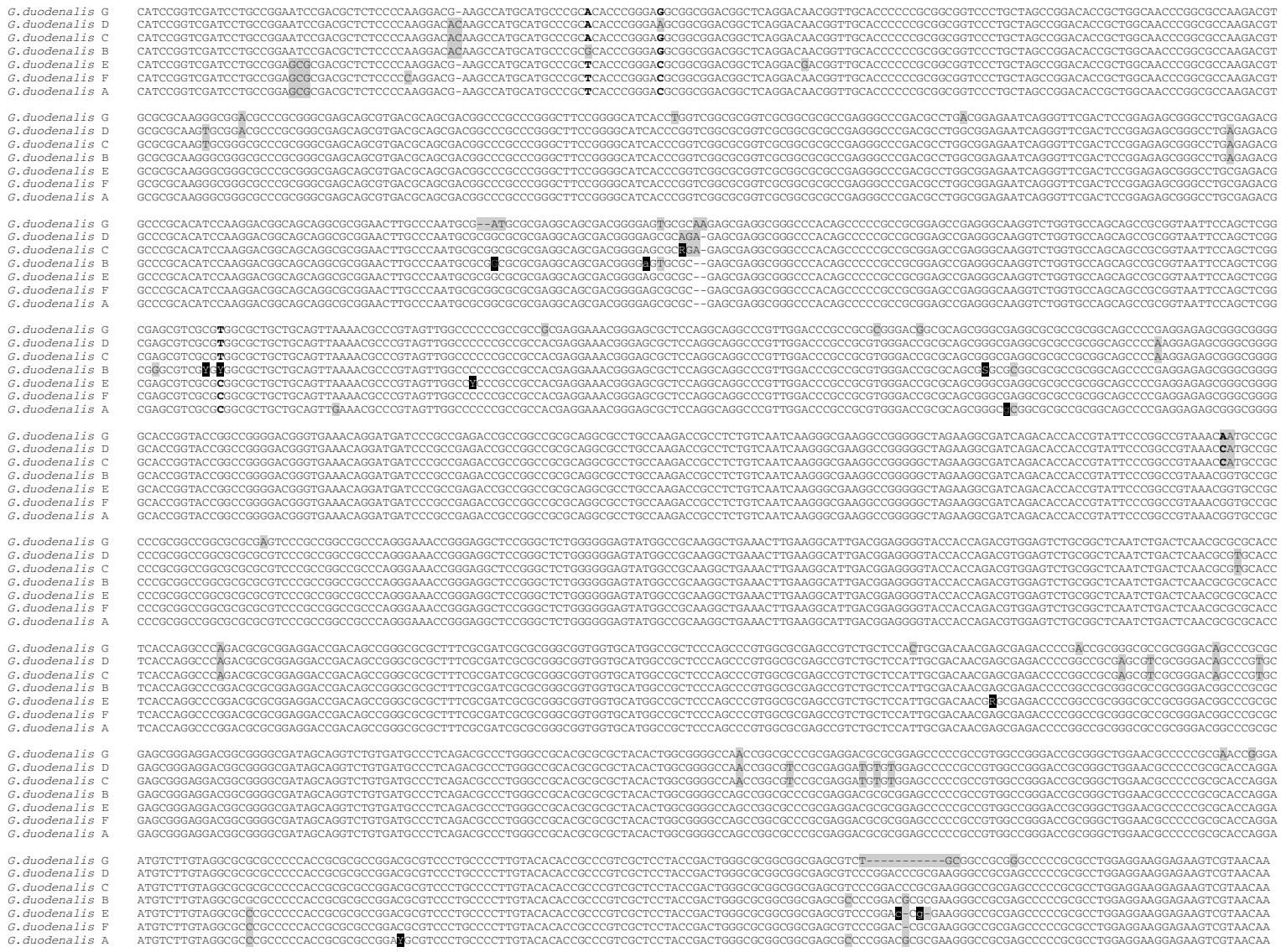

Fig. 1. Full-length *SSU rDNA* consensus sequence alignment (1407 to 1420 base pairs). Substitutions (from the majority) are shaded grey. Large variations within an assemblage (greater than 25% or subgrouping of AI/AII) are shown with a degenerate base shaded black with white text. Small variations within an assemblage are shown with the predominant base in lower case shaded black with white text. Two bases shaded black with grey text represent positions where half (assemblage B) or a quarter (assemblage A) of the isolates had a deletion at that site. Bold nucleotides indicate positions with increased variation.

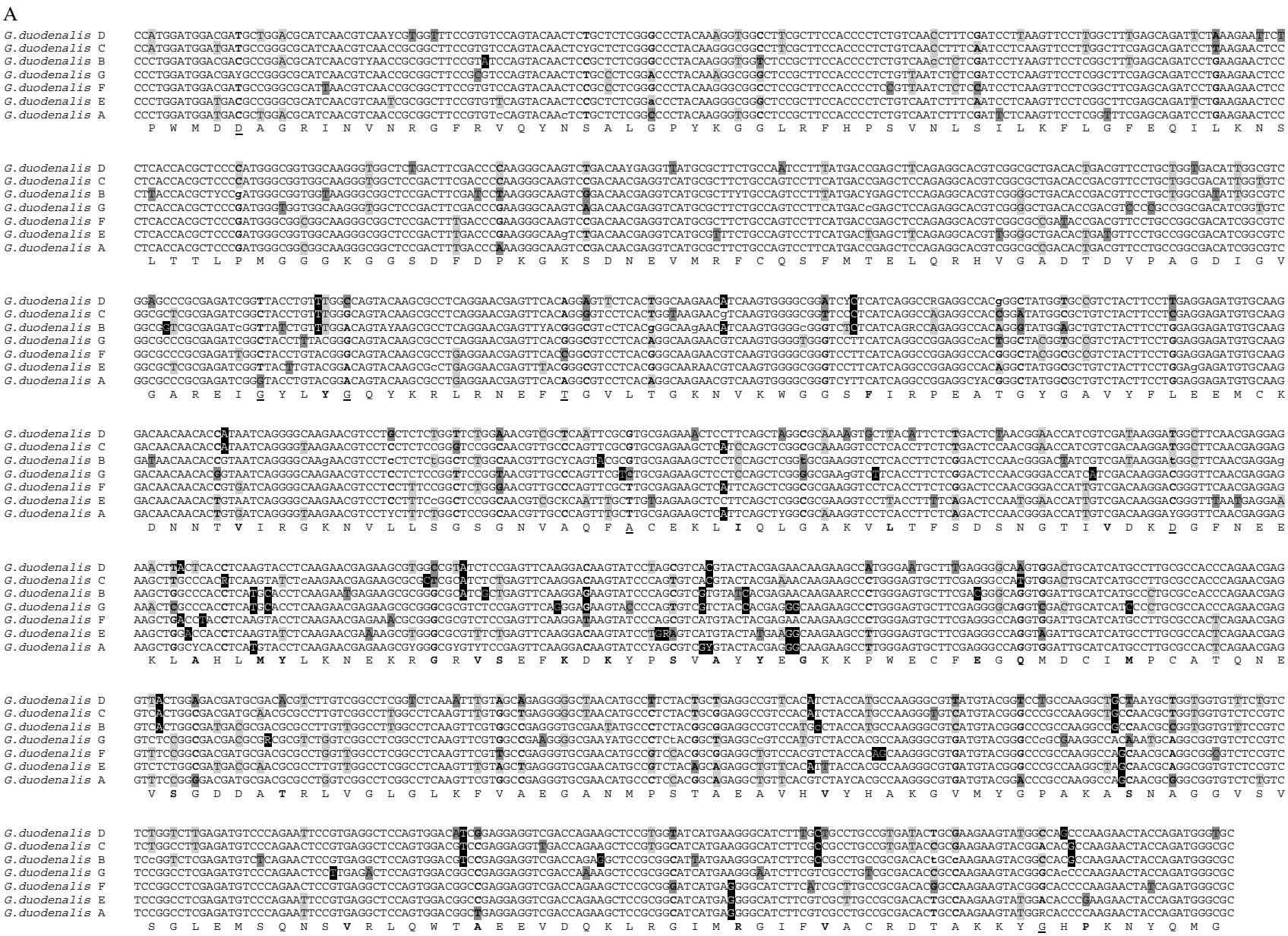

Fig. 2. For legend see p. 1811.

