INFECTION AND IMMUNITY, Feb. 2007, p. 846–851 Vol. 75, No. 2
0019-9567/07/$08.00+0 doi:10.1128/IAI.01205-06
Copyright © 2007, American Society for Microbiology. All Rights Reserved.

Bioinformatic Identification of Tandem Repeat Antigens of the *Leishmania donovani* Complex[▽]

Yasuyuki Goto, Rhea N. Coler, and Steven G. Reed*

Infectious Disease Research Institute, Seattle, Washington 98104

Received 31 July 2006/Returned for modification 8 September 2006/Accepted 30 October 2006

With large amounts of parasite gene sequence available, additional bioinformatic tools to screen these sequences for identifying genes encoding antigens are needed. Proteins containing tandem repeat (TR) domains are often B-cell antigens, and antibody responses toward TR domains of the proteins are dominant in human infected with certain parasites. We hypothesized that antigens of serological significance could be identified with a search for TR domains. Here we show the result of bioinformatic screening of the gene sequence database of the parasitic protozoan *Leishmania infantum*. Of 8,191 genes scanned, 64 genes contained TR domains. Of the 64 genes, 22 encoded previously characterized antigens; the remaining 42 genes were previously uncharacterized. By using sera from Sudanese visceral leishmaniasis patients, we confirmed that the TR domains of LinJ11.0070, LinJ25.1100, LinJ27.0400, and LinJ29.0110, which were from the 42 uncharacterized proteins, are also antigenic. The results suggest the validity of this approach for identifying leishmanial antigens of serological significance.

Parasitic protozoa, such as the causative agents for leishmaniasis, malaria, and trypanosomiasis, are important human pathogens. Among the diseases caused by *Leishmania* is a severe form known as visceral leishmaniasis (VL). Diagnostic methods for human VL often rely on detection of parasite-specific antibodies (27, 30, 34). Among defined leishmanial antigens reported previously, rK39 (7) is the most widely antigen for serodiagnosis of VL in terms of both sensitivity and specificity, particularly in Brazil, India, and Nepal (3, 33, 35). However, new diagnostic antigens are needed to complement rK39 for developing more sensitive diagnostics for VL, particularly in Africa (37).

Proteins containing tandem repeats (TR) are known targets of B-cell responses (21, 28). Genes encoding proteins with TR, consisting of two or more copies of a pattern of nucleotides, have been found in many protozoan parasites, usually by expression cloning methods, although no systematic search for TR-containing proteins has been reported. Antibody responses toward the encoded TR domains have been found in various parasitic diseases such as leishmaniasis, malaria, and Chagas' disease (5, 7–11, 16, 17, 19, 22, 29, 32, 36).

In a previous study, we have found that serological screening of a DNA library revealed a disproportional number of serological antigens containing TR (16). Because dominant antigens often contain TR domains, we hypothesized that a bioinformatic approach to identify TR proteins according to their sequences could be useful for antigen identification. In the present study, we computationally searched the database of *L. infantum*, one of the causative agents of VL, resulting in the identification of 64 TR genes from 8,191 genes analyzed. These 64 genes contained 22 genes encoding previously characterized antigens; the remaining 42 genes were previously uncharacterized. Furthermore, we confirmed that VL patient sera recognized some of the novel TR proteins. Taken together, the results shown here suggest that *L. infantum* TR proteins may be antigenic and that a bioinformatic approach to discover TR proteins is useful for identifying such antigens.

MATERIALS AND METHODS

Bioinformatic screening of TR genes. For comparative purposes, we analyzed available DNA sequence data of *L. major* (20), *L. infantum*, *Trypanosoma brucei* (4), *Plasmodium falciparum* (14), and *Theileria annulata* (25) obtained from GeneDB (http://www.genedb.org/) (18). Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.html), a program to locate and display TR in DNA sequences, was used for this analysis (2). The program calculates the score according to the characteristics of the TR genes such as the period size of the repeat, the number of copies aligned with the consensus pattern, and the percentage of matches between adjacent copies overall. A high score indicates that the gene possesses a large TR sequence and that the repeat is highly conserved among the copies. For example, a gene with 10 copies of a 30-bp repeat and a gene with 5 copies of a 60-bp repeat, both of which have a 300-bp TR domain, have a score of 600 ($=300 \times 2$). In the present study, the genes were regarded as TR genes if the scores from the Tandem Repeats Finder analysis were 500 or higher. The cutoff value of 500 is likely to eliminate genes with repeat domains whose sizes are less than 250 bp. When more than one TR domain was found within a gene, only the domain with the highest score was listed or used for further analyses and protein production. Spliced DNA sequences were used for the analysis in order to ensure that the nucleotide repeats found are likely to reflect repeats in peptide sequence.

Expression of recombinant proteins. Cloning of TR domains of LinJ11.0070, LinJ21.2010, LinJ25.1100, LinJ27.0400, LinJ29.0110, and LinJ32.3710, and expression and purification of the encoded proteins were performed as described previously (16). In brief, sequences encoding whole or partial TR domains were amplified by PCR with *L. infantum* total DNA using primer sets as following, LinJ11.0070, 5′-CAA TTA CAT ATG CTC CGC CAC CAG CTG GCC and 3′-CAA TTA AAG CTT CTA CTG CTC CAG CTC CTC TGC; LinJ25.1100, 5′-CAA TTA CAT ATG GAG GAC ACG AGG ATA ACC and 3′-CAA TTA AAG CTT CTA TTC AGG CTC CTC GGC TGA C; LinJ27.0400, 5′-CAA TTA CAT ATG CGC GCG CAC GAC CTT GCG and 3′-CAA TTA AAG CTT CTA GTC GTT CAT CCT CCT CTC; and LinJ29.0110, 5′-CAA TTA CAT ATG GAG ATT CAA GCG CTA CGC and 3′-CAA TTA AAG CTT CTA AAC CTC CTC CAG ACC ACC. Parasites were dissolved in Tris-EDTA buffer containing 1% sodium dodecyl sulfate, and the total DNA was purified by phenol-chloroform purification following sequential RNase and proteinase K treatment for use as a template for PCRs. The amplified PCR products were inserted in-frame with

* Corresponding author. Mailing address: Infectious Disease Research Institute, 1124 Columbia St, Suite 400, Seattle, WA 98104. Phone: (206) 330-2520. Fax: (206) 381-3678. E-mail: sreed@idri.org.

