Multilocus Sequence Typing Tool for Cyclospora cayetanensis

Yaqiong Guo, Dawn M. Roellig, Na Li, Kevin Tang, Michael Frace, Ynes Ortega, Michael J. Arrowood, Yaoyu Feng, Yvonne Qvarnstrom, Lin Wang, Delynn M. Moss, Longxian Zhang, Lihua Xiao

Because the lack of typing tools for *Cyclospora cayetanensis* has hampered outbreak investigations, we sequenced its genome and developed a genotyping tool. We observed 2 to 10 geographically segregated sequence types at each of 5 selected loci. This new tool could be useful for case linkage and infection/contamination source tracking.

Cyclospora cayetanensis is an emerging parasitic pathogen responsible for numerous foodborne outbreaks of cyclosporiasis in North America, primarily associated with imported fresh produce from cyclosporiasis-endemic areas (1). The lack of genotyping tools has hampered case linkage and infection/contamination source tracking (2). In this study, we developed a multilocus sequence typing (MLST) tool to help with identification of this protozoan.

The Study

To identify potential genotyping markers, we sequenced the genome of 1 *C. cayetanensis* isolate (CHN_HEN01) from Henan, China (*3*), and searched for microsatellite and minisatellite sequences among the first 40 of 4,811 assembled contigs by using Tandem Repeat Finder software (http://tandem.bu.edu/trf/trf.html). We designed primers for nested PCR analysis of the targets based on flanking nucleotide sequences.

The total volume of PCR mixture was 50 mL, which contained 1 mL of DNA (for primary PCR) or 2 mL of the primary PCR product, 250 nmol/L primers, 3 mmol/L magnesium chloride, 200 μ mol/L deoxynucleotide triphosphates, 1× GeneAmp PCR buffer (Applied Biosystems, Foster City, CA, USA), and 1.5 U of Taq polymerase (Promega, Madison, WI, USA). The amplification consisted of

Author affiliations: Centers for Disease Control and Prevention, Atlanta, Georgia, USA (Y. Guo, D.M. Roellig, K. Tang, M. Frace, M.J. Arrowood, Y. Qvarnstrom, D.M. Moss, L. Xiao); East China University of Science and Technology, Shanghai, China (Y. Guo, N. Li, Y. Feng, L. Wang); University of Georgia, Griffin, Georgia, USA (Y. Ortega); Henan Agricultural University, Zhengzhou, China (L. Zhang)

DOI: http://dx.doi.org/10.3201/eid2208.150696

an initial denaturation at 94°C for 5 min; 35 cycles at 94°C for 45 s; a specified annealing temperature (Table; online Technical Appendix Table 1, http://wwwnc.cdc.gov/EID/article/22/8/15-1696-Techapp1.pdf) for 45 s and 72°C for 1 min; and a final extension at 72°C for 7 min. The secondary PCR products were sequenced in both directions on an ABI 3130 Genetic Analyzer (Applied Biosystems).

The sequences obtained from each locus were aligned by using ClustalX version 2.1 (http://www.clustal.org). A neighbor-joining analysis was used to assess the genetic relatedness of various *C. cayetanensis* sequences for each locus and concatenated sequences of 5 loci. Unique sequences generated from the 5 MLST loci were deposited in GenBank (accession nos. KP723491–KP723518).

Altogether, 15 loci were chosen for evaluations (Table; online Technical Appendix Table 1). These loci included 13 microsatellite and 2 minisatellite loci. Six specimens from China and Peru were used in the initial evaluation of the PCR primers designed. Five microsatellite loci (CYC3, CYC13, CYC15, CYC21, and CYC22) exhibiting high PCR amplification efficiency and nucleotide sequence polymorphism in the initial evaluation were chosen for further evaluations of the nature of nucleotide sequence polymorphism by using a total of 64 C. cayetanensis specimens from China (n = 26), Nepal (n = 3), Indonesia (n =1), Guatemala (n = 2), Peru (n = 8), Spain (n = 1), and the United States (n = 23) (online Technical Appendix Table 2). Of these, 63 specimens were amplified by PCR at the CYC3 locus, 61 at the CYC13 locus, 63 at the CYC15 locus, 62 at the CYC21 locus, and 64 at the CYC21 locus (Table). However, 1-11 specimens did not produce readable sequences at each locus.