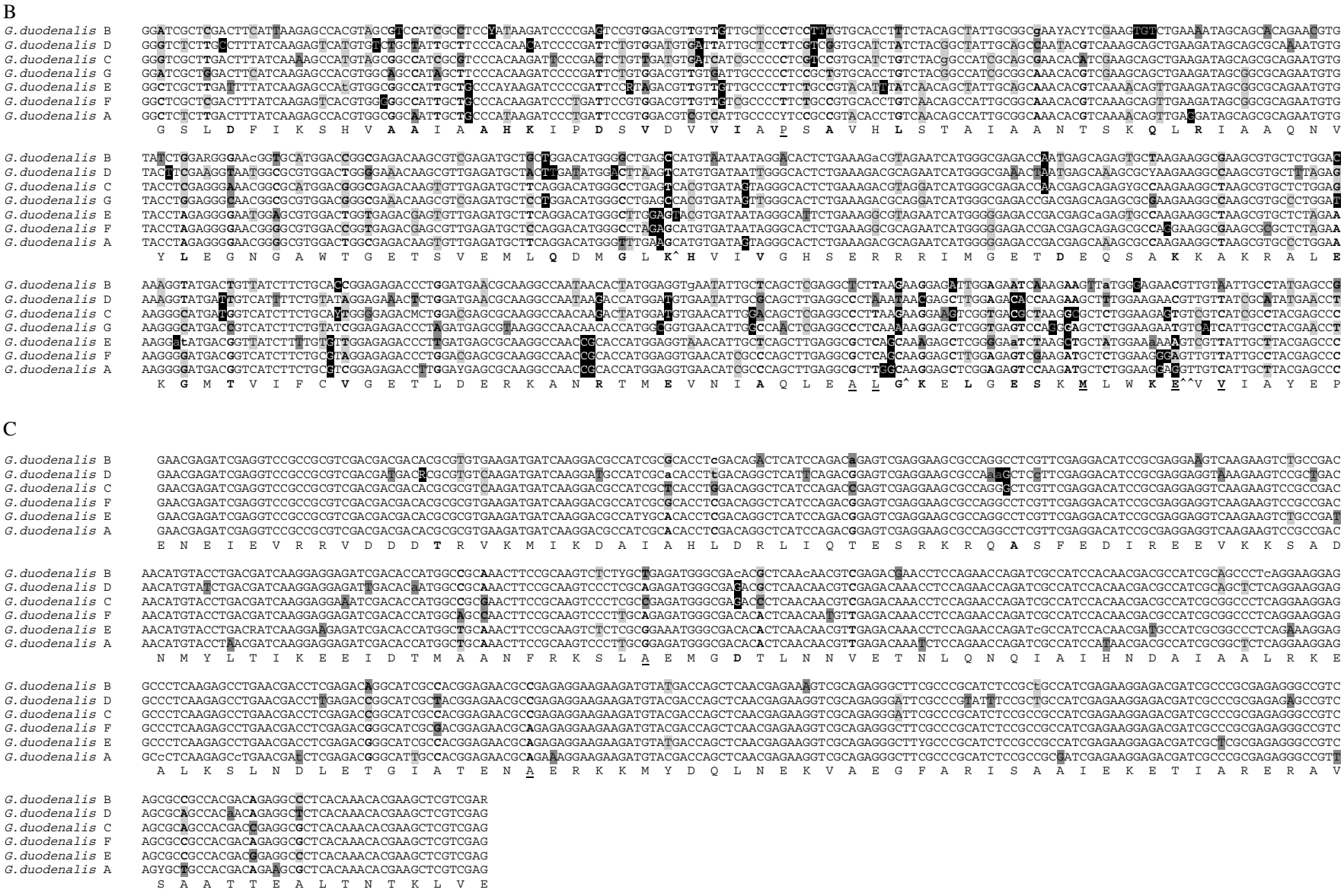

Fig. 2. For legend see p. 1811.

shared evolution (developed before their divergence). Non-synonymous substitutions are also more likely to be more linear (changing one-way) over time, whereas the synonymous substitutions are more likely to revert, also potentially confusing the evaluation of the shared substitutions.

Due to the giardin proteins being *Giardia* specific, an out group had to be sourced from within the *Giardia* family, and at the time of collating the data, *G. muris* was the only available sequence for *β giardin*; however, this sequence was not available for the *gdh* locus. As there was not a common out-group available for all of the loci, the trees were left un-rooted to facilitate comparison with the concatenated sequences. Two combinations of concatenated nucleotide sequences were constructed, one without assemblage G to facilitate the *β giardin* sequences and one without *β giardin* to include the assemblage G sequences. For the concatenated amino acid sequences, only *gdh* and *tpi* were used as *β giardin* had limited expressed variation and this allowed the inclusion of assemblage G.

RESULTS AND DISCUSSION

GenBank survey

In the initial process of collecting sequences from GenBank the most striking fact was how little sequence data was available for the *SSU rDNA* locus of *Giardia*. As this is typically the original gene of choice for identification and phylogenetic analysis, it was surprising that there were only 12 full-length *G. duodenalis* sequences for the *SSU rDNA* on GenBank and most of these were several years old (Table 1). There were an additional 9 sequences greater than 350 base pairs (out of 1400 bp) totalling only a third of the gene. It was therefore interesting to note that the original sequence results for most of the genotypes have not been repeated and reproduced. The remaining sequences were genotyping sequences covering the first fifth or the last tenth of the gene, using primers from Weiss *et al.* (1992), van Keulen *et al.* (1995) or Hopkins *et al.* (1997). However, the products produced by these primer sets are too small to genotype all of the current assemblages. Notably assemblage F (of cats) is identical to assemblage A (zoonotic) until base 499 (Fig. 1). This has led to some inconclusive genotyping and potentially inaccurate results being reported. In the study by Berrilli *et al.* (2004), the cat isolate was genotyped as assemblage A using the 292 bp-product RH primers of Hopkins *et al.* (1997) and in a study by Fayer *et al.* (2006), the cat isolates were genotyped as assemblage F using the same RH primers. This discrepancy may be due to an unsubstantiated substitution at the 5′ end of the AF199444 (a C at position 38). As there were only 2 assemblage F sequences available on GenBank (that had been verified as such at another locus) and one of them started near this position, it was difficult to determine if this was an artefact (error due to signal noise) or an actual substitution. At the time these data were being compiled, the sequences from Fayer *et al.* (2006) had not yet been deposited on GenBank for comparison. Ironically the primers described by van Keulen *et al.* (2002) to produce an *SSU rDNA* product for restriction enzyme digestion would have been able to discriminate all of the assemblages by sequencing, but not by digestion.

The other loci were comparatively well represented over the alignment length examined. The majority of sequences available were from the zoonotic assemblages A and B (predominantly from humans/human waste water), followed by the domestically significant assemblages C and D (dogs) and E (hoofed livestock) and the main weakness in representation was found for assemblages F and G (notably assemblage G in *β giardin*). Increasing some of these data sets in the future would improve the accuracy and sensitivity of further analyses. The coverage of the alignments over the genes also varied.

Fig. 2. Nucleotide consensus sequences. (A) Glutamate dehydrogenase. (B) Triose phosphate isomerase (C) Beta giardin. (A) Glutamate dehydrogenase: 80% coverage of gene, 1106 base pairs (bp) in length of the 1350 bp total (including stop codon) from the first base of the 63rd amino acid codon (187th bp) to the second base of the 431st codon (1292nd bp). (B) Triose phosphate isomerase: 60% coverage of gene, 468 bp long of a total of 774 bp (including stop) from the first base of the 16th amino acid codon (46th bp) to the third base of the 171st codon (513th bp). (C) Beta giardin: 60% coverage of gene, 511 bp long of a total of 819 bp (including stop) from the third base of the 46th amino acid codon (138th bp) to the third base of the 216th codon (648th bp). Expressed nucleotide substitutions (from the majority) are shaded black with white text (unique expressed, black with grey text). Silent nucleotide substitutions (from the majority) are shaded grey (unique silent, dark grey). Large variations within an assemblage (greater than 25% or subgrouping of AI/AII or BIII/BIV) are shown with a degenerate base. Lower case text is used to show the predominant nucleotide in positions with small variations. Bold nucleotides indicate positions with increased variation. Bold amino acids indicate sites of expressed variation. Positions of even majority of nucleotides are noted with underlined amino acids. Amino acid sequence given beneath the nucleotide sequence is from assemblage A. ^Codons with near even majorities of expressed substitutions (*tpi*). ^^Codon with too much variation to decipher substitutions (*tpi*).