▽ Published ahead of print on 6 November 2006.

TABLE 1. Numbers of TR genes in protozoan parasites

Parasite species	No. of genes tested	No. of TR genes (%)[a]	TR score (%)[b]				
			500–1000	1000–1999	2000–49999	5000–9999	≥10000
L. infantum	8,191	64 (0.78)	10 (16)	17 (27)	20 (31)	11 (17)	6 (9)
L. major	9,218	59 (0.64)	15 (25)	16 (27)	15 (25)	10 (17)	3 (5)
T. brucei	10,955	73 (0.67)	14 (19)	19 (26)	24 (33)	8 (11)	8 (11)
P. falciparum	5,513	169 (3.07)	130 (77)	29 (17)	8 (5)	2 (1)	0 (0)
T. annulata	3,795	11 (0.29)	9 (82)	1 (9)	1 (9)	0 (0)	0 (0)

[a] The percentages in this column represent the ratio of the number of TR genes identified to the number of genes tested.
[b] The identified TR genes were sorted according to the TR scores. The percentages represent the ratio of the number of TR genes in each range to the number of total TR genes identified.

a His_6 tag into the vector pET-28a, and sequences of the inserts were confirmed against the *L. infantum* GeneDB database. The vectors were then transformed into *Escherichia coli*, and the recombinant proteins were purified as His_6-tagged proteins.

ELISA. The expressed TR-containing proteins were analyzed for seroreactivity using panels of patient and control sera. *L. infantum* soluble lysate antigen (SLA) was used as a positive control and a *Mycobacterium leprae* antigen ML2331 was used as an irrelevant antigen (26). Proteins were diluted in an enzyme-linked immunosorbent assay (ELISA) coating buffer, and 96-well plates were coated with 1 μg of *L. infantum* SLA or 200 ng of individual recombinant antigens, followed by blocking with phosphate-buffered saline containing 0.05% Tween 20 and 1% bovine serum albumin. Plates were incubated with VL patient sera (n = 16, tested individually, not pooled, human immunodeficiency virus negative), as well as sera from healthy donors in the United States (n = 8) at a 1:100 dilution and then with horseradish peroxidase-conjugated anti-human immunoglobulin G (Rockland Immunochemicals, Inc., Gilbertsville, PA). The plates were developed with tetramethylbenzidine peroxidase substrate (Kirkegaard & Perry Laboratories, Gaithersburg, MD) and read by a microplate reader at 450 nm (570-nm reference).

Analysis of amino acid compositions of TR proteins. The *L. infantum* TR proteins were analyzed to determine their isoelectric points (pIs) by using the EditSeq software package (DNASTAR, Inc., Madison, WI). As a control, 108 genes, randomly selected from the *L. infantum* gene database, were also analyzed for the pI and amino acid compositions of their deduced amino acid sequences. TR proteins were further analyzed for amino acid composition in the whole proteins, TR domains, and non-TR regions by using the EditSeq.

RESULTS

Identification of TR genes from a *L. infantum* gene database. The database used contained a number of putative genes on which we performed analyses. Of 8,191 *L. infantum* gene sequences analyzed by Tandem Repeats Finder, 64 genes (0.78%) were identified as genes containing TR regions based on an arbitrary cutoff score of 500 (Table 1). The ratio of TR genes is similar to that observed in *L. major* (59 of 9,218 [0.64%]) and *Trypanosoma brucei* (73 of 10,955 [0.67%]). The *Plasmodium falciparum* genome is rich in TR genes (169 of 5,513 [3.07%]), whereas *Theileria annulata* has only 11 TR genes (11 of 3,795 [0.29%]).

When these selected TR genes were sorted by their TR scores, the trypanosomatid and the apicomplexa showed different patterns. *P. falciparum* and *T. annulata* were rich in TR genes with TR scores of <1000 (Table 1). In contrast, *L. major*, *L. infantum*, and *T. brucei* were rich in large TR genes, with their peaks between scores 2000 and 4999. Although the number of total TR genes was greater in *P. falciparum*, TR genes with a TR score of 2000 or higher were found more in *L. infantum* (39 and 10 in *L. infantum* and *P. falciparum*, respectively).

These 64 genes included 5 genes encoding the previously well-characterized antigens, K26, K39, A2, and Lt-1 (5, 7, 13, 15), as well as 17 genes also identified by serological screening in our recent study (16) (Table 2). The remaining 42 genes, however, were previously uncharacterized. Molecular masses of the TR proteins were 180 kDa in average, ranging from 24 to 687 kDa. Individual copy of the repeats ranged in size from 6 to 483 bp (2 to 161 amino acids [aa]). The repeat of each TR gene was highly conserved among copies, 95% on average ranging from 75 to 99% in nucleotide sequence, and more highly in amino acid sequence identity. These TR genes were found on 26 of 36 chromosomes, and the highest number of TR genes was found on chromosomes 14, 22, and 35. A number of putative genes from the database either did not have start or stop codons or had stop codons within the genes. These are shown as "incomplete" genes in Table 2. The other genes, which have both start and stop codons, are shown as "complete" genes. Of 64 genes identified, 50 were complete and 14 were incomplete.

