2102–2108 Nucleic Acids Research, 1994, Vol. 22, No. 11

Allele-specific expression of a variant-specific surface protein (VSP) of *Giardia lamblia*

Yongmin Yang and Rodney D.Adam*
Infectious Disease Section, Arizona Health Sciences Center, 1501 N. Campbell Avenue, Tucson, AZ 85724, USA

Received February 7, 1994; Revised and Accepted May 2, 1994 GenBank accession nos L25058–L25061 (incl.)

ABSTRACT

The surfaces of *Giardia lamblia* trophozoites demonstrate variable expression of a set of cysteine-rich surface proteins, called variant-specific surface proteins (VSP). The cloned *Giardia* line, WBA6, expresses a 170 kD VSP (VSPA6 or CRP170) which contains approximately 18 to 23 copies of a 65 amino acid repeat. We have cloned the expressed vspA6 gene containing 23 repeats from a genomic library as well as copies of the vspA6 gene with only 8 or 9 repeats from both WBA6 and from WB1269, a cloned line derived from WBA6 which has lost the expressed copy of the gene. The recombinant clones containing the genes with only 8 or 9 repeats have 8 nucleotide substitutions in the coding region. All the recombinant clones map to the same chromosomal location, yet RNA sequencing and comparison with the transcript size indicate that only the clone with 23 repeats contains a gene producing a stable transcript. The most likely interpretation of these data is that *G.lamblia* trophozoites contain multiple alleles of the vspA6 gene of which only one is expressed.

INTRODUCTION

Giardia lamblia is a flagellated protozoan parasite that causes diarrhea in people throughout the world. The clinical manifestations of infection are highly variable with the majority of infected humans remaining asymptomatic, while a substantial minority develop diarrhea and malabsorption. The duration of infection is also variable with spontaneous resolution of some infections within weeks while other infections may last for years (1). Host (2–4) and parasite (4–6) factors are important determinants of the outcome of infection. One of the parasite factors may be the family of variant-specific surface proteins (VSP's) found on the surface of the trophozoites. These surface proteins have also been called excretory-secretory products (7), cysteine-rich proteins (CRP) (8), Trophozoite surface antigens (TSA) (9), and Trophozoite surface proteins (TSP) (10), but will be called VSP's as suggested by Mowatt, *et al* (11). The upper case will be used to refer to the surface protein and the lower case for the gene encoding the VSP. The VSP's have a cysteine content of approximately 12% and a recurring motif of CXXC. With the exception of the CXXC motif, the N-termini are variable, but the deduced 36–38 C-terminal amino acids are 80–90% identical for all vsp genes that have been sequenced. Cloned trophozoites demonstrate variable expression of the VSP's (4,12) and are able to change from expression of one to another VSP every six to 12 generation times (13). Cloned lines derived from the same isolate and expressing different VSP's differ in their virulence for humans (6). In addition, similar isolates that express different VSP's differ in protease susceptibility (14).

The multiply cloned A6 line of the WB isolate expresses a 170 kD VSP, VSPA6 (also called CRP170). Antigenic variants were derived from WBA6 by incubation of the trophozoites with a monoclonal antibody that is cytotoxic for trophozoites expressing VSPA6 (15). In one of the cloned variants, WB1269, VSPA6 is replaced by another VSP (VSP1269 or CRP72) and an expression-associated copy of the vspA6 gene (represented by a six kb HindIII fragment) is lost (16). The loss of this expression-associated copy has provided a useful tool for studying the mechanism whereby expression of the vspA6 gene is lost in WB1269.

G.lamblia trophozoites contain two nuclei that are transcriptionally active and are equal in all ways by which they have been compared (17). Five distinct chromosomal bands can be identified on PFGE separations, ranging in size from one to four million base pairs (18). The ploidy of *G.lamblia* is not known, but preliminary data suggests that *G.lamblia* trophozoites are polyploid with approximately 6 to 10 copies of each chromosome (18,19). In this report, we provide evidence for the existence of multiple alleles of the vspA6 gene, consistent with the polyploid nature of *G.lamblia*. We also provide evidence that only one of the alleles is expressed. Allele-specific expression of the vspA6 gene, provides potential clues for the mechanism of antigenic variation in *G.lamblia*.