As mentioned, the *SSU rDNA* locus had several full-length sequences available and the *gdh* gene was also well represented with 80% coverage of the gene by the alignment. One quarter of the samples covered the whole alignment (with all assemblages represented) and the remaining samples covered the first 40% to 60% of the alignment. However, the *tpi* and *β giardin* alignments were shortened to 60% of the length of the genes to maximize the number of assemblages included in the analyses. Therefore, the resulting alignments were well represented over their entire length by the sample sets. Both of these alignment lengths (468 and 511 bp respectively) were sufficient for genotyping samples to the assemblage and possibly subassemblage level; however, the phylogenetic analyses may have been affected by truncating these sequence alignments. The *tpi* and *β giardin* gene segments that were omitted had similar rates of substitutions overall as those that were included, however the *tpi* segment had a reduced rate of non-synonymous substitutions (between the sequences available – WB, JH and GS/M) and the *β giardin* segment had an extra shared substitution (between assemblages B and D). These differences were of note because changes in the substitution properties of an alignment can alter the amount of information gathered from it and, similarly, additional shared substitutions would provide further information to increase the support for relationships.

Alignments and subgroups

Aligning the sequences to establish the consensus sequences (Tables 2–4) was fairly straightforward until the subgroups of the zoonotic assemblages A and B were examined. Subgrouping within these zoonotic assemblages is of interest as there may be a link between specific subgroups and their hosts, with particular attention on those that may affect humans. It is therefore important to establish the specific subgroups, based on their intra-genotypic substitution patterns, and the continuity of these subgroups across the loci. The AI and AII subgroups originally described as groups 1 and 2 (Nash and Keister, 1985) were quite robust. In the *SSU rDNA* they were differentiated at the 3′ end by a single substitution (Weiss *et al*. 1992). At the *gdh* gene there were numerous AI/AII specific polymorphisms and the majority of isolates were easily characterized (Table 2A). There were only a few exceptions with single substitutions not matching the pattern. Since the *gdh* gene had so many intra-assemblage substitutions, it could be assumed that not all of these would specifically represent 'AI/AII-substitutions' and that some may represent other groups not yet identified. As the amount of sequence data increases it may be possible to determine which substitutions are more (or less) significant for differentiating the core subgroups and lesser subgroups. One sample was quite different however; the 'roe deer' has been previously noted for its divergence from either AI or AII (van der Giessen *et al*. 2006) and is potentially a new subgroup. The *tpi* gene had only 2 AI/AII specific polymorphisms in the region examined and there were no outliers (Table 3). It was interesting to note that this variable locus produced such a clean and neat distinction between these close subgroups. The conserved *β giardin* alignment in contrast had only 1 substitution clearly differentiating the AI/AII subgroups and then several others producing further groups, the significance of which were not yet apparent (Table 4). In fact the *β giardin* gene consistently demonstrated numerous subgroups within all assemblages represented by sufficient samples. Their significance, however, was difficult to determine because those samples that had been assessed at other loci did not always segregate into the same subgroups as they had with the *β giardin* locus (see below).

The BIII and BIV subgroups originally described by allozyme electrophoretic studies (Andrews *et al*. 1989) were not so reproducible. At the *SSU rDNA* there were insufficient full-length samples characterized previously as BIII or BIV to assess if there were in fact any BIII/BIV specific polymorphisms present. There were, however, several polymorphic sites within assemblage B that might warrant further investigation (Fig. 1). Establishing subgroups at the *SSU rDNA* locus would add weight to their distinction as the conserved loci typically only discern the older more significant groups. In the *β giardin* alignment there were also insufficient reference isolates previously characterized as BIII/BIV, to determine which substitutions may be BIII/BIV specific (Table 4). Attempts to use isolates characterized at other loci (and *β giardin*) as references lead to conflicting results, where samples grouped differently at the different loci, as found in a previous study (Robertson *et al*. 2006). This may be because the impromptu 'reference' isolates were not 'true' BIII or BIV representatives and therefore grouped differently again with another gene or perhaps the subgroups within this assemblage have not been comprehensively established. It was also tempting to suspect that the *β giardin* gene was incapable of subgrouping consistently (relative to the other loci) because of the diversity also shown within assemblage A. However, as the *β giardin* gene is conserved, it should demonstrate all of the basal relationships well (the assemblages and the main subassemblages) and so perhaps there are still insufficient data. In the *gdh* and *tpi* alignments the BIII and BIV isolates appeared to form the most distant subgroups within assemblage B (Tables 2B and 3). The subgrouping of assemblage B was clouded with extra substitutions forming additional groupings to those demonstrated by the BIII/BIV reference isolates alone – as well as

less (and less complete) reference isolates available when compared to assemblage A. The *tpi* alignment had only a slightly higher ratio of BIII/BIV specific substitutions to other substitutions than did the *gdh* alignment. As there were many isolates that were difficult to characterize it implied there might be more subgroups involved than just BIII and BIV. It has been noted previously that the BIII/BIV classification originated from only single isolates and that they may only represent a preliminary view of the diversity within assemblage B (Mayrhofer *et al*. 1995). If this is the case, attempting to characterize all assemblage B isolates as either BIII or BIV may be problematic. In order to delineate the basal relationships within this diverse assemblage more work is required using conserved loci. As mentioned earlier, it would be interesting to learn how many subgroups could be found at the *SSU rDNA* locus, as those would be expected to be as robust as the AI/AII divergence. Isolates representing the new subgroups could then be established as reference isolates for subsequent analyses of the more complex substitution patterns of the variable loci. For comparative purposes, the *ef1α* sequences for the BIII and BIV subgroups show no divergence over the 50% of the gene covered and so there are the possibilities that either polymorphisms may be found in the other half of the *ef1α* gene or that the older assemblage B sequences at the *SSU rDNA* locus had errors and not polymorphisms. Overall it is expected that continuing increases in sample numbers and reference isolates for this assemblage (to allow for its apparently greater variability) should eventually establish subgroups with as much continuity and reproducibility across the loci as the AI/AII subgroups and the assemblages demonstrate.

It was also interesting to note, that with the increasing number of samples available, subgroups have been identified in assemblage E. Assemblage E, like assemblages A and B, has a broad host range (hoofed livestock) and so it was plausible that it may comprise host-specific subgroups. The subgrouping was most obvious at the *β giardin* locus again, but also between the 2 full-length *SSU rDNA* sequences. The nucleotide sequences do not yet appear to segregate in the host-specific manner that the allozyme data have demonstrated (Monis *et al*. 2003), but the number of representative sequences from the different livestock hosts were still low and disproportionate.

It was not possible to speculate on the subgroups forming in the other assemblages (C, D, F and G) as there was too little data available; small sample numbers, insufficient reference isolates and limited host variability. For assemblage D, there was 1 apparently host-specific polymorphism in the *β giardin* alignment where samples have included dogs and a coyote described by Santin *et al*. (2003), but for assemblage G, including rats and a mouse in the *gdh* alignment, there was no host-specific polymorphism.

Consensus sequence substitutions

The consensus sequence alignments at the different loci show the differences in total, synonymous, non-synonymous and unique substitution rates between the genes and gene regions (Figs 1 and 2). Different substitution rates and types allow for the detection and comparison of different groups formed at different times. The slower evolving (conserved) genes are usually targeted to detect and compare older (more distant) groups whereas the quicker evolving (variable) loci can detect and compare new and emerging groups that may not yet have polymorphisms at the conserved loci. The substitution rates found for the *SSU rDNA*, *gdh* and *tpi* alignments (0·01, 0·06 and 0·12 substitutions per nucleotide respectively) were similar to those previously described (Monis *et al*. 1999). The typical applications of these loci have therefore varied accordingly. For *Giardia* the conserved *SSU rDNA* is traditionally used for species and assemblage/subassemblage level genotyping (Sogin *et al*. 1989; van Keulen *et al*. 1991, 1993, 1995; Hopkins *et al*. 1997), where as the most variable locus, *tpi*, is frequently used for subtyping clinical samples (Lu *et al*. 1998; Amar *et al*. 2002) and the *gdh* locus, with a substitution rate midway between them, has a broad application spectrum (Monis *et al*. 1996, 1999).

The newer gene in use, *β giardin* (Mahbubani *et al*. 1992; Caccio *et al*. 2002), has interesting properties but has yet to demonstrate consistent subgenotyping with respect to the other loci. It is both conserved in total number of substitutions (0·03 substitutions per nucleotide) as well as non-synonymous substitution rate (5% of total compared to 15% for *gdh* and 25% for *tpi* – representing only 2 amino acid changes in assemblages C and D) and yet demonstrates prominent subgrouping within the assemblages and clear polymorphisms between the assemblages. More than half of the total substitutions for an assemblage were unique to the particular assemblage (60% compared to 30% in the other loci). This was interesting because both the variable *tpi* and conserved *SSU rDNA* had the same 'rate' of unique substitutions. The two main differences in the giardin genes that may be responsible for this characteristic are their age and their function. The giardin genes are *Giardia* specific (more recently evolved) and they are structural (rather than metabolic). The increased rate of unique substitutions appears more likely to be the result of the gene's age than function, as the majority of the substitutions were synonymous. In any event, these few and deliberate polymorphisms may eventually prove useful in determining and differentiating the subgroups of the assemblages.