Recognition of *Leishmania* TR proteins by Sudanese VL patient sera. Although some proteins containing TR, which we identified by the computational screening, were antigens previously identified by serological screening, this did not guarantee the antigenicity of the remaining, previously uncharacterized TR proteins. We next examined the antigenicity of TR proteins previously uncharacterized and identified solely from the computational screening. Because the TR domains are often B-cell epitopes, we focused on the TR regions of these genes instead of entire open reading frames. Of the 42 previously unidentified genes, 10 were incomplete genes and were excluded from the list of proteins to be pursued. LinJ09.0950 (polyubiquitin) showed similarity to ubiquitin in mammals and was excluded from further study. Of the remaining 31 complete genes, some had very large TR domains which were not practical to clone in full. Also, it was difficult to sequence the cloned TR if larger than 1 kb because internal primers could match with multiple sites within the repeats.

Thus, we cloned entire TR regions if they were smaller than 1 kb and the partial TR regions if they were larger than 1 kb. For cloning of TR of less than 1 kb, primers matching with sequences flanking outside the TR domain were used for PCR. In this case, a single band was expected for each gene. For cloning of TR of more than 1 kb, primers matching with both ends of the TR were used for PCR by which ladder bands corresponding to a single or multiple repeats were amplified. To avoid losing possible epitope(s) which may lie between repeats, a band corresponding to not a single repeat but multiple copies of TR was used for cloning. If one copy of TR was

TABLE 2. *L. infantum* TR genes identified by the Tandem Repeats Finder[a]

Gene ID[b]	C/I[c]	Product	Size (kDa)	PS (bp)	CN	Score[d]	Reference[e]
LinJ03.0120	C	Hypothetical protein	237	117	31.8	7033	16
LinJ05.0340	C	Viscerotropic leishmaniasis antigen	95	99	13.8	2545	13
LinJ05.0380	C	Microtubule-associated protein	165	114	28.5	6336	16
LinJ09.0950	C	Polyubiquitin	74	228	8.0	3621	
LinJ11.0070	C	Hypothetical protein	147	138	12.9	2435	
LinJ13.0780	C	Hypothetical protein	107	63	14.2	1637	
LinJ14.0370	C	Hypothetical protein	302	84	10.9	1475	
LinJ14.1160	C	Kinesin K39	242	117	27.9	5237	7
LinJ14.1180	I	Kinesin K39		168	8.2	2671	
LinJ14.1190	I	Kinesin K39		315	6.1	2828	16
LinJ14.1200	C	Kinesin K39	79	468	3.4	1971	7
LinJ14.1210	I	Kinesin K39		483	10.9	3676	
LinJ14.1540	C	Hypothetical protein	112	72	6.1	806	16
LinJ15.0490	I	Tb-292 membrane-associated protein-like protein		105	31.6	6027	16
LinJ15.1570	I			105	29.9	5588	
LinJ16.1540	C	Kinesin	230	42	138.5	10588	16
LinJ16.1750	C	Hypothetical protein	346	219	8.7	3691	16
LinJ18.1030	C	Hypothetical repeat protein	46	21	30.4	1036	
LinJ19.0940	C		24	6	95.0	1076	
LinJ19.1560	I			81	21.1	3094	
LinJ20.1220	C	Calpain-like cysteine peptidase	112	39	11.3	826	
LinJ21.2010	C	Hypothetical protein	306	192	5.3	2003	
LinJ22.0410	C	Hypothetical protein	130	183	15.9	5779	
LinJ22.0680	C	Hypothetical protein	45	216	5.9	1240	15
LinJ22.1510	C	Hypothetical protein	179	81	13.5	1984	
LinJ22.1520	C		72	39	42.9	3197	
LinJ22.1550	C		126	81	10.4	1504	
LinJ22.1560	I			267	16.9	8614	
LinJ22.1570	C		210	81	23.5	3230	
LinJ22.1580	C		175	267	17.1	8591	
LinJ22.1590	C	Hypothetical protein	234	84	29.2	3993	16
LinJ23.1180	C	Hydrophilic surface protein	26	42	11.2	832	5
LinJ25.1100	C	Hypothetical protein	91	66	9.5	1142	
LinJ25.1910	C	Hypothetical protein	91	369	2.0	1443	
LinJ26.2140	C	Hypothetical protein	215	48	63.4	5289	
LinJ27.0140	I	Kinetoplast-associated protein-like protein		30	19.9	1086	
LinJ27.0170	C	Kinetoplast-associated protein-like protein	95	30	62.1	3283	
LinJ27.0400	C	Calpain-like cysteine peptidase	687	204	43.8	17362	
LinJ28.2310	C	Glycoprotein 96-92	61	315	2.2	1398	16
LinJ28.3170	C	Hypothetical protein	75	60	23.4	2546	16
LinJ29.0110	C	Hypothetical protein	278	24	28.6	967	
LinJ30.0400	C	Hypothetical protein	56	117	7.4	1716	
LinJ31.1820	C	Hypothetical protein	49	75	4.1	581	
LinJ31.1840	C	Hypothetical protein	52	24	18.1	814	
LinJ31.2660	C	Hypothetical protein	247	456	2.2	1973	
LinJ31.3360	C	Hypothetical protein	71	30	11.1	556	
LinJ32.2730	C	Hypothetical protein	173	150	10.3	2916	16
LinJ32.2780	C	Membrane associated protein-like protein	132	30	60.9	3125	16
LinJ32.3710	C	Hypothetical protein	292	99	3.9	730	
LinJ33.2870	C	Hypothetical protein	413	444	7.0	6041	16
LinJ34.0710	I	Hypothetical protein		336	9.5	4517	16
LinJ34.2140	C	Hypothetical protein	296	249	7.4	3604	16
LinJ34.4250	C	Hypothetical protein	168	168	6.1	1960	
LinJ35.0590	C	Proteophosphoglycan ppg4	536	45	246.1	10667	16
LinJ35.0600	I	Proteophosphoglycan ppg3		135	37.8	8773	16
LinJ35.0610	C	Proteophosphoglycan ppg4	291	45	183.2	13275	
LinJ35.0620	I	Proteophosphoglycan 5		90	152.5	15050	
LinJ35.0630	I	Proteophosphoglycan ppg4		45	176.6	10813	
LinJ35.0640	I	Hypothetical protein		45	58.4	4766	
LinJ35.1530	C	Hypothetical protein	328	141	2.4	661	
LinJ35.1620	I	Hypothetical protein		126	8.7	1855	
LinJ35.4500	C	Hypothetical protein	60	165	4.5	1438	
LinJ36.0320	C	Histidine secretory acid phosphatase	71	72	6.5	861	
LinJ36.5810	C	Hypothetical protein	365	276	4.3	2341	