MATERIALS AND METHODS

Strains and cultivation

Giardia lamblia trophozoites WBA6 and WB1269 expressing VSPA6 (CRP170) and VSP1269 (CRP72), respectively, have been described (20). WBA6 was initially cloned twice on soft

*To whom correspondence should be addressed

agar and twice again by limiting dilution. The variant WB1269 was selected by incubation of WBA6 with a monoclonal antibody cytotoxic for WBA6 trophozoites (15) and the resultant variants were cloned by limiting dilution. *G.lamblia* trophozoites were grown in modified TYI-S-33 medium (21).

Nucleic acid preparation

Giardia DNA was prepared by cesium chloride ultracentrifugation as previously described (22). Bacteriophage DNA was prepared by a liquid culture lysate and cesium chloride ultracentrifugation (23). Plasmid DNA and M13 DNA was prepared by standard methods (23). Total RNA was prepared by guanidinium isothiocyanate followed by guanidinium HCl treatment and ethanol precipitation (24).

Genomic cloning of vspA6 genes

Genomic DNA of WBA6 and WB1269 was digested with BglII and cloned into BamHI-digested λEMBL3 (16). The genomic libraries were screened by using an oligonucleotide based on the sequence of the CRP170 repeat region (16) (oligo # 135; GGT-CGCTGGGATTCGTCTTCTTCA) as a probe.

For construction of the cosmid library, WBA6 DNA between 150 kb and 200 kb was prepared by gentle cell lysis and sucrose gradient ultracentrifugation (25). The highly viscous DNA was dialyzed and partially digested with Sau3AI. The fraction around 50 kb was collected and extracted with phenol/chloroform followed by ethanol precipitation. Dephosphorylation of partially digested WBA6 DNA, and preparation of Supercos cosmid vector were performed using protocols provided by the manufacturer (Stratagene, La Jolla, CA, USA).

DNA and RNA sequence analysis

DNA sequencing was performed by the chain termination method (26) and Sequenase 2.0 (U.S. Biochemical, Cleveland, OH), with M13 and plasmid vectors. RNA sequencing was performed as described (27) using 50 micrograms of total RNA and end-labeled antisense oligonucleotide # 3532 (AGGCACATCCCCTTCGAG) as the primer. Acrylamide gels with 50% formamide were used for resolution of GC compression regions. Sequence analysis performed by using programs in the GCG package of the University of Wisconsin (28).

Pulsed field gel electrophoresis (PFGE)

Giardia chromosomal DNA was prepared and separated by PFGE as previously described (18). For restriction digests of individual chromosomal bands, blocks were equilibrated in the appropriate enzyme buffer. Twenty units of enzyme were added, followed by incubation at 4°C for one hour and at the recommended digestion temperature for one hour. The short incubation time was chosen to minimize nonspecific degradation of the chromosomal DNA. The blocks were then separated by PFGE as described (19) using a contour-clamped homogeneous electric field (CHEF) apparatus.

Probes and hybridization

DNA was transferred to nylon membrane in alkaline buffer (0.4M NaOH, 1M NaCl). The blots were washed in 0.1×SSC with 0.1% SDS at 65°C for double-stranded DNA probes and in 6×SSC with 0.1% SDS at 50°C for oligonucleotide probes. Double-stranded DNA was labeled by random priming (29) or by linear amplification of randomly primed probe (Bios Laboratories; New Haven, CT). Oligonucleotides were labeled by T4 polynucleotide kinase using [γ^{32}P] ATP. Unincorporated nucleotides were removed by Sephadex G50 (double-stranded probes) or Sephadex G25 (oligonucleotide probes) column chromatography.

Immunoblotting

For Western immunoblotting, recombinant XL1 Blue *Escherichia coli* cells or *G.lamblia* trophozoites were sonicated or lysed with SDS. The recombinant cells were not treated with a reducing agent, but were subjected to heat denaturation for 3 minutes. For the trophozoites, neither a reducing agent or heat denaturation were used. The samples were separated by 10% SDS-PAGE, and transferred to a PVDF membrane (Millipore, San Francisco, CA, USA) in 10 mM CAPS buffer (10 mM CAPS, pH 11.0, 5% methanol). The blots were reacted with MAb 6E7 (15) using hydrogen peroxidase-conjugated goat anti-mouse IgG as the second antibody.