Nucleotide sequence alignment led to the identification of 4 sequence types at locus CYC3, 10 at locus CYC13, 2 at locus CYC15, 8 at locus CYC21, and 4 at locus CYC22 (online Technical Appendix Table 2). As expected, all 5 loci showed differences in the number of microsatellite repeats. In addition, single nucleotide polymorphisms were present at all loci (online Technical Appendix Figures 1–5). Sequences from CYC3, CYC13, CYC21, and CYC22 formed 2–3 major groups in neighbor-joining trees (Figure 1). Clear geographic clustering of sequences was observed at most loci, with specimens from China largely clustering together and US outbreak specimens often clustering with specimens from Peru (Figure 1). Of the 9 specimens from a 2013 Texas outbreak, 1 had a different sequence from the remaining specimens at CYC3, 2 had different sequences

	Contig	Targeted	· · · · ·	Annealing	Expected	Amplification efficiency,
Locus	no.	repeat*	Primer sequence, $5' \rightarrow 3'^{\dagger}$	temp, °C	size, bp	no. positive/no. analyzed
CYC3	00003	TGTA ₆₃	F1: GAAGATGAAGCGTTGGTACG;	55	598	4/4
		and	R1: TACCGCTGCTGGAGTGCAT;			
		TATA ₂₃	F2: TTGTGCATGGCACCCAATGC;			
			R2: CCAGACAGTAGTTCGTGTCTT			
CYC13	80000	GAT ₁₅	F1: TTGGAGCAGGACGAGTTTCG;	58	595	4/4
			R1: ATGGAAGCGGCTATGAAATTGG;			
			F2: CCTCGGAGTCCTCTGAGTG;			
			R2: AGCCGTCGCAGTGTGTAGCA			
CYC15	00009	TGC ₁₁	F1: AGTAGCTACGTGCCAAGACGA;	58	609	4/4
			R1: TCGTTCTATCTGACCATAGTAGTG;			
			F2: CGCTGTGCAAGAGGCGATCTA;			
			R2: AAGCACTGCAGGGTCCGTAAC			
CYC21	00036	AT ₃₁	F1: TAGTGGCGACTGCGACATG;	55	471	4/4
			R1: GCACCTTGCTGATGAGGCA;			
			F2: CTA AGGCTGTCTTGAGCGG;			
			R2: CGCCCACATGCTTCGTATAC			
CYC22	00037	AC ₂₀	F1: CACTATGCCGTGTGACACGT;	55	512	4/4
			R1: GTAGATTTGCAAGAACTCATGCTA;			
			F2: ATAGTATTCAGGCGCAAACTAAG;			
			R2: GAGGCTTTCCAAAGGTCTAGTT			

Table. Primer sequences of microsatellite loci used in multilocus sequence analysis of Cyclospora cayetanensis

*Tandem repeat identified in the sequence from whole genome sequencing. †Six *C. cayetanensis* specimens were used in initial evaluation of PCR primers: specimens 22231, 22234, 22238, and 28709 were used to evaluate PCR primers from loci CYC3, whereas specimens 22231, 22234, 24550, and 24552 were used for PCR primers from the remaining loci.

from the remaining specimens at CYC13 and CYC21, and at CYC22, PCR products from 7 specimens produced unreadable sequences (online Technical Appendix Table 2).

A total of 34 specimens had complete sequence data at 5 loci, forming 25 MLST types (online Technical Appendix Table 2). Most of the MLST types had only 1 specimen, except for 4 MLST types (MS3, MS15, MS16, and MS17), which had 3 or 4 specimens (online Technical Appendix Table 2). A neighbor-joining analysis of the concatenated sequences of 2,317 bp showed clear geographic clustering of MLST types (Figure 2). Most specimens from China clustered together in 1 major group, whereas specimens from outbreaks in the United States formed 2 other groups with specimens from Peru. The specimen from Spain appeared to be distinct.

Conclusions

In this study, we sequenced the genome of *C. cayetanensis* protozoa and developed a genotyping tool. Noticeable geographic clustering was observed at some of the loci, with specimens from China forming 1–2 groups at each of these loci. In contrast, the US outbreak specimens mostly grouped together with Peru specimens, probably because of the imported nature of pathogens from Central and South America. The geographic clustering pattern of specimens from the same country at 1 locus does not conform to patterns at other loci, probably because of the occurrence of genetic recombination among parasites in a particular area. Therefore, the use of a single genetic marker is probably not useful in geographic tracking of infection sources of this species.