[a] Data for the number of copies aligned with the consensus pattern (CN), the period size of the repeat (PS), and the score are from a program analysis using the Tandem Repeats Finder.
[b] Identification (ID) numbers in GeneDB are temporary and may vary.
[c] C, complete gene; I, incomplete gene. See Results.
[d] Genes with a TR score of 500 or higher are listed.
[e] The antigenicities of the proteins were reported in the indicated references.

small, 60 bp or less, the TR is not suitable to be cloned by PCR with primers matching both ends of the TR. Thus, TR genes with more than 1 kb of TR domain and 60 bp or less of TR unit, such as LinJ22.1520, LinJ26.2140, and LinJ35.0590, were excluded. Based on these selections, 19 individual genes were chosen for cloning by PCR. Of the 19 genes, 12 of them were successfully cloned by PCR amplification. Of these, six (LinJ13.0780, LinJ20.1220, LinJ22.1510, LinJ22.1570, LinJ31.1820, and LinJ36.0320) did not express in *E. coli*. For these reasons, we chose six TR proteins for a further serological study.

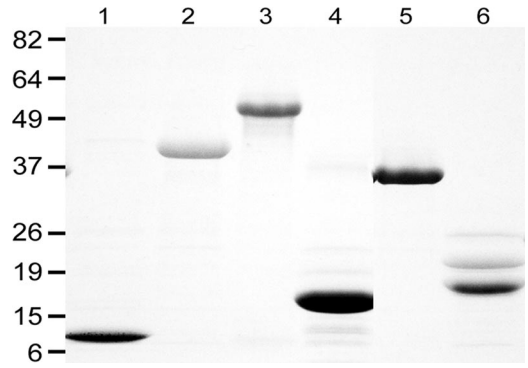

FIG. 1. Recombinant *L. infantum* TR proteins. Lane 1, rLinJ11.0070r2; lane 2, rLinJ21.2010TR; lane 3, rLinJ25.1100TR; lane 4, rLinJ27.0400r2; lane 5, rLinJ29.29.0110TR; lane 6, LinJ32.3710TR. Sizes are shown in kilodaltons on the left.

By sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis, rLinJ11.0070r2 (with 2 copies of 46 aa), rLinJ21.2010TR (with 5.3 copies of 64 aa), rLinJ27.0400r2 (with 2 copies of 68 aa), rLinJ29.29.0110TR (with 28.8 copies of 8 aa), and LinJ32.3710TR (with 3.9 copies of 33 aa) showed apparent molecular masses similar to those expected (12, 38, 18, 31, and 17 kDa, respectively; Fig. 1). The apparent molecular mass of rLinJ25.1100TR (9.6 copies of 22 aa) was around 54 kDa and was larger than the expected size (27 kDa).

We then examined the presence of antibodies in Sudanese VL patient sera to these TR proteins. Two, rLinJ27.0400r2 and rLinJ29.0110TR, showed good reactivity to the VL patient sera with higher peak responses than that of *L. infantum* SLA (Fig. 2). rLinJ11.0070r2 and rLinJ25.1100TR showed intermediate reactivity to the VL patient sera; none of the four antigens were recognized by sera from healthy donors. VL patient sera showed only a weak antibody response to an irrelevant *Mycobacterium leprae* antigen ML2331 (26). Compared to the reactivity of the irrelevant antigen, rLinJ11.0070r2, rLinJ25.1100TR, rLinJ27.0400r2, and rLinJ29.29.0110TR, as well as *L. infantum* SLA, showed significantly stronger reactivity to the VL patient sera, whereas rLinJ21.2010TR or LinJ32.3710TR did not detect VL-specific antibodies ($P < 0.05$ on rLinJ11.0070r2, $P < 0.01$ on rLinJ25.1100TR, and $P < 0.001$ on rLinJ27.0400r2, rLinJ29.29.0110TR, and *L. infantum* SLA by unpaired *t* test).