Within a gene, regions of different substitution rates may also be targeted for different applications. For example, the variable 5′ and 3′ ends of the *SSU rDNA* locus are targeted for genotyping the relatively closely related assemblages, whereas the more conserved regions would only provide sufficient information for the differentiation of *Giardia* species (and above) but not for within *G. duodenalis*. Using the current consensus sequence information, different genes and gene regions can be targeted for different genotyping applications. Older primers can be reassessed for their suitability for their original task – potentially changing the location of the primers to change their scope and/or specificity. For example, the use of the RH primers on cat isolates may not be appropriate (as previously mentioned) due to their limited scope. In another example, the use of the *gdh* primer 'GDHeF' described for the discrimination of all genotypes (Read *et al.* 2004) may no longer be optimal, because a further 2 substitutions in assemblages D, E and F have been discovered, potentially altering the primer's specificity. Similarly, it has been demonstrated how inadequate sequence knowledge in the development of RFLP protocols can affect the accuracy of results significantly (Monis and Andrews, 1998), and therefore that updates on intra-assemblage variation are regularly required.

Consensus sequence phylograms

The relationships between the assemblages of *G. duodenalis* were investigated to understand the history of their divergence and their evolution and adaptability to the different hosts. So far, 7 assemblages have been described for *G. duodenalis* – assemblages A and B (zoonotic), C and D (dogs), E (hoofed livestock), F (cats) and G (rodents). In *G. duodenalis* the dog associated assemblages C and D cluster together in most phylogenetic analyses as do assemblages A, E and F (Monis *et al.* 1999).

The grouping of dog-associated assemblages C and D was well supported in all nucleotide phylograms except for *tpi* (Fig. 3). As seen previously, assemblage C groups with assemblage G at the *tpi* locus (Monis *et al.* 1999) and assemblage D with B (Fig. 3d). Since the *tpi* locus was such a variable locus it was hypothesized that the relationships inferred from the nucleotide sequences may be obscured by the increased substitution rate and that an additional alignment, incorporating only the effects of non-synonymous substitutions, would moderate the results. Conversion of *tpi* alignment to amino acid sequences resulted in assemblage C clustering with assemblage D with strong bootstrap support (89%, Fig. 4b).

The relationships demonstrated between assemblages A (zoonotic), E (hoofed livestock) and F (cats) by the different loci were conflicting, amongst the loci and within them – all combinations of clustering between assemblages A, E and F were presented. The only supported relationship, however, was the clustering of assemblages E and F as sister taxa, this occurred both at the *SSU rDNA* locus (92%, Fig. 3a) and at the *tpi* locus (93%, Fig. 4b). As seen previously, only the *tpi* amino acid sequences were able to elucidate the phylogenetic relationships with strong bootstrap support. The remaining loci (*gdh* and *β giardin*) were unable to resolve the terminal relationship with any support. For the *gdh* locus, the nucleotide and amino acid phylograms clustered differently (A/E, 59%, Fig. 3c and A/F, 39%, Fig. 4c respectively) due to the conflicting signals arising from the different nucleotide positions within the codon. Separate analysis of these positions showed the third (most variable) nucleotide position clustered similar to the original nucleotide sequences (A/E, ∼55%, data not shown) and the first nucleotide position (effecting most non-synonymous change) clustered similar to the amino acid sequences (A/F, ∼30%, data not shown). The opposing signals were presumably due to the effects of random-shared substitutions (homoplastic) obscuring the true-shared substitutions (developed before the divergence, synapomorphic) leaving no single strong phylogenetic signal. Interestingly, the second (most conserved) nucleotide position clustered E with F (∼60%, data not shown), as in the *SSU rDNA* and *tpi* phylograms. The older relationships appeared to have enough synapomorphic substitutions to outweigh the homoplastic substitutions, as their bootstrap support was strong and their phylogenies constant. The *β giardin* locus was less able to infer strong phylogenetic relationships than the other loci because of the higher rate of unique substitutions over shared substitutions producing generally weaker bootstrap values (Fig. 3b). In addition to this, the *β giardin* gene had no non-synonymous substitutions amongst assemblages A, E and F and hence the amino acid sequences (and first and second nucleotide position analyses) provided no information (Fig. 4a). Phylograms from the nucleotide sequences and third nucleotide positions were conflicting and poorly supported (A/F, 45%, Fig. 3b and A/E, ∼50%, data not shown) with the difference in clustering due to the influence of among-site variation (variation within codons or along the gene) on the calculations. When the parameters for this were adjusted, the same clustering was produced for the nucleotide sequences as in the original third nucleotide position phylogram but still with poor support (A/E ∼50%, data not shown). The inability of the *β giardin* locus to resolve the A/E/F relationship was presumably due to the high rate of unique substitutions introducing enough homoplastic substitutions to the sequences to obscure the signal from the synapomorphic substitutions. Similar to the *gdh* locus, the older relationships were better defined,

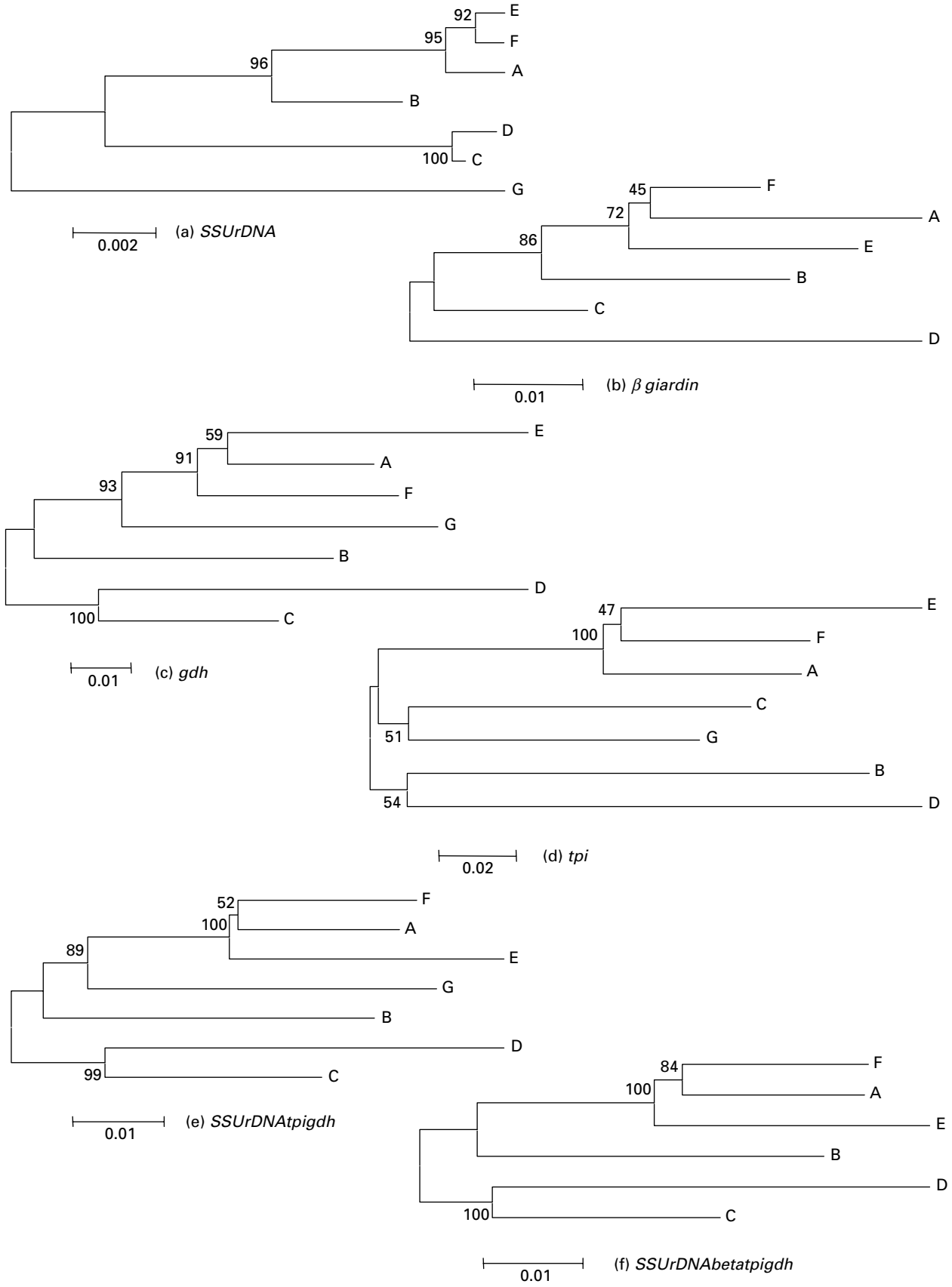

Fig. 3. Nucleotide consensus sequence phylograms. (a) *SSU rDNA* (b) *β giardin* (c) *gdh* (d) *tpi* (e) concatenated sequences (*SSU rDNA*, *tpi* and *gdh*) and (f) concatenated sequences (*SSU rDNA*, *β giardin*, *tpi* and *gdh*). Scale represents substitutions per nucleotide, bootstrapping given as a percentage of 1000 replicates. Tamura-Nei model with uniform rates among sites.