Data generated from this study have demonstrated the high resolution of the MLST tool. Although genotyping

resulted in complete data at all 5 loci for only 34 of the 64 specimens, 25 MLST types were detected. The failure in obtaining informative sequences from some amplicons was mainly attributable to the presence of PCR products with different repeat lengths, leading to overlapped signals following the tandem repeat region. This highlights some potential challenges in investigations of cyclosporiasis outbreaks using genotyping tools. Only 2 of the 9 specimens from the 2013 outbreak of cyclosporiasis in Texas produced complete MLST data because of inability to obtain readable sequences from CYC22. Sequence analysis at other loci suggested that at least 3 types of C. cayetanensis protozoa were present in specimens from the outbreak. The occurrence of mixed C. cayetanensis populations probably led to unreadable sequences for most specimens from the outbreak at CYC22.

The occurrence of mixed C. cayetanensis populations in large outbreaks is expected because divergent MLST types are apparently present in a small community or geographic area. For example, the Peru specimens in this study were from a small shantytown, Pampas de San Juan de Miraflores, in Lima (4), but the specimens had at least 5 MLST types among them. Similarly, 14 MLST types were detected among the 26 Chinese specimens collected from 2 neighboring cities (Kaifeng and Zhengzhou) in Henan Province (5). Fresh produce is frequently contaminated by C. cayetanensis protozoa through irrigation water (6) and thus has a higher probability of containing multiple C. cayetanensis genotypes. It might be possible to use only 2 or 3 loci that are highly polymorphic and easier to sequence in C. cayetanensis genotyping, such as CYC13 and CYC21.

DISPATCHES



Figure 1. Phylogenetic relationships among sequence types of *Cyclospora cayetanensis* at 5 microsatellite loci: A) CYC3, B) CYC13, C) CYC15, D) CYC21, and E) CYC22. Tree was constructed on the basis of neighbor-joining analyses of the nucleotide sequences, using genetic distances calculated by the Kimura 2-parameter model. Numbers on branches are bootstrap values from 1,000 replicate analyses. Only values >50% are displayed on the left of each node. Scale bars indicate substitution rates per nucleotide. HN, Henan.



Figure 2. Phylogenetic relationships among concatenated multilocus sequence types of *Cyclospora cayetanensis* as assessed by a neighbor-joining analysis of the nucleotide sequences, using genetic distances calculated by the Kimura 2-parameter model. Numbers on branches are bootstrap values from 1,000 replicate analyses. Only values >50% are displayed on the left of each node. Scale bar indicates substitution rates per nucleotide. HN, Henan.

In summary, whole-genome sequence data from *C. cayetanensis* protozoa enabled the development of a MLST tool for characterizing isolates in outbreak investigations. The high resolution of the typing tool and the apparent presence of geographic clusters might facilitate the identification of outbreaks and infection sources. Nevertheless, extensive characterization of specimens from diverse areas and wide application of the developed tool in outbreak investigations are needed to better understand *C. cayetanensis* transmission.

Acknowledgments

We thank Robert H. Gilman and Olga Gonzalez-Moreno Portugal for providing specimens and Lori A. Rowe and Kristine Knipe for technical assistance.

This work was supported in part by the National Natural Science Foundation of China (grant nos. 31425025 and 31229005), the China Postdoctoral Science Foundation (grant no. 2014M560310), and the US Centers for Disease Control and Prevention.

Ms. Guo is a lecturer at East China University of Science and Technology. Her research interests include molecular epidemiology and genomics of foodborne, waterborne, and zoonotic parasites.