Abundance of strongly acidic amino acids in TR domains. Since a number of TR domains of *L. infantum* TR proteins, including those in the present study, have been found to be recognized by VL patient sera, we sought characteristics of the TR domains. The 50 "complete" TR genes in Table 2 were analyzed for the isoelectric point (pI) of their deduced amino acid sequences and compared to those of *L. infantum* proteins randomly selected from the database. Randomly selected proteins showed various pIs with a normal distribution (according to the KS normality test), 7.7 as the mean pI (with a 95% confidence interval of 7.3 to 8.0), which is close to the physiological pH (Fig. 3). In contrast, the pIs of TR proteins showed dichotomous distribution. The mean pI of the 50 "complete" TR proteins was 6.0, which is statistically lower than that of the randomly selected proteins ($P < 0.0001$ according to the Mann-Whitney test). The 50 "complete" TR proteins contained putative proteins whose expression or antigenicity has not been characterized. When 22 TR proteins, including 18 identified in previous studies (see references in Table 2) and 4 whose antigenicities were characterized in the present study (i.e., rLinJ11.0070r2, rLinJ25.1100TR, rLinJ27.0400r2, and rLinJ29.29.0110TR), were analyzed, the mean pI was 5.5, which is statistically lower than that of the randomly selected proteins ($P < 0.0001$ [Mann-Whitney test]), whereas no difference was observed compared to the 50 "complete" TR proteins. A total of 37% (40 of 108) of the randomly selected proteins were acidic, with pIs of <7, whereas most of the antigenic TR proteins were acidic (19 of 22 [86%]).

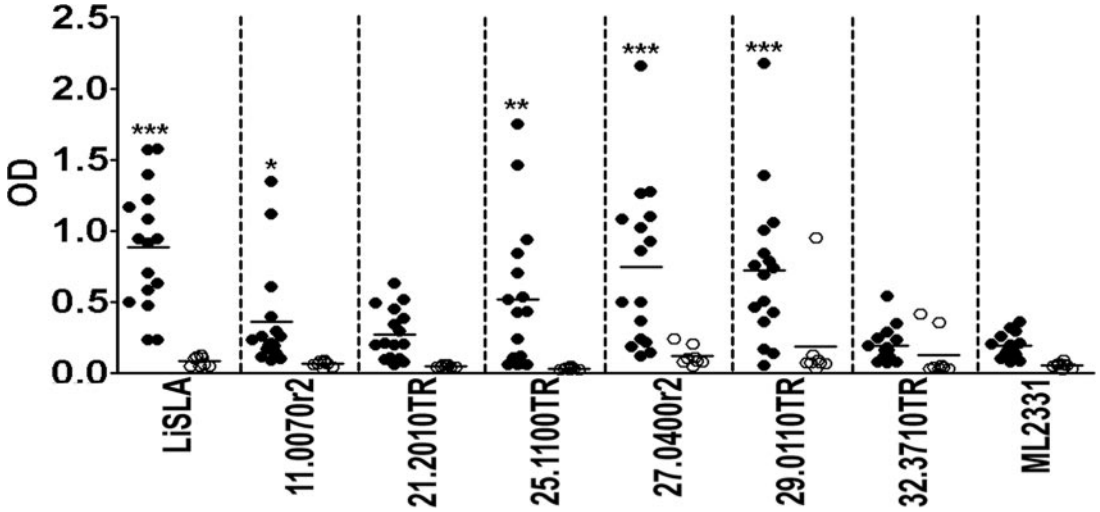

FIG. 2. Antibody responses of VL patient sera to TR proteins identified by bioinformatic screening. Sera from VL patients (●; $n = 16$) and healthy donors (○; $n = 8$) were tested individually by ELISA, and optical density values for each individual are shown. Bars represent means of each group. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ (as determined by unpaired *t* tests between reactivity of VL patient sera to leishmanial antigens and to ML2331).

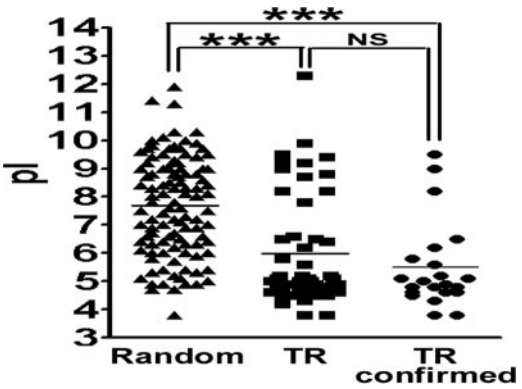

FIG. 3. Isoelectric points of *Leishmania* TR proteins. Isoelectric points of proteins randomly selected from the database, with all 50 TR proteins shown as "complete" in Table 2 or 22 TR proteins with confirmed antigenicity are shown. NS, not significant; ***, $P < 0.001$ (as determined by Mann-Whitney tests).

DISCUSSION

Although antigenic TR proteins have been identified in protozoan parasites, no systematic bioinformatic approach to identify and characterize such proteins has been reported. Therefore, we approached antigen identification by computational screening of TR proteins, focusing especially on *Leishmania*. In the present study, 64 of 8,191 *L. infantum* genes (0.78%) were identified as containing TR domains. In a previous study, we identified 43 genes encoding antigenic proteins by serological screening, 19 of which (44%) contained TR (16). This indicates the potency of TR proteins as antigens recognized by patient sera. In addition, 64 genes identified in the present study included 22 genes previously characterized as coding for antigens. We identified, through bioinformatic analysis of TR domains, previously uncharacterized antigens with serodiagnostic potential. Taken together, these results demonstrate the usefulness of the bioinformatic analysis for finding parasite antigens.