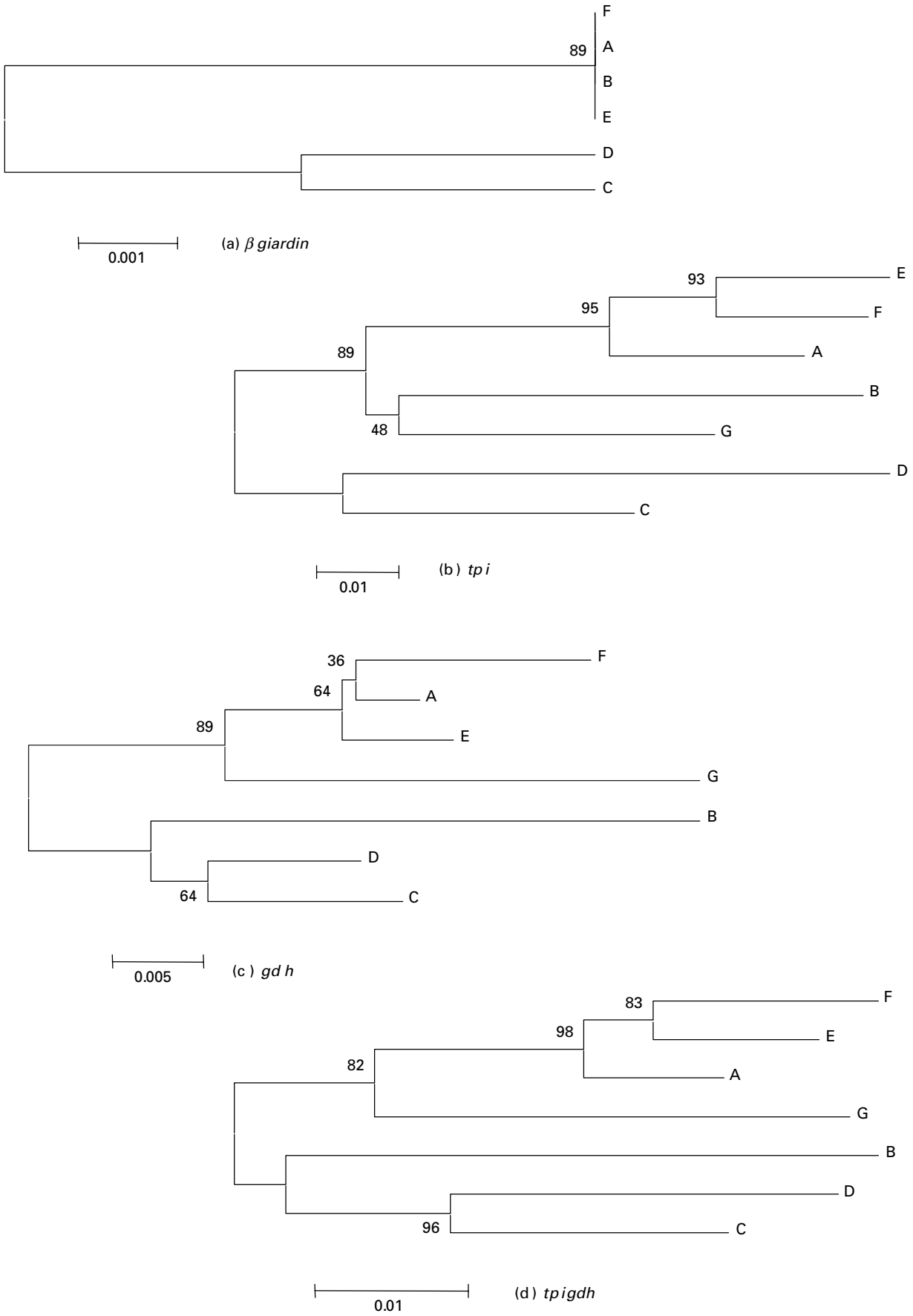

Fig. 4. Amino acid consensus sequence phylograms. (a) *β giardin*, (b) *tpi*, (c) *gdh*, (d) concatenated (*tpi* and *gdh*). Scale represents substitutions per nucleotide, bootstrapping given as a percentage of 1000 replicates. Poisson-corrected model.

with consistent phylogenies and stronger bootstrap values.

Concatenating the sequences from the different loci proved difficult due to the difference in their substitution rates. For the amino acid sequences only the *gdh* and *tpi* genes could be equitably concatenated as the *β giardin* locus had no substitutions (and the *SSU rDNA* is not translated). This combination clustered assemblages E and F with strong bootstrap support (83%, Fig. 4d) due to the strong influence of the *tpi* sequences. For the nucleotide sequences, the variation in substitution rates between and within the genes was very large – as demonstrated by the difference in the sequences and nucleotide positions employed for the original analyses. As a consequence of this marked variation, equitable analysis of the concatenated sequences was not possible with the software package utilized. Basic analysis of the concatenated sequences (assuming uniform substitution rates) clustered assemblages A and F with strong support (84%, Fig. 3f), adjusting these analyses for among-site variation, however, resulted in weak support (~65%, data not shown). Accurate analyses would require specialized software that could include sequence weighting and nucleotide/amino acid combinations. In this way the *SSU rDNA* locus would not be overshadowed by the influence of the more variable loci (*β giardin*, *gdh* and *tpi*) and the translated *tpi* sequence (producing strong phylogenetic signals) could be used. For comparative purposes, the *ef1α* nucleotide alignment also grouped assemblages E and F (Monis *et al*. 1999), initially with poor bootstrap support, however upon translation and combination with the *tpi* and *gdh* loci, with strong support (Monis *et al*. 1999). Conversely, a recent multiple loci analysis (53 isolates over 21 enzymes) employing allozyme electrophoresis, grouped assemblage A with assemblage E (Monis *et al*. 2003).

The significance of this apparently close relationship between the cat (F) and hoofed livestock (E) assemblages is unclear. As both assemblages are currently ascribed to domesticated hosts, further samples from wild animals of similar hosts may prove enlightening. Both of the wild artiodactyl sequences available to date were genotyped as assemblage A (Trout *et al*. 2003; van der Giessen *et al*. 2006). Although the white tailed deer had no greater substitutions away from the majority in the *β giardin* and *tpi* alignments, the roe deer sequence contained numerous isolated synonymous substitutions in the *gdh* alignment. These results may indicate that the white tailed deer was host to the zoonotic assemblage A, but the roe deer was host to a new subgroup of assemblage A. More samples are required to speculate further on the relationships between the cat and hoofed livestock assemblages and the wild and domesticated host samples.

The relationships of assemblage B (zoonotic) and G (rodent), relative to the assemblage A/E/F cluster or C/D cluster, were also conflicting. With both the *SSU rDNA* and the *gdh* phylograms (Fig. 3a and c) the bootstrap support was strong for opposite relationships (assemblage B closer to A/E/F in *SSU rDNA* and assemblage G closer to A/E/F in *gdh*) and in the *tpi* phylogram (amino acid, Fig. 4b) they were weakly clustered together. Concatenating the sequences only resulted in the relationship of the 'stronger' locus being represented – in both the nucleotide and amino acid alignments the *gdh* locus dominated the results with more informative sites than the *SSU rDNA* locus and a stronger phylogenetic signal than the *tpi* locus respectively (Figs 3e and 4d). The order of these relationships may become clearer in future analyses incorporating more of the *tpi* gene as well as a *β giardin* assemblage G sequence. A longer *tpi* sequence may provide more informative sites (potentially clarifying the order of these relationships) and input from the *β giardin* loci on distant relationships is predicted to be more instructive than that on the newer relationships. In addition, future analyses employing a suitable out-group would also provide some perspective on these apparently older lineages.