References

- Hall RL, Jones JL, Herwaldt BL. Surveillance for laboratoryconfirmed sporadic cases of cyclosporiasis—United States, 1997–2008. MMWR Surveill Summ. 2011;60:1–11.
- desVignes-Kendrick M, Reynolds K, Lee T, Gaul L, Klein K, Irvin K, et al. Outbreaks of cyclosporiasis—United States, June– August 2013. MMWR Morb Mortal Wkly Rep. 2013;62:862.
- Tang K, Guo Y, Zhang L, Rowe LA, Roellig DM, Frace MA, et al. Genetic similarities between *Cyclospora cayetanensis* and cecum-infecting avian *Eimeria* spp. in apicoplast and mitochondrial genomes. Parasit Vectors. 2015;8:358. http://dx.doi.org/10.1186/ s13071-015-0966-3
- Cama VA, Bern C, Roberts J, Cabrera L, Sterling CR, Ortega Y, et al. *Cryptosporidium* species and subtypes and clinical manifestations in children, Peru. Emerg Infect Dis. 2008;14:1567– 74. http://dx.doi.org/10.3201/eid1410.071273
- Zhou Y, Lv B, Wang Q, Wang R, Jian F, Zhang L, et al. Prevalence and molecular characterization of *Cyclospora cayetanensis*, Henan, China. Emerg Infect Dis. 2011;17:1887–90. http://dx.doi.org/10.3201/eid1710.101296
- Mota P, Rauch CA, Edberg SC. Microsporidia and *Cyclospora*: epidemiology and assessment of risk from the environment. Crit Rev Microbiol. 2000;26:69–90. http://dx.doi.org/10.1080/10408410091154192

Address for correspondence: Yaoyu Feng, State Key Laboratory of Bioreactor Engineering, School of Resources and Environmental Engineering, East China University of Science and Technology, 130 Meilong Rd, Shanghai 200237, China; email: yyfeng@ecust.edu.cn

Letters

Letters commenting on recent articles as well as letters reporting cases, outbreaks, or original research are welcome. Letters commenting on articles should contain no more than 300 words and five references; they are more likely to be published if submitted within four weeks of the original article's publication. Letters reporting cases, outbreaks, or original research should contain no more than 800 words and ten references. They may have one Figure or Table and should not be divided into sections. All letters should contain material not previously published and include a word count.



Multilocus Sequence Typing Tool for *Cyclospora cayetanensis*

Technical Appendix

Technical Appendix Table 1. Primer sequences of additional microsatellite and minisatellite loci selected for initial analysis of *Cyclospora cayetanensis*

						Amplification
	Contig			Annealing	Expected	efficiency, no.
Locus	no.	Targeted repeat*	Primer sequence, $(5' \rightarrow 3')^{\dagger}$	temp, °C	size, bp	positive/no. analyzed
CYC7	00007	AC ₄₀	F1: CGATCGTATGGAGAGCTCCT; R1:	55	593	4/4
			AAGATGCCCTTCAGGCGGA; F2:			
			TTCACAATGTACGACTTGGCAC; R2:			
			GCGCATCATATACAGATGGTC			
CYC8	80000	ACCACC ₁₆	F1: ATTCCACAGTCAAGACTGCG; R1:	55	435	4/4
			CTGAGAGATGGTTTCTCTCCA; F2:			
			CTTCCTCGAGGGTATCTTCCAT; R2:			
			GCGCAGAGCTCTCTGCAGA			
CYC10	00010	TGTG ₆₂	F1: GTCGATTTCACAGTAGAGGGAT;	55	690	4/4
			R1: CTGCTTCGTGGCGAGCAAT; F2:			
			ATGCATGCGTGTTTAGGGCTT; R2:			
			ACAGCGAAATCATTTGCTGAGG			
CYC12	00012	ACTTCTTCTCC	F1: GAAGGTGCTACGGTGACAACT;	55	701	4/4
		TTCTGTTCT ₃₀	R1: TTGAAGCCTCGGATGAGACTT;			
			F2: CGGAGGCGCCAAAGTTGTCT; R2:			
			CTCCACTGTTCCAGCACC			
CYC14	00016	TG ₂₂	F1: TTGGTGTGTCGTCCCTACTAGA;	55	341	4/4
			R1: ATACAAAGACGACTAGCATGGC;			
			F2: CCATTCTGCCATCTAAGCATTAA;			
			R2:			
			AACGCATACTTCTGATTGAGGCAT			
CYC16	00018	TGC ₁₂	F1: CACATAAAGAAGCACTCAAGACG;	58	553	4/4
			R1: GGTCTCGCATTTAGACACTCG;			
			F2: CAACTCAGGACCCTTGTCGA; R2:			
			CCTCGTAGTAAAAGACCCTTGT			
CYC17	00020	AGC ₁₂	F1: AGCATTGAGCATCCCTCCTG; R1:	58	623	4/4
			TCCGCTTCCCATGCAGCTTC; F2:			
			CCTACAGTGTCAGCCGCTTC; R2:			
			ACAGTTGCATGCACATCAGTTC			
CYC18	00022	GTGTTTGTTT ₄₅	F1: GCATGCAACTTCTGGTGCATC;	55	947	4/4
			R1: AAGTGATTGCAGCAGCGAGGT;			
			F2: TGGAAGGGATGCTCGCTGC; R2:			
			CCCAAGTGCTGCAGCAATAA			
CYC19	00034	AG ₂₅	F1: TCTTAGAGCTCCCTAGTCGG; R1:	55	361	4/4
			CTCGTAGGCATCGACTGGAA; F2:			