This screening approach may be applicable to other protozoan parasites such as *Plasmodium* and *Trypanosoma*. Indeed, we found genes encoding previously characterized TR antigens such as *Plasmodium* CSP, FIRA, RESA, and S antigen (9–11, 32) by screening the parasite database using the Tandem Repeats Finder. Although we did not test the antigenicity of *Plasmodium* or *Trypanosoma* TR proteins found using this bioinformatic method but which had not been characterized previously, the data on *Leishmania* suggest the potential antigenicity of those as well. Furthermore, it is of interest that some cancer antigens to which patients show antibody responses contain TR domains (23, 24), suggesting that TR domains tend to be antigenic despite the origin.

With the exception of peptide epitope prediction, there have been a limited number of bioinformatic approaches to antigen discovery. One approach has been to identify sequences likely to encode secreted or surface proteins (1, 6). However, this approach has not led to the discovery of the most effective antigens. For example, rK39, the best diagnostic antigen of VL, is a kinesin-related protein, which does not have predicted signal sequences or transmembrane domains. The results in the present study suggest that our unique computational approach can be very useful to complement existing screening methods, including serological expression cloning to find antigens.

TR domains of *L. infantum* proteins could be highly antigenic for a variety of reasons. The existence of multiple copies of antigenic units may result in increased exposure to the immune system. Besides that, in the present study we have identified the tendency of *L. infantum* TR proteins to possess charges. Charged (hydrophilic) proteins are likely to be more potent as B-cell antigens than hydrophobic proteins. In fact, most of previously reported antigens of *L. donovani* complex, not only TR proteins but also non-TR proteins such as acidic ribosomal proteins or heat shock proteins (12, 31), are highly charged. TR domains seem to contribute to the acidic or basic character of the proteins, since there is a higher prevalence of strongly charged amino acids (D, E, K, and R) in the TR domains than in the non-TR domains (data not shown). These two factors, repetition and hydrophilicity, may explain the antigenicity of the TR domains.

It is intriguing that trypanosomatid parasites, which include *Leishmania* and *Trypanosoma* species, are rich in relatively large TR genes compared to the apicomplexa, which include the malarial parasites *Plasmodium*, even though a large amount of nucleotide repeats are found in both of these parasite groups in the genomic DNA sequence. In contrast to *Leishmania*, *P. falciparum* is rich in a large number of small TR genes. When the cutoff value of the TR score was decreased to 150 instead of 500, 1,316 of 5,513 *P. falciparum* genes would be regarded as TR genes versus only 99 in *L. infantum* (data not shown). Exon-intron splicing often occurs in the apicomplexa, which disturbs the translation of repeat sequences in the genome to repetitive proteins. In contrast, splicing is rare in the trypanosomatid, reflecting repeats in genome and in the corresponding proteins. Thus, it is of interest how these parasites utilize such different patterns of TR, i.e., abundant small TR versus fewer but larger TR sequences.

In summary, we have demonstrated the usefulness of the bioinformatic analysis to identify antigenic parasite proteins. This study might contribute to a better understanding of immunological control, or lack thereof, during parasitic infection and possibly to antigen discovery using other pathogens as well.

ACKNOWLEDGMENTS

Sequence data were produced by the Pathogen Sequencing Unit at the Wellcome Trust Sanger Institute and were obtained from GeneDB (http://www.genedb.org). We thank Matthew Berriman and Chris Peacock, The Wellcome Trust Sanger Institute, for help with manuscript preparation. We thank Darrick Carter and Gregory Ireton for critical comments and Jeffrey Guderian and Garrett Poshusta for technical assistance.

This study was partly supported by the National Institutes of Health grant AI25038 and a grant from the Bill and Melinda Gates Foundation.

REFERENCES

1. **Araoz, R., N. Honore, S. Cho, J. P. Kim, S. N. Cho, M. Monot, C. Demangel, P. J. Brennan, and S. T. Cole.** 2006. Antigen discovery: a postgenomic approach to leprosy diagnosis. Infect. Immun. **74:**175–182.
2. **Benson, G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27:**573–580.
3. **Bern, C., S. N. Jha, A. B. Joshi, G. D. Thakur, and M. B. Bista.** 2000. Use of the recombinant K39 dipstick test and the direct agglutination test in a setting endemic for visceral leishmaniasis in Nepal. Am. J. Trop. Med. Hyg. **63:**153–157.
4. **Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D. C. Bartholomeu, N. J. Lennard, E. Caler, N. E. Hamlin, B. Haas, U. Bohme, L.**

Hannick, M. A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U. C. Alsmark, C. Arrowsmith, R. J. Atkin, A. J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T. J. Chillingworth, C. Churcher, L. N. Clark, C. H. Corton, A. Cronin, R. M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M. C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B. R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A. X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P. J. Mooney, S. Moule, D. M. Martin, G. W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C. S. Peacock, J. Peterson, M. A. Quail, E. Rabbinowitsch, M. A. Rajandream, C. Reitter, S. L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A. J. Simpson, L. Tallon, C. M. Turner, A. Tait, A. R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M. D. Adams, T. M. Embley, K. Gull, E. Ullu, J. D. Barry, A. H. Fairlamb, F. Opperdoes, B. G. Barrell, J. E. Donelson, N. Hall, C. M. Fraser, et al. 2005. The genome of the African trypanosome *Trypanosoma brucei*. Science **309:**416–422.