Mixed templates

One of the advantages of the *SSU rDNA* locus for genotyping is its ability to easily detect mixed templates. Preliminary data from our laboratory (unpublished observations), using combinations of cloned isolates from assemblages A and B, suggest a detection limit of the minority group at around 20–30% with our current *SSU rDNA* protocols (Read *et al*. 2002).

There is, however, some uncertainty about the origin of mixed templates. DNA sequence-based studies that have included both cultured and environmental samples, have only found evidence of mixed templates (multiple assemblages detected) in the environmental samples. For example, in the study by Weiss *et al*. (1992), multiple probes specific for different assemblages bound with single samples in nearly half of the samples analysed and in the study by Read *et al*. (2004), one quarter of the samples amplified at two loci produced a different assemblage result per locus – 'assemblage swapping' (results for repeated analyses were not provided). This has lead many researchers to assume that mixed templates arise from mixed/concurrent infections. However, allozyme studies have shown mixed banding patterns in cultured (Meloni *et al*. 1988; Andrews *et al*. 1989; Stranden *et al*. 1990; Monis *et al*. 2003) and cloned (Meloni *et al*. 1989) isolates, suggesting not only potentially mixed samples but also possible allelic polymorphisms and/or post-translational modifications. In addition to this, mechanisms of

introgression and retention of ancestral genes have been proposed as possible explanations for examples of 'assemblage swapping' detected in some environmental samples (Traub *et al*. 2004).

For the purposes of genotyping (using direct-sequencing methods), qualities like allelic polymorphism and the retention of ancestral genes (that have the ability to generate mixed genotypes at a single locus) and introgression (that has the ability to change genotypes at different loci) are of concern and require investigation. In environmental samples that produce conflicting results, the ability to selectively culture each suspected genotype individually from the original sample (for cloning and reanalysis) would be useful to examine these possibilities. Presently, however, the only genotype that is easily cultured is assemblage A (Nash and Keister, 1985; Meloni and Thompson, 1987; Andrews *et al*. 1989), due to its wide host range and nutrient adaptability, and few people have investigated in detail the more specific conditions required by the other assemblages (Binz *et al*. 1992). In the cultured and cloned isolates producing mixed (allozyme) banding patterns (notably the original BIII and BIV reference isolates), genetic sequence analyses of those mixed loci to establish the presence of allelic polymorphisms, and consequently their segregation (assemblage-specific or not), would be useful. Although most of these loci have not been genetically characterised in *Giardia* with a set of reference isolates for comparison, this would be straightforward with the aid of the *Giardia* genome project (McArthur *et al*. 2000), the *Giardia lamblia* Genome Database – GiardiaDB (http://gmod.mbl.edu/perl/site/giardia?page=intro) and their contributions to GenBank. If assemblage-segregating alleles were found, it would be valuable to identify the genes that were stable enough for use in genotyping.

Further to these uncertain traits, the *Giardia* genome has been demonstrated to be quite plastic – with a high degree of recombination in the telemetric regions of the chromosomes containing the variable surface protein (vsp) and ribosomal repeat unit (rDNA unit) genes (Adam *et al*. 1988, 1992; Adam, 1992; Le Blancq *et al*. 1991*b*, 1992) as well as an example of a truncated rDNA unit (Le Blancq *et al*. 1991*a*). The main concern here would be the effect of recombination on the rDNA repeat units used in genotyping. For direct sequencing methods, low levels of variation amongst the rDNA repeat units would go undetected and high levels would result in excessive noise. Since the excess noise is not found, it could be suggested that the recombination events are primarily directed at effecting change in the variable surface protein genes and not the rDNA repeat units. In fact, as this is a standard mechanism for boosting genetic diversity in host defence proteins (Roitt *et al*. 1996; Buchanan *et al*. 2000) it could be expected that analogous mechanisms would be utilized for pathogen surface proteins.

Conclusions

The current data set was limited in several key areas, namely, there were few lengthy *SSU rDNA* sequences available, there were few assemblage F and G sequences available (none at some loci), there were no full-length *tpi* sequences for assemblages C, D, E, F and G and there were too few assemblage B sequences to gauge the extent of the intra-assemblage variability.

All loci were found capable of genotyping to the assemblage and subassemblage (AI/AII) level, although each gene was evidently suited to different applications. The *tpi* gene, as the most variable, was ideal for its usual application in research focusing on strain identification and was also capable of establishing the closer (more recent) phylogenetic relationships. The *gdh* gene, with moderate variability, was also suited for analyses on strain identification as well as all genotyping applications (providing numerous informative sites) and some phylogenetic analyses (of the more distantly related assemblages). The *β giardin* gene, as a unique and conserved locus, has great potential for defining the core subgroups within the assemblages; however, the continuity of these subgroups across all loci must be established. The *SSU rDNA* was suitable for both subassemblage (AI/AII) level genotyping and phylogenetic analyses and also has potential for defining the core assemblage subgroups (notably within Assemblage B). It remains a well-suited locus for routine genotyping from environmental samples due to its high copy number.

The phylogenetic relationships within *Giardia duodenalis*, although apparently resolved for the closer (more recently diverged) assemblages, requires more sequence data to establish those remaining. More complete sequences for *tpi* and *β giardin* (representing all assemblages) and analyses incorporating an out-group may be sufficient, but more loci may be required.

The current issues surrounding mixed templates demonstrate a need for more research on culturing and cloning different assemblages of *Giardia*. Cloned isolates will be required to investigate the potential allelic polymorphisms and introgression and so representatives from each assemblage would provide a more comprehensive analysis.

REFERENCES

Abe, N., Kimata, I. and Tokoro, M. (2005*a*). Genotyping of *Giardia* isolates from humans in Japan using the small subunit ribosomal RNA and glutamate dehydrogenase gene sequences. *Japanese Journal of Infectious Diseases* **58**, 57–58.

Abe, N., Read, C., Thompson, R. C. and Iseki, M. (2005*b*). Zoonotic genotype of *Giardia intestinalis* detected in a ferret. *The Journal of Parasitology* **91**, 179–182.

Adam, R. D. (1992). Chromosome-size variation in *Giardia lamblia*: the role of rDNA repeats. *Nucleic Acids Research* **20**, 3057–3061.

Adam, R. D. (2001). Biology of *Giardia lamblia*. *Clinical Microbiology Reviews* **14**, 447–475.

Adam, R. D., Aggarwal, A., Lal, A. A., De La Cruz, V. F., McCutchan, T. and Nash, T. E. (1988). Antigenic variation of a cysteine-rich protein in *Giardia lamblia*. *Journal of Experimental Medicine* **167**, 109–118.

Adam, R. D., Yang, Y. M. and Nash, T. E. (1992). The cysteine-rich protein gene family of *Giardia lamblia*: loss of the CRP170 gene in an antigenic variant. *Molecular and Cellular Biology* **12**, 1194–1201.

Amar, C. F., Dear, P. H., Pedraza-Diaz, S., Looker, N., Linnane, E. and McLauchlin, J. (2002). Sensitive PCR-restriction fragment length polymorphism assay for detection and genotyping of *Giardia duodenalis* in human feces. *Journal of Clinical Microbiology* **40**, 446–452.

Andrews, R. H., Adams, M., Boreham, P. F., Mayrhofer, G. and Meloni, B. P. (1989). *Giardia intestinalis*: electrophoretic evidence for a species complex. *International Journal for Parasitology* **19**, 183–190.

Baruch, A. C., Isaac-Renton, J. and Adam, R. D. (1996). The molecular epidemiology of *Giardia lamblia*: a sequence-based approach. *The Journal of Infectious Diseases* **174**, 233–236.

Berrilli, F., Di Cave, D., De Liberato, C., Franco, A., Scaramozzino, P. and Orecchia, P. (2004). Genotype characterisation of *Giardia duodenalis* isolates from domestic and farm animals by SSU-rRNA gene sequencing. *Veterinary Parasitology* **122**, 193–199.

Binz, N., Thompson, R. C., Lymbery, A. J. and Hobbs, R. P. (1992). Comparative studies on the growth dynamics of two genetically distinct isolates of *Giardia duodenalis in vitro*. *International Journal for Parasitology* **22**, 195–202.

Boothroyd, J. C., Wang, A., Campbell, D. A. and Wang, C. C. (1987). An unusually compact ribosomal DNA repeat in the protozoan *Giardia lamblia*. *Nucleic Acids Research* **15**, 4065–4084.