						Amplification
	Contig			Annealing	Expected	efficiency, no.
Locus	no.	Targeted repeat*	Primer sequence, $(5' \rightarrow 3')^{\dagger}$	temp, °C	size, bp	positive/no. analyzed
			CGCACGCTGCTGGGTATG; R2:			
			TCTTTCTAAGCGCATGCTGTTAA			
CYC20	00035	CA ₁₃	F1: CTTGCTTCGATTTCCCTACGG;	55	421	4/4
			R1: CTGCAGGTTTACGAGCCTG; F2:			
			CATCGTCAGTCACTCGGAGTT; R2:			
			GATGGCTCGGCAACAAGTG			

*Tandem repeat identified in the sequence from whole genome sequencing.

†Six *C. cayetanensis* specimens were used in initial evaluation of PCR primers: specimens 22231, 22234, 22238, and 28709 were used to evaluate PCR primers from CYC7, CYC8, CYC10, and CYC12, whereas specimens 22231, 22234, 24550, and 24552 were used for the primers from the remaining loci.

0

					Sequence type	9		
Specimen	Host	Study	CYC3	CYC13	CYC15	CYC21	CYC22	MLST type
19134	Human	China, Henan, 2006	C1	C5	C1	C1	C3	MS1
22231	Human	China, Henan, 2007	C1	C1	C2	C1	C3	MS2
22232	Human	China, Henan, 2007	C1	C1	C1	C1	C3	MS3
22234	Human	China, Henan, 2007	C1	C1	C2	C1	C4	MS4
22236	Human	China, Henan, 2007	C1	C1	C1	C4	C3	MS5
22237	Human	China, Henan, 2007	C1	C6	C1	C1	C4	MS6
22238	Human	China, Henan, 2007	C1	C1	C1	C1	C3	MS3
22239	Human	China, Henan, 2007	C1	C1	C1	C1	C3	MS3
22240	Human	China, Henan, 2007	C4	C1	C1	C1	C3	MS7
22241	Human	China, Henan, 2007	C1	C1	C1	C3	C3	MS8
26245	Human	China, Henan, 2008	C1	C7	C1	C2	C3	MS9
28707	Human	China, Henan, 2009	Noisy	C1	C1	C4	C3	
28708	Human	China, Henan, 2009	Noisy	C1	C1	C4	Noisy	
28709	Human	China, Henan, 2009	C1	C2	C1	C2	C4	MS10
28710	Human	China, Henan, 2009	Noisy	C2	C1	C1	C4	
28711	Human	China, Henan, 2009	Noisy	C7	C1	C1	C4	
28712	Human	China, Henan, 2009	Noisy	C7	C1	C1	C3	
28713	Human	China, Henan,	C1	C7	C1	C4	C4	MS11
28715	Human	China, Henan,	C1	Noisy	C1	C3	C3	
28716	Human	China, Henan, 2009	C1	C5	C1	C1	C4	MS12
28717	Human	China, Henan, 2009	C1	C1	C1	C2	C4	MS13

Technical Appendix Table 2. Specimens used in the study and their sequence identity at the 5 selected loci