5. **Bhatia, A., N. S. Daifalla, S. Jen, R. Badaro, S. G. Reed, and Y. A. Skeiky.** 1999. Cloning, characterization, and serological evaluation of K9 and K26: two related hydrophilic antigens of *Leishmania chagasi*. Mol. Biochem. Parasitol. **102:**249–261.
6. **Bhatia, V., M. Sinha, B. Luxon, and N. Garg.** 2004. Utility of the *Trypanosoma cruzi* sequence database for identification of potential vaccine candidates by in silico and in vitro screening. Infect. Immun. **72:**6245–6254.
7. **Burns, J. M., Jr., W. G. Shreffler, D. R. Benson, H. W. Ghalib, R. Badaro, and S. G. Reed.** 1993. Molecular characterization of a kinesin-related antigen of *Leishmania chagasi* that detects specific antibody in African and American visceral leishmaniasis. Proc. Natl. Acad. Sci. USA **90:**775–779.
8. **Burns, J. M., Jr., W. G. Shreffler, D. E. Rosman, P. R. Sleath, C. J. March, and S. G. Reed.** 1992. Identification and synthesis of a major conserved antigenic epitope of *Trypanosoma cruzi*. Proc. Natl. Acad. Sci. USA **89:**1239–1243.
9. **Coppel, R. L., A. F. Cowman, R. F. Anders, A. E. Bianco, R. B. Saint, K. R. Lingelbach, D. J. Kemp, and G. V. Brown.** 1984. Immune sera recognize on erythrocytes *Plasmodium falciparum* antigen composed of repeated amino acid sequences. Nature **310:**789–792.
10. **Cowman, A. F., R. B. Saint, R. L. Coppel, G. V. Brown, R. F. Anders, and D. J. Kemp.** 1985. Conserved sequences flank variable tandem repeats in two S-antigen genes of *Plasmodium falciparum*. Cell **40:**775–783.
11. **Dame, J. B., J. L. Williams, T. F. McCutchan, J. L. Weber, R. A. Wirtz, W. T. Hockmeyer, W. L. Maloy, J. D. Haynes, I. Schneider, D. Roberts, et al.** 1984. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. Science **225:**593–599.
12. **de Andrade, C. R., L. V. Kirchhoff, J. E. Donelson, and K. Otsu.** 1992. Recombinant Leishmania Hsp90 and Hsp70 are recognized by sera from visceral leishmaniasis patients but not Chagas' disease patients. J. Clin. Microbiol. **30:**330–335.
13. **Dillon, D. C., C. H. Day, J. A. Whittle, A. J. Magill, and S. G. Reed.** 1995. Characterization of a *Leishmania tropica* antigen that detects immune responses in Desert Storm viscerotropic leishmaniasis patients. Proc. Natl. Acad. Sci. USA **92:**7981–7985.
14. **Gardner, M. J., N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M. S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell.** 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature **419:**498–511.
15. **Ghedin, E., W. W. Zhang, H. Charest, S. Sundar, R. T. Kenney, and G. Matlashewski.** 1997. Antibody response against a *Leishmania donovani* amastigote-stage-specific protein in patients with visceral leishmaniasis. Clin. Diagn. Lab. Immunol. **4:**530–535.
16. **Goto, Y., R. N. Coler, J. Guderian, R. Mohamath, and S. G. Reed.** 2006. Cloning, characterization, and serodiagnostic evaluation of *Leishmania infantum* tandem repeat proteins. Infect. Immun. **74:**3939–3945.
17. **Gruber, A., and B. Zingales.** 1993. *Trypanosoma cruzi*: characterization of two recombinant antigens with potential application in the diagnosis of Chagas' disease. Exp. Parasitol. **76:**1–12.
18. **Hertz-Fowler, C., C. S. Peacock, V. Wood, M. Aslett, A. Kerhornou, P. Mooney, A. Tivey, M. Berriman, N. Hall, K. Rutherford, J. Parkhill, A. C. Ivens, M. A. Rajandream, and B. Barrell.** 2004. GeneDB: a resource for prokaryotic and eukaryotic organisms. Nucleic Acids Res. **32:**D339–D343.
19. **Ibanez, C. F., J. L. Affranchino, R. A. Macina, M. B. Reyes, S. Leguizamon, M. E. Camargo, L. Aslund, U. Pettersson, and A. C. Frasch.** 1988. Multiple *Trypanosoma cruzi* antigens containing tandemly repeated amino acid sequence motifs. Mol. Biochem. Parasitol. **30:**27–33.
20. **Ivens, A. C., C. S. Peacock, E. A. Worthey, L. Murphy, G. Aggarwal, M. Berriman, E. Sisk, M. A. Rajandream, E. Adlem, R. Aert, A. Anupama, Z. Apostolou, P. Attipoe, N. Bason, C. Bauser, A. Beck, S. M. Beverley, G. Bianchettin, K. Borzym, G. Bothe, C. V. Bruschi, M. Collins, E. Cadag, L. Ciarloni, C. Clayton, R. M. Coulson, A. Cronin, A. K. Cruz, R. M. Davies, J. De Gaudenzi, D. E. Dobson, A. Duesterhoeft, G. Fazelina, N. Fosker, A. C. Frasch, A. Fraser, M. Fuchs, C. Gabel, A. Goble, A. Goffeau, D. Harris, C. Hertz-Fowler, H. Hilbert, D. Horn, Y. Huang, S. Klages, A. Knights, M. Kube, N. Larke, L. Litvin, A. Lord, T. Louie, M. Marra, D. Masuy, K. Matthews, S. Michaeli, J. C. Mottram, S. Muller-Auer, H. Munden, S. Nelson, H. Norbertczak, K. Oliver, S. O'Neil, M. Pentony, T. M. Pohl, C. Price, B. Purnelle, M. A. Quail, E. Rabbinowitsch, R. Reinhardt, M. Rieger, J. Rinta, J. Robben, L. Robertson, J. C. Ruiz, S. Rutter, D. Saunders, M. Schafer, J. Schein, D. C. Schwartz, K. Seeger, A. Seyler, S. Sharp, H. Shin, D. Sivam, R. Squares, S. Squares, V. Tosato, C. Vogt, G. Volckaert, R. Wambutt, T. Warren, H. Wedler, J. Woodward, S. Zhou, W. Zimmermann, D. F. Smith, J. M. Blackwell, K. D. Stuart, B. Barrell, et al.** 2005. The genome of the kinetoplastid parasite, *Leishmania major*. Science **309:**436–442.