Buchanan, B., Gruissem, W. and Jones, R. (2000). *Biochemistry and Molecular Biology of Plants*, American Society of Plant Physiologists Maryland, USA.

Caccio, S. M., De Giacomo, M. and Pozio, E. (2002). Sequence analysis of the beta-giardin gene and development of a polymerase chain reaction-restriction fragment length polymorphism assay to genotype *Giardia duodenalis* cysts from human faecal samples. *International Journal for Parasitology* **32**, 1023–1030.

Caccio, S. M., Thompson, R. C., McLauchlin, J. and Smith, H. V. (2005). Unravelling *Cryptosporidium* and *Giardia* epidemiology. *Trends in Parasitology* **21**, 430–437.

Edlind, T. D. and Chakraborty, P. R. (1987). Unusual ribosomal RNA of the intestinal parasite *Giardia lamblia*. *Nucleic Acids Research* **15**, 7889–7901.

Erlandsen, S. L. and Bemrick, W. J. (1987). SEM evidence for a new species, *Giardia psittaci*. *The Journal of Parasitology* **73**, 623–629.

Erlandsen, S. L., Bemrick, W. J., Wells, C. L., Feely, D. E., Knudson, L., Campbell, S. R., Van Keulen, H. and Jarroll, E. L. (1990). Axenic culture and characterization of *Giardia ardeae* from the great blue heron (*Ardea herodias*). *The Journal of Parasitology* **76**, 717–724.

Ey, P. L., Andrews, R. H. and Mayrhofer, G. (1993). Differentiation of major genotypes of *Giardia intestinalis* by polymerase chain reaction analysis of a gene encoding a trophozoite surface antigen. *Parasitology* **106**, 347–356.

Ey, P. L., Khanna, K., Andrews, R. H., Manning, P. A. and Mayrhofer, G. (1992). Distinct genetic groups of *Giardia intestinalis* distinguished by restriction fragment length polymorphisms. *Journal of General Microbiology* **138**, 2629–2637.

Ey, P. L., Mansouri, M., Kulda, J., Nohynkova, E., Monis, P. T., Andrews, R. H. and Mayrhofer, G. (1997). Genetic analysis of *Giardia* from hoofed farm animals reveals artiodactyl-specific and potentially zoonotic genotypes. *The Journal of Eukaryotic Microbiology* **44**, 626–635.

Fayer, R., Santin, M., Trout, J. M. and Dubey, J. P. (2006). Detection of *Cryptosporidium felis* and *Giardia duodenalis* Assemblage F in a cat colony. *Veterinary Parasitology* **140**, 44–53.

Filice, F. P. (1952). Studies on the cytology and life history of a *Giardia* from the laboratory rat. *University of California Publications in Zoology* **57**, 53–146.

Guy, R. A., Xiao, C. and Horgen, P. A. (2004). Real-time PCR assay for detection and genotype differentiation of *Giardia lamblia* in stool specimens. *Journal of Clinical Microbiology* **42**, 3317–3320.

Healey, A., Mitchell, R., Upcroft, J. A., Boreham, P. F. and Upcroft, P. (1990). Complete nucleotide sequence of the ribosomal RNA tandem repeat unit from *Giardia intestinalis*. *Nucleic Acids Research* **18**, 4006.

Holberton, D. V. and Marshall, J. (1995). Analysis of consensus sequence patterns in *Giardia* cytoskeleton gene promoters. *Nucleic Acids Research* **23**, 2945–2953.

Homan, W. L., Van Enckevort, F. H., Limper, L., Van Eys, G. J., Schoone, G. J., Kasprzak, W., Majewska, A. C. and Van Knapen, F. (1992). Comparison of *Giardia* isolates from different laboratories by isoenzyme analysis and recombinant DNA probes. *Parasitology Research* **78**, 316–323.

Hopkins, R. M., Meloni, B. P., Groth, D. M., Wetherall, J. D., Reynoldson, J. A. and Thompson, R. C. (1997). Ribosomal RNA sequencing reveals differences between the genotypes of *Giardia* isolates recovered from humans and dogs living in the same locality. *The Journal of Parasitology* **83**, 44–51.

Hunt, C. L., Ionas, G. and Brown, T. J. (2000). Prevalence and strain differentiation of *Giardia intestinalis* in calves in the Manawatu and Waikato regions of North Island, New Zealand. *Veterinary Parasitology* **91**, 7–13.

Hunter, P. R. and Thompson, R. C. (2005). The zoonotic transmission of *Giardia* and *Cryptosporidium*. *International Journal for Parasitology* **35**, 1181–1190.

Itagaki, T., Kinoshita, S., Aoki, M., Itoh, N., Saeki, H., Sato, N., Uetsuki, J., Izumiyama, S., Yagita, K. and Endo, T. (2005). Genotyping of *Giardia intestinalis* from domestic and wild animals in Japan using glutamate dehydrogenase gene sequencing. *Veterinary Parasitology* **133**, 283–287.

Kumar, S., Tamura, K. and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150–163.

Lalle, M., Jimenez-Cardosa, E., Caccio, S. M. and Pozio, E. (2005*a*). Genotyping of *Giardia duodenalis* from humans and dogs from Mexico using a beta-giardin nested polymerase chain reaction assay. *The Journal of Parasitology* **91**, 203–205.

Lalle, M., Pozio, E., Capelli, G., Bruschi, F., Crotti, D. and Caccio, S. M. (2005*b*). Genetic heterogeneity at the beta-giardin locus among human and animal isolates of *Giardia duodenalis* and identification of potentially zoonotic subgenotypes. *International Journal for Parasitology* **35**, 207–213.

Le Blancq, S. M., Kase, R. S. and Van Der Ploeg, L. H. (1991*a*). Analysis of a *Giardia lamblia* rRNA encoding telomere with [TAGGG]n as the telomere repeat. *Nucleic Acids Research* **19**, 5790.

Le Blancq, S. M., Korman, S. H. and Van Der Ploeg, L. H. (1991*b*). Frequent rearrangements of rRNA-encoding chromosomes in *Giardia lamblia*. *Nucleic Acids Research* **19**, 4405–4412.

Le Blancq, S. M., Korman, S. H. and Van Der Ploeg, L. H. (1992). Spontaneous chromosome rearrangements in the protozoan *Giardia lamblia*: estimation of mutation rates. *Nucleic Acids Research* **20**, 4539–4545.

Leonhard, S., Pfister, K., Beelitz, P., Wielinga, C., Traub, R. and Thompson, R. C. A. (2006). The molecular characterisation of *Giardia* from dogs in Munich. In *22nd Annual Convention of the German Society for Parasitology, University of Veterinary Medicine Vienna*, pp. 163, Abstract in programme.

Lu, S. Q., Baruch, A. C. and Adam, R. D. (1998). Molecular comparison of *Giardia lamblia* isolates. *International Journal for Parasitology* **28**, 1341–1345.

Mahbubani, M. H., Bej, A. K., Perlin, M. H., Schaefer, F. W. 3rd, Jakubowski, W. and Atlas, R. M. (1992). Differentiation of *Giardia duodenalis* from other *Giardia spp*. by using polymerase chain reaction and gene probes. *Journal of Clinical Microbiology* **30**, 74–78.

Matsubayashi, M., Kimata, I. and Abe, N. (2005). Identification of genotypes of *Giardia intestinalis* isolates from a human and calf in Japan. *The Journal of Veterinary Medical Science* **67**, 337–340.

Mayrhofer, G., Andrews, R. H., Ey, P. L. and Chilton, N. B. (1995). Division of *Giardia* isolates from humans into two genetically distinct assemblages by electrophoretic analysis of enzymes encoded at 27 loci and comparison with *Giardia muris*. *Parasitology* **111**, 11–17.

McArthur, A. G., Morrison, H. G., Nixon, J. E., Passamaneck, N. Q., Kim, U., Hinkle, G., Crocker, M. K., Holder, M. E., Farr, R., Reich, C. I., Olsen, G. E., Aley, S. B., Adam, R. D., Gillin, F. D. and Sogin, M. L. (2000). The *Giardia* genome project database. *FEMS Microbiology Letters* **189**, 271–273.

Meloni, B. P., Lymbery, A. J. and Thompson, R. C. (1988). Isoenzyme electrophoresis of 30 isolates of *Giardia* from humans and felines. *The American Journal of Tropical Medicine and Hygiene* **38**, 65–73.

Meloni, B. P., Lymbery, A. J. and Thompson, R. C. (1989). Characterization of *Giardia* isolates using a non-radiolabeled DNA probe, and correlation with the results of isoenzyme analysis. *The American Journal of Tropical Medicine and Hygiene* **40**, 629–637.