RESULTS

Analysis of multiple clones of the vspA6 gene from WBA6 and WB1269

The vspA6 (CRP170) cDNA clone previously reported contained 21 copies of a 195 bp tandem repeat (16). An oligonucleotide specific to the repeat, hybridized to multiple bands of WBA6 genomic DNA for all of the enzymes chosen (Figure 1A) (including 14, ~6, 3.9, 3.7, and 1.5 kb HindIII fragments). The presence of multiple bands with the repeat probe suggested a multigene family, with each member hybridizing to the repeat sequence. We had shown in previous work that the ~6 HindIII fragment was the expression-associated fragment, and that it was deleted in the WB1269 *Giardia* line (16). The variability of the repeat copy number for the expression-associated fragment

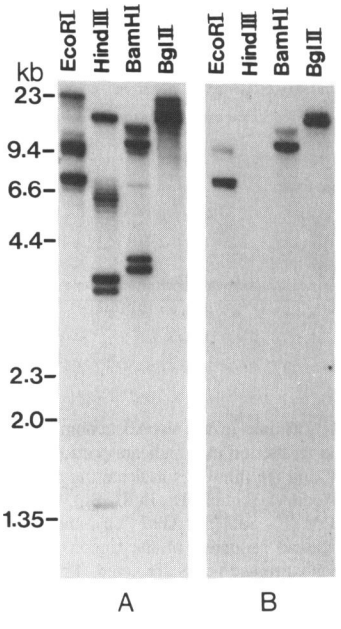
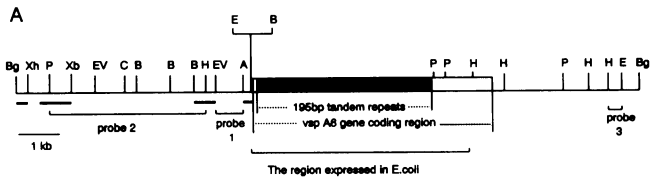


Figure 1. Restriction digests of WBA6 genomic DNA probed with (**A**) repeat oligo # 135, and (**B**) probe 3 (See Figure 2A). The lack of hybridization to the HindIII lane in 1B is due to the small size of the HindIII fragment that contains probe 3 (see Figure 2A). The blot was stripped between hybridizations.

(18−23) was most likely due to population heterogeneity, since the repeat-containing HindIII fragments from different DNA preparations showed some variation around the 6 kb size range; in contrast, the 3.7 and 3.9 kb fragments were constant for all DNA preparations (data not presented).

In order to look for differences among the vspA6 gene(s) containing the 195 bp repeat that might explain why the gene represented by the ~6 HindIII fragment was expressed while others were not, we screened λEMBL3 libraries of BglII-digested WBA6 and WB1269 genomic DNA using the repeat oligo (# 135) as a probe. Twenty-three clones from the WBA6 library and 10 clones from the WB1269 library were obtained; eleven from WBA6 and four from WB1269 were randomly chosen for further analysis. It initially appeared that the clones differed at the 5′ and 3′ ends (16); however, the apparent difference was a cloning artifact due to the presence of multiple BglII inserts in some clones. Therefore, analysis was limited to the BglII fragments containing the 195 bp repeat. The clones were analyzed by restriction mapping and limited sequence analysis (Figure 2). The repeat copy numbers were determined by partial digestion with enzymes digesting once in the repeat (SalI, NarI). The clones obtained from the λEMBL3 libraries had 3.7, 3.9, or 6.6 kb HindIII repeat-containing fragments; none contained the 1.5 or 14 kb HindIII-containing fragments. The restriction maps outside the repeat-containing regions were identical for each of the clones obtained. Subsequently, clones with the 1.5 and 14 kb repeat-containing HindIII fragments have been obtained from a cosmid library. Initial characterization of these cosmid clones has indicated that they contain regions of sequence similarity, but differ from the clones with the 3.7, 3.9, or ~6 kb HindIII fragments in genomic location, restriction patterns, and in the sequence of the coding regions 5′ and 3′ to the repeats (Yang and Adam, unpublished results). Thus, the genes represented by the 1.5 and 14 kb HindIII fragments are members of a family of genes related to, but distinct from the vspA6 gene, and will not be discussed further in this manuscript. The clones containing the 3.7, 3.9, and ~6 kb HindIII fragments are referred to as vspA6 clones.