		Sequence type*						
Specimen	Host	Study	CYC3	CYC13	CYC15	CYC21	CYC22	MLST type
28718	Human	China, Henan,	C1	C4	C1	C4	C3	MS14
		2009						
28719	Human	China, Henan,	Noisy	C8	C1	C1	C3	
		2009			• ·			
28721	Human	China, Henan,	C1	C7	C1	C1	Noisy	
20722	Llumon	2009 China Uanan	Noiov	Naiov	C1	C1	<u></u>	
20122	Human		NOISY	INDISY	CI	64	03	
28723	Human	China Henan	Noisy	Noisy	C1	C4	C3	
20125	riaman	2009	INDISY	Noisy	01	04	05	
24550	Human	Peru. 2006	C1	C1	C1	C2	C1	MS15
24552	Human	Peru, 2006	C3	C3	C1	C2	C1	MS16
24556	Human	Peru, 2006	C3	C1	C1	C2	C1	MS17
24557	Human	Peru, 2006	C1	C1	C1	C2	C1	MS15
24560	Human	Peru, 2006	C3	C3	C1	C2	C3	MS18
24568	Human	Peru, 2006	_	C9	C1	C2	C1	
24578	Human	Peru, 2006	C3	Noisv	C1	C2	C1	
24600	Human	Peru, 2006	C1	C1	C1	C4	C1	MS19
1733	Human	United States.	C3	_	C1	C8	C2	
		Georgia, 1996			•			
24798	Human	United States.	C3	_	C1	_	C2	
200		Georgia 1996	•••		0.		-	
39948	Human	United States.	C2	Noisy	C1	C2	C1	
00010		Georgia, 1996	01	. teley	•	-	•	
24800	Human	United States	C3	C3	C1	C2	C1	MS16
24000	rianan	Rhode Island	00	00	01	02	01	Mere
		1997						
41398	Human	United States	C3	C3	C1	C2	C1	MS16
41000	rianan	Rhode Island	00	00	01	02	01	Mere
		1997						
24802	Human	United States	C3	C1	C1	C2	C1	MS17
24002	rianan	Virginia 1998	00	01	01	02	01	MOTI
39952	Human	United States	C2	C.4	C1	C2	C1	MS20
00002	riaman	Virginia 1998	02	04	01	02	01	1020
1731	Human	United States	C3	C1	C1	C3	C1	MS21
	rianan	New York	00	01	01	00	01	11021
		1998						
40064	Human	United States	Noisy	C1	C1	C2	Noisy	
10001	rianan	New York	Noloy	01	01	02	iteley	
		1998						
41399	Human	United States	Noisy	C3	C1	C2	Noisy	
		New York			•	-		
		1997						
41400	Human	United States.	C3	C3	C1	Noisv	C2	
		New York.	•••		•		-	
		1997						
41405	Human	United States.	C1	C1	C1	C2	C1	MS15
		Texas, 2013	-	-	-	-	-	
41406	Human	United States	C1	C1	C1	C2	C1	MS15
		Texas, 2013			•		•	
41407	Human	United States.	C1	C1	C1	C2	Noisv	
		Texas, 2013			•			
41408	Human	United States.	C1	C1	C1	C2	Noisv	
		Texas, 2013	-	-	-	-		
41409	Human	United States.	C1	C1	C1	C2	Noisv	
		Texas. 2013		2.				
41410	Human	United States.	C3	C1	C1	C2	Noisv	
		Texas. 2013		2.				
41411	Human	United States	C1	C9	C1	C5	Noisv	
		Texas, 2013	-		-			
41412	Human	United States.	C1	C9	C1	C5	Noisv	
		Texas, 2013					,	

					Sequence type	9*		
Specimen	Host	Study	CYC3	CYC13	CYC15	CYC21	CYC22	MLST type
41414	Human	United States,	C1	C1	C1	C2	Noisy	
		Texas, 2013						
41397	Human	United States,	C3	C3	C1	C2	C2	MS22
		Florida, 1996						
41401	Human	United States,	C3	C1	C1	C2	C1	MS17
		Nevada, 1997						
41402	Human	United States,	C3	C10	C1	Noisy	C2	
		Pennsylvania,						
		2000						
41395	Human	Nepal, 1997	C1	C1	C1	Noisy	C1	
39951	Human	Nepal, 1997	C3	C4	C1	C2	C1	MS23
24799	Human	Nepal, 1997	C2	C1	C1	C6	C1	MS24
39955	Human	Indonesia,	C3	C1	C1	Noisy	C1	
		2012						
33354	Human	Spain, 2011	C2	C1	C1	C7	C2	MS25
24801	Human	Guatemala,	C3	_	_	_	C1	
		1997						
1732	Human	Guatemala,	C2	C3	Noisy	C2	C1	
		1997						

*Dashes indicate no PCR amplification; "noisy" represents unreadable sequence. MLST, multilocus sequence typing.