Meloni, B. P. and Thompson, R. C. (1987). Comparative studies on the axenic *in vitro* cultivation of *Giardia* of human and canine origin: evidence for intraspecific variation. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **81**, 637–640.

Monis, P. T. and Andrews, R. H. (1998). Molecular epidemiology: assumptions and limitations of commonly applied methods. *International Journal for Parasitology* **28**, 981–987.

Monis, P. T., Andrews, R. H., Mayrhofer, G. and Ey, P. L. (1999). Molecular systematics of the parasitic protozoan *Giardia intestinalis*. *Molecular Biology and Evolution* **16**, 1135–1144.

Monis, P. T., Andrews, R. H., Mayrhofer, G. and Ey, P. L. (2003). Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* **3**, 29–38.

Monis, P. T., Andrews, R. H., Mayrhofer, G., Mackrill, J., Kulda, J., Isaac-Renton, J. L. and Ey, P. L. (1998). Novel lineages of *Giardia intestinalis* identified by genetic analysis of organisms isolated from dogs in Australia. *Parasitology* **116**, 7–19.

Monis, P. T., Mayrhofer, G., Andrews, R. H., Homan, W. L., Limper, L. and Ey, P. L. (1996). Molecular genetic analysis of *Giardia intestinalis* isolates at the glutamate dehydrogenase locus. *Parasitology* **112**, 1–12.

Mowatt, M. R., Weinbach, E. C., Howard, T. C. and Nash, T. E. (1994). Complementation of an *Escherichia coli* glycolysis mutant by *Giardia lamblia* triosephosphate isomerase. *Experimental Parasitology* **78**, 85–92.

Nash, T. (1992). Surface antigen variability and variation in *Giardia lamblia*. *Parasitology Today* **8**, 229–234.

Nash, T. E. and Keister, D. B. (1985). Differences in excretory-secretory products and surface antigens among 19 isolates of *Giardia*. *The Journal of Infectious Diseases* **152**, 1166–1171.

Page, R. D. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences. *CABIOS* **12**, 357–358.

Read, C., Walters, J., Robertson, I. D. and Thompson, R. C. (2002). Correlation between genotype of *Giardia duodenalis* and diarrhoea. *International Journal for Parasitology* **32**, 229–231.

Read, C. M., Monis, P. T. and Thompson, R. C. (2004). Discrimination of all genotypes of *Giardia duodenalis* at the glutamate dehydrogenase locus using PCR-RFLP. *Infection, Genetics and Evolution: Journal*

of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases **4**, 125–130.

Robertson, L. J., Hermansen, L., Gjerde, B. K., Strand, E., Alvsvag, J. O. and Langeland, N. (2006). Application of genotyping during an extensive outbreak of waterborne giardiasis in Bergen, Norway, during autumn and winter 2004. *Applied and Environmental Microbiology* **72**, 2212–2217.

Roitt, I., Brostoff, J. and Male, D. (1996). *Immunology*, 4th Edn., Times Mirror International Publishers Limited London, UK.

Santin, M., Ludwig, K., Fayer, R. and Trout, J. M. (2003). First report of *Giardia* in coyotes (*Canis latrans*). *The Journal of Eukaryotic Microbiology* **50** (Suppl.) S709.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. and Peattie, D. A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* **243**, 75–77.

Stranden, A. M., Eckert, J. and Kohler, P. (1990). Electrophoretic characterization of *Giardia* isolated from humans, cattle, sheep, and a dog in Switzerland. *The Journal of Parasitology* **76**, 660–668.

Sulaiman, I. M., Fayer, R., Bern, C., Gilman, R. H., Trout, J. M., Schantz, P. M., Das, P., Lal, A. A. and Xiao, L. (2003). Triosephosphate isomerase gene characterization and potential zoonotic transmission of *Giardia duodenalis*. *Emerging Infectious Diseases* **9**, 1444–1452.

Sulaiman, I. M., Jiang, J., Singh, A. and Xiao, L. (2004). Distribution of *Giardia duodenalis* genotypes and subgenotypes in raw urban wastewater in Milwaukee, Wisconsin. *Applied and Environmental Microbiology* **70**, 3776–3780.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

Thompson, J. D., Higgins, D. G. and 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.

Thompson, R. C., Hopkins, R. M. and Homan, W. L. (2000). Nomenclature and genetic groupings of *Giardia* infecting mammals. *Parasitology Today* **16**, 210–213.

Thompson, R. C., Lymbery, A. J. and Meloni, B. P. (1990). Genetic variation in Giardia Kunstler, 1882: taxonomic and epidemiological significance. *Protozoological Abstracts* **14**, 1–28.

Thompson, R. C. and Monis, P. T. (2004). Variation in *Giardia*: implications for taxonomy and epidemiology. *Advances in Parasitology* **58**, 69–137.

Traub, R. J., Monis, P. T., Robertson, I., Irwin, P., Mencke, N. and Thompson, R. C. (2004). Epidemiological and molecular evidence supports the zoonotic transmission of *Giardia* among humans and dogs living in the same community. *Parasitology* **128**, 253–262.

Trout, J. M., Santin, M. and Fayer, R. (2003). Identification of assemblage A *Giardia* in white-tailed deer. *The Journal of Parasitology* **89**, 1254–1255.

Trout, J. M., Santin, M., Greiner, E. and Fayer, R. (2004). Prevalence of *Giardia duodenalis* genotypes in pre-weaned dairy calves. *Veterinary Parasitology* **124**, 179–186.

Upcroft, J. A., Healey, A. and Upcroft, P. (1994). A new rDNA repeat unit in human *Giardia*. *The Journal of Eukaryotic Microbiology* **41**, 639–642.

Upcroft, J. A. and Upcroft, P. (1994). Two distinct varieties of *Giardia* in a mixed infection from a single human patient. *The Journal of Eukaryotic Microbiology* **41**, 189–194.

Van Der Giessen, J. W., De Vries, A., Roos, M., Wielinga, P., Kortbeek, L. M. and Mank, T. G. (2006). Genotyping of *Giardia* in Dutch patients and animals: a phylogenetic analysis of human and animal isolates. *International Journal for Parasitology* **36**, 849–858.

Van Keulen, H., Campbell, S. R., Erlandsen, S. L. and Jarroll, E. L. (1991). Cloning and restriction enzyme mapping of ribosomal DNA of *Giardia duodenalis, Giardia ardeae* and *Giardia muris*. *Molecular and Biochemical Parasitology* **46**, 275–284.

Van Keulen, H., Feely, D. E., Macechko, P. T., Jarroll, E. L. and Erlandsen, S. L. (1998). The sequence of *Giardia* small subunit rRNA shows that voles and muskrats are parasitized by a unique species *Giardia microti*. *The Journal of Parasitology* **84**, 294–300.

Van Keulen, H., Gutell, R. R., Gates, M. A., Campbell, S. R., Erlandsen, S. L., Jarroll, E. L., Kulda, J. and Meyer, E. A. (1993). Unique phylogenetic position of Diplomonadida based on the complete small subunit ribosomal RNA sequence of *Giardia ardeae, G. muris, G. duodenalis and Hexamita sp*. *The FASEB Journal* **7**, 223–231.

Van Keulen, H., Homan, W. L., Erlandsen, S. L. and Jarroll, E. L. (1995). A three nucleotide signature sequence in small subunit rRNA divides human *Giardia* in two different genotypes. *The Journal of Eukaryotic Microbiology* **42**, 392–394.

Van Keulen, H., Macechko, P. T., Wade, S., Schaaf, S., Wallis, P. M. and Erlandsen, S. L. (2002). Presence of human *Giardia* in domestic, farm and wild animals, and environmental samples suggests a zoonotic potential for giardiasis. *Veterinary Parasitology* **108**, 97–107.

Weiss, J. B., Van Keulen, H. and Nash, T. E. (1992). Classification of subgroups of *Giardia lamblia* based upon ribosomal RNA gene sequence using the polymerase chain reaction. *Molecular and Biochemical Parasitology* **54**, 73–86.

Yee, J. and Dennis, P. P. (1992). Isolation and characterization of a NADP-dependent glutamate dehydrogenase gene from the primitive eucaryote *Giardia lamblia*. *The Journal of Biological Chemistry* **267**, 7539–7544.

Yong, T. S., Park, S. J., Hwang, U. W., Yang, H. W., Lee, K. W., Min, D. Y., Rim, H. J., Wang, Y. and Zheng, F. (2000). Genotyping of *Giardia lamblia* isolates from humans in China and Korea using ribosomal DNA Sequences. *The Journal of Parasitology* **86**, 887–891.