21. **Kemp, D. J., R. L. Coppel, and R. F. Anders.** 1987. Repetitive proteins and genes of malaria. Annu. Rev. Microbiol. **41:**181–208.
22. **Koenen, M., A. Scherf, O. Mercereau, G. Langsley, L. Sibilli, P. Dubois, L. Pereira da Silva, and B. Muller-Hill.** 1984. Human antisera detect a *Plasmodium falciparum* genomic clone encoding a nonapeptide repeat. Nature **311:**382–385.
23. **Kotera, Y., J. D. Fontenot, G. Pecher, R. S. Metzgar, and O. J. Finn.** 1994. Humoral immunity against a tandem repeat epitope of human mucin MUC-1 in sera from breast, pancreatic, and colon cancer patients. Cancer Res. **54:**2856–2860.
24. **Mollick, J. A., F. S. Hodi, R. J. Soiffer, L. M. Nadler, and G. Dranoff.** 2003. MUC1-like tandem repeat proteins are broadly immunogenic in cancer patients. Cancer Immun. **3:**3.
25. **Pain, A., H. Renauld, M. Berriman, L. Murphy, C. A. Yeats, W. Weir, A. Kerhornou, M. Aslett, R. Bishop, C. Bouchier, M. Cochet, R. M. Coulson, A. Cronin, E. P. de Villiers, A. Fraser, N. Fosker, M. Gardner, A. Goble, S. Griffiths-Jones, D. E. Harris, F. Katzer, N. Larke, A. Lord, P. Maser, S. McKellar, P. Mooney, F. Morton, V. Nene, S. O'Neil, C. Price, M. A. Quail, E. Rabbinowitsch, N. D. Rawlings, S. Rutter, D. Saunders, K. Seeger, T. Shah, R. Squares, S. Squares, A. Tivey, A. R. Walker, J. Woodward, D. A. Dobbelaere, G. Langsley, M. A. Rajandream, D. McKeever, B. Shiels, A. Tait, B. Barrell, and N. Hall.** 2005. Genome of the host-cell transforming parasite *Theileria annulata* compared to *T. parva*. Science **309:**131–133.
26. **Reece, S. T., G. Ireton, R. Mohamath, J. Guderian, W. Goto, R. Gelber, N. Groathouse, J. Spencer, P. Brennan, and S. G. Reed.** 2006. ML0405 and ML2331 are antigens of *Mycobacterium* leprae with potential for diagnosis of leprosy. Clin. Vaccine Immunol. **13:**333–340.
27. **Reed, S. G.** 1996. Diagnosis of leishmaniasis. Clin. Dermatol. **14:**471–478.
28. **Reeder, J. C., and G. V. Brown.** 1996. Antigenic variation and immune evasion in *Plasmodium falciparum* malaria. Immunol. Cell Biol. **74:**546–554.
29. **Schofield, L.** 1991. On the function of repetitive domains in protein antigens of *Plasmodium* and other eukaryotic parasites. Parasitol. Today **7:**99–105.
30. **Singh, S., and R. Sivakumar.** 2003. Recent advances in the diagnosis of leishmaniasis. J. Postgrad. Med. **49:**55–60.
31. **Skeiky, Y. A., D. R. Benson, M. Elwasila, R. Badaro, J. M. Burns, Jr., and S. G. Reed.** 1994. Antigens shared by Leishmania species and *Trypanosoma cruzi*: immunological comparison of the acidic ribosomal P0 proteins. Infect. Immun. **62:**1643–1651.
32. **Stahl, H. D., P. E. Crewther, R. F. Anders, G. V. Brown, R. L. Coppel, A. E. Bianco, G. F. Mitchell, and D. J. Kemp.** 1985. Interspersed blocks of repetitive and charged amino acids in a dominant immunogen of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **82:**543–547.
33. **Sundar, S., R. Maurya, R. K. Singh, K. Bharti, J. Chakravarty, A. Parekh, M. Rai, K. Kumar, and H. W. Murray.** 2006. Rapid, noninvasive diagnosis of visceral leishmaniasis in India: comparison of two immunochromatographic strip tests for detection of anti-K39 antibody. J. Clin. Microbiol. **44:**251–253.
34. **Sundar, S., and M. Rai.** 2002. Laboratory diagnosis of visceral leishmaniasis. Clin. Diagn. Lab. Immunol. **9:**951–958.
35. **Sundar, S., S. G. Reed, V. P. Singh, P. C. Kumar, and H. W. Murray.** 1998. Rapid accurate field diagnosis of Indian visceral leishmaniasis. Lancet **351:**563–565.
36. **Vergara, U., M. Lorca, C. Veloso, A. Gonzalez, A. Engstrom, L. Aslund, U. Pettersson, and A. C. Frasch.** 1991. Assay for detection of *Trypanosoma cruzi* antibodies in human sera based on reaction with synthetic peptides. J. Clin. Microbiol. **29:**2034–2037.
37. **Zijlstra, E. E., Y. Nur, P. Desjeux, E. A. Khalil, A. M. El-Hassan, and J. Groen.** 2001. Diagnosing visceral leishmaniasis with the recombinant K39 strip test: experience from the Sudan. Trop. Med. Int. Health **6:**108–113.

Editor: W. A. Petri, Jr.