		10	20	30	40	50	60	70	80	90
		.			.				.	
28716-C1	TGCCTGCC	CTCCTTGCA	FTCCTTGCAT	rectrigettee	CTGCAGCGCG	TGTCTGTGTG	JTATGTATGTA	TGTATGTATO	TGTCTGTATC	TATG
28715-C1		•••••		• • • • • • • • • • • •		********				
22239-C1										• • • •
22240-C4				A						
24799-C2										
33354-C2										
39952-C2										
24798-C3							G	.C.G.A.G.A	G	
24801-C3							G	.C.G.A.G.A	G	
24560-C3						c	G	.C.G.A.G.A	G	
		110	120	130	140	150	160	170	180	190
28716-C1	TCTGTATG	TGTGTATGT.	ATGTGTCTGT.	ATGTATGTGTC	TGTATGTGTG	TGTATGTGT	CTGTATGTGTC	TGTATGTAT	TGTGTCTGT/	ATGTG
28715-C1										
22239-C1										
22240-C4										
24799-C2										
33354-C2										
39952-C2										
24798-C3	.A	c	G.A.	G.A.G.AG	A.C	G.A	AG.AA	GA	4G.C.C	3.A
24801-C3	.A	c	G.A.	G.A.G.AG	A.C	G.A	AG.AA		G.C.C	3.A
24560-C3	.A	c	G.A.	G.A.G.AG	A.C	9G.A)	AG.AA	G	4G.C.C	3.A
		210	220 	230	240	250 	260 	270	280	290

						200 200		
		.						
28716-C1	TGTGTATG	TATGTGTCT	GTATGTGTG	TATGTATATGT	TATATGTAGAT	ATATATATGTATGTAT	ATATATATATA	TGTATGTATATATATA
28715-C1								
22239-C1								
22240-C4								
24799-C2							TA	
33354-C2							T A	
39952-C2							ТА	
24798-C3				A.		. G	G	
24801-C3				A.		G	G	
24560-C3				A.			G	

Technical Appendix Figure 1. Variations in nucleotide sequences among Cyclospora cayetanensis

specimens at the CYC3 locus. Dots denote nucleotides identical to those in the first sequence of the

sequence alignment, and dashes denote nucleotide deletions.

	220	320	330	340	350	360	370	380	390
		-		.					
22231-C1	CGAGGAGGCAGACA	AGATGATGA	TGACGATGA	TGATGACAGA	CTTGATGGCC	TCTCAGATCT	TGACGGTTT	AGGCGAGGAT	GATGAT
28710-C2	•••••							0	3
24552-C3						• • • • • • • • • • •			
39951-C4									
19134-C5		C						· · · · · · · · · · · · · · · · · · ·	3
22237-C6		C							3
26245-C7									3
28719-C8									
24568-C9									
41402-C10	•••••	c	•••••	••••••	••••••		•••••••••		
	410	420	430	440	450	460	470	480	490
00001 01	CARCARCARCARCARCI								nmm a m a m
22231-01	CATOATOATOATOATO	ATOATOATOAT	GCAGAI	I CAAAT GACG	ACGAGGAATA	TOOCOGIACO	ACCICICACO	JOGIOCACAI	TIMIMI
20710-02				••••••					
24052-03									
10124 05				••••••••••••					
19134-05									
22237-00				••••••					
26245-07		• • • • • • • • • • • •		••••••				•••••••••	
28/19-08	••••••	• • • • • • • • • • • • •	GAT	•••••••				• • • • • • • • • • •	
24568-09				• • • • • • • • • • •		•••••		• • • • • • • • • • •	
41402-CIU	·····C·····			••••••	.T	••••••		• • • • • • • • • • • •	•••••
	510	520	530	540	550				
				.					
22231-C1	TTTACATTTACA	TATATATATAT	ATTTATTA'	TGTTTATATO	TTGGACGGAC				
28710-C2	T	G		• • • • • • • • • • •	•••••				
24552-C3	TA								
39951-C4									
19134-C5	TA			.AA					
22237-C6	TA			.AA					
26245-C7				.AA					
28719-C8									
24568-C9	ТА								
41402-C10	TATA			.AA					

Technical Appendix Figure 2. Variations in nucleotide sequences among *Cyclospora cayetanensis* specimens at the CYC13 locus. Dots denote nucleotides identical to those in the first sequence of the sequence alignment, and dashes denote nucleotide deletions.