B

```
-372 CCCCATGCCACTCCATGAGCATCACGATCGCTGTGTCCCCCACGATGGTCACTCAGGCAA
-312 TCAGACAGATTACAGGACACGGTGGCTACGGATGTCAAGGGCTACAAGCGCCAGCTGGGT
-252 TAATGTAGCAATGGATGCACTCATTATCTATGTATACACTTATGCCTTCTCAGAGCAAGG
-192 AAAGCCAGCGCTGGGCTTCCTCAGTCGTGCTTGTTGAGGGCGGATGCTGCATTAACGGGA
-132 CTCCGGTGTAGCTTGAGGATGTGCCCTCGCTTCCGTTGGCCGCTGTGTTCCATTGAATTC
 -72 CGTGACATGAAGCAGGATCCCGTGCGCCTCTTAATGCAAACCAGAGCGGTGCTGTGCCAG
 -12 CTTCTGGCCTCAATGTTGGTAGGAATTGGTTTGATTTTTGCAGTCGCACTGGCTGATGCA
                M L V G I G L I F A V A L A D A

  49 TGCAGTCCATCGATCGATGGTTGCGCTGAATGCGATTCAACTGGGAAAAAATGCACCAAA
     C S P S I D G C A E C D S T G K K C T K

 109 TGCGACGCAAACGGCGCCCGCGCCCTACCTGAAGAAGACGAATCCCAGCGACCCGACGGGG
                  AA-A
     C D A N G A A P Y L K K T N P S D P T G
     - - - - - N T - - - - - - - - - - - - -

 169 ACGTGCGTCAGCGCGGTCGAC
```

C

```
       1          2                    3          4 5
265 E GGAACGTGCCTTACTGCTGAAGAATGCACCTCCGATACGACGCACTTCACCAAAGAGAAG
    N ---G---------C----------------------A-----------T-G-------
    E G T C L T A E E C T S D T T J F T K E K
    N   A                      K           R
```

Figure 2. A. Restriction map of the vspA6 genomic clone, A6.GBE3-4. The heavy lines under the restriction map indicate portions of the 5′ flanking region that were sequenced, and the thin lines indicate the probes used. The restriction enzymes shown are AccI(A), BamHI (B), BglII (Bg), ClaI (C), EcoRI (E), EcoRV (EV), HindIII (H), PstI (P), SalI (S), XbaI (Xb), and XhoI (Xh). **B.** Nucleotide and translated amino acid sequence of the region from −372 to +189. The substitutions in the nonexpressed genes are noted. The NarI site in the first repeat is underlined. **C.** Nucleotide and translated amino acid sequence of the region from 265 to 324 (numbering restarted after the repeats to keep the numbering constant in clones with different repeat copy numbers) of the downstream region for the expressed (E) clones (A6.GBE3-4, CRP170-L5 cDNA clone) and WBA6 RNA. The substitutions found in the nonexpressed (N) clones (A6.GBE13-2, A6.GBE5-5, 1269.GBEm, 1269.GBEl, 1269.GBEc, 1269.GBEp)are noted immediately below. The numbers of the five nucleotide substitutions correspond to those of the RNA sequencing gel (Figure 4).

Genomic DNA was probed with a downstream region of a vspA6 clone (probe 3 shown in Figure 2A) specific to the vspA6 genes (Figure 1B). The three EcoRI fragments identified (6.9 kb and 9.4 kb), and the BamHI and BglII fragments are the sizes predicted by the sizes of the comparable fragments in the vspA6 clones. The 6.9 kb EcoRI fragment (which corresponds to the 3.7 and 3.9 kb HindIII fragments) appeared as a doublet in lighter exposures, consistent with a 200 bp size difference between the fragments of the doublet.

Differences among the vspA6 clones

The vspA6 clones were found to differ in three ways: (1) Repeat copy number, (2) Presence or absence of 3 nucleotide substitutions in the first repeat, and (3) Presence or absence of 5 nucleotide substitutions in the region 3′ to the repeat (Table 1).