	210	220	230	240	250	260	270	280	290
OAFET OF	CA TO CO CO CA CO					000000000000000000000000000000000000000	magagammamm	ammammaama	
24337-CI	CATCAGCGAG	GGCAGC I GAU	SCAGICGCIIG	199919CIGCIGCI	Gendendenden	GCIGCIGCIGC	IGCCGIIGII	GIIGIIGCIC	SCIGCICC
33354-C1			• • • • • • • • • • • • •	• • • • • • • • • • • •	• • • • • • • • • • • • • • •				
28722-C1									
22231-C2			.T						
22234-C2			.T						

Technical Appendix Figure 3. Variations in nucleotide sequences among *Cyclospora cayetanensis* specimens at the CYC15 locus. Dots denote nucleotides identical to those in the first sequence of the sequence alignment, and dashes denote nucleotide deletions.

	10	110	120	130	140	150	160	170	180
			.						
22232-C1	GGCAAACGTAGG	ATCTATA	TATATTTATTA	TATATTTATTA	TATATT	TAT-ATATAT	ATATATAT		TTATGG.
24550-C2							AT		
28715-C3							AT	'AT	
28718-C4	C	A			CTTA	ATAT . TA	T TT	ATATATATAT	
41411-C5							AT	'ATAT	
24799-C6							AT	ATATATATAT	
33354-C7		ACTATCT.	A.T.AT	G.T.AT	TATATA				
1733-C8		ACTATCT.	A.T.AT	ACT . AT	TATATA	T	T AT	'AT	
	210	220	230	310	320	330 • • • • • • • •	340	350	360
22232-C1	CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	CAACGCCGGTACGTC	AGTGGGTT	CAGTAGCAGC	AGCAGATCTA	GCCACTGCTG	CTGCTGCGAC	TTGCGGAGAA	CACGCA
24550-C2									
28715-C3									
28718-C4		T		G		A		A	
41411-C5									
24799-C6									
33354-C7						A		A	
1733-C8						A			

Technical Appendix Figure 4. Variations in nucleotide sequences among *Cyclospora cayetanensis* specimens at the CYC21 locus. Dots denote nucleotides identical to those in the first sequence of the sequence alignment, and dashes denote nucleotide deletions.

	60	70	80	90	100	110	120	130	140
24552-C1	GCACATCAGGGT	CCAGACAAAAGG	GGAGATGGAG	CACCTAGGA-	GTC	GCGTAAATA	ACTCCCAAAC	CACTTCGGGG	CATCACAG
24600-C1									
41397-C2	C.						T		
33354-C2	C.						T		
28718-C3				C	ACCTAGGA		T	C	T
28715-C3				C	ACCTAGGA		T	C	T
28709-C4				C	ACCTAGGA		T	C	T
28716-C4				C	ACCTAGGA		T	c	T
	160	170	180	190	200	210	220	230	240
24552-C1	ACACAGCATATA	ACACACACACACA	ACA			GCATTGAGC	AGTTTCGTGT"	FGAGCTAATT	GTACCCTA
24600-C1									
41397-C2			CACACA						
33354-C2			CACACA-						
28718-C3			CACACAC	ACACACACAC	ACACACACA				
28715-C3			CACACAC	ACACACACAC	ACACACACA				
28709-C4			CACACAC	ACACACACAC	ACACACACACA				
28716-C4			CACACAC	ACACACACAC	ACACACACAC	A			

Technical Appendix Figure 5. Variations in nucleotide sequences among *Cyclospora cayetanensis* specimens at the CYC22 locus. Dots denote nucleotides identical to those in the first sequence of the sequence alignment, and dashes denote nucleotide deletions.