Repeat copy number. One of the WBA6 genomic clones (A6.GBE3-4) contained 23 repeats, while the CRP170-L5 cDNA clone contained 21 repeats. The remainder of the WBA6 genomic clones and all the WB1269 genomic clones contained 8 or 9 repeats. An additional 10 vspA6 clones were obtained by screening a WBA6 cosmid library with the 195 bp repeat. Of the 10 vspA6 cosmid clones obtained, 2 contained 8 repeats, 4 contained 9 repeats, and 4 contained 18− ~23 repeats. The repeat-containing fragments of the clones with 8, 9, and 21−23 repeats comigrated with the 3.7, 3.9, and ~6 kb genomic HindIII fragments seen in Figure 1A, and the 6.9 kb doublet and 9.4 kb EcoRI fragments seen in Figure 1B (data not shown). The difference in repeat copy number also provided an explanation for the relative difference in intensity between the ~6 kb HindIII ladder in Figure 1A and the 9.4 kb EcoRI fragment in Figure 1B, since the probe used in Figure 1A was from the repeat region, while the probe used in Figure 1B was present as a single copy.

Substitutions in the first repeat. In the first repeat, the sequence of the WBA6 genomic clone containing 23 repeats (A6.GBE3-4) was identical to that of the CRP170-L5 cDNA clone, while all the WBA6 and WB1269 clones with 8 or 9 repeats contained 3 nucleotide substitutions (Figure 2B). Since the substitutions in the first repeat result in the loss of a NarI site in the clones containing 8 or 9 repeats, the presence of this NarI site may be a marker for the expressed gene. In order to evaluate this possibility, probe 1 (Figure 2A) was hybridized to HindIII/NarI digests of WBA6 and WB1269 DNA (Figure 3).

Table 1.

Clones	Repeat copy #	HindIII Frag (Kb)	NarI site in first repeat	Matches RNA Sequence[2]	5 nt Substitution[3]
WBA6 cDNA clones					
CRP170-L5	21	NA[1]	YES	YES	NO
WBA6 genomic clones					
A6.GBE 3-4	23	6.6	YES	YES	NO
A6.GBE 13-2	9	3.9	NO	NO	YES
A6.GBE 5-5	9	3.9	NO	NO	YES
A6.GBE 20-1	9	3.9	NO	NO	YES
A6.GBE 19-1	9	3.9	NO	NO	YES
A6.GBE 3-3	9	3.9	NO	NO	ND
A6.GBE 7-2	9	3.9	NO	ND	ND
A6.GBE 7-2.2	9	3.9	NO	ND	ND
A6.GBE 11-1	8	3.7	NO	ND	ND
A6.GBE 11-2	9	3.9	NO	ND	ND
A6.GBE 11-3	8	3.7	NO	ND	ND
WB1269 genomic clones					
1269.GBEm	8	3.7	NO	NO	YES
1269.GBEl	8	3.7	NO	NO	YES
1269.GBEp	9	3.9	NO	NO	YES
1269.GBEc	9	3.9	NO	NO	YES

[1]The size of the HindIII fragment could not be directly determined, but is calculated to be 6.2 kb.
[2]Matches the RNA sequence in the downstream region containing the nucleotide substitutions.
[3]Contains the 5 nucleotide substitutions as compared to the RNA. NA not applicable, ND not done.

The distance from the HindIII site to the NarI site of the first repeat is 1.4 kb, while the distance to the NarI site of the second repeat is 1.6 kb. The absence of the 1.4 kb fragment from WB1269 genomic DNA correlates with the absence of the ~6 kb expression-linked HindIII fragment, and suggests that the first repeat contains the NarI site only in the expressed gene. The fragments larger than 1.6 kb have not been analyzed, but may be from members of the family of genes related to the vspA6 gene (those corresponding to the 1.5 and 14 kb HindIII fragments of Figure 1A).

Substitutions in the 3' region. Sequence analysis of a 900 bp segment of DNA 3' to the repeat showed identity between the A6.GBE3-4 genomic and CRP170-L5 cDNA clones. The CRP170-4 cDNA clone (16) had a single nucleotide substitution in comparison to the CRP170-L5 clone, possibly indicating a substitution that occurred in the vspA6 gene between the times of preparation of the two different cDNA libraries from which these clones were derived or an error by the DNA polymerase during construction of the cDNA library from which the CRP170-4 clone was derived. In contrast, all the WBA6 and WB1269 clones with 8 or 9 repeats contained 5 nucleotide substitutions (Figure 2C). The sequences of the clones were compared to that of the RNA transcript by direct sequencing of RNA prepared from WBA6 trophozoites (Figure 4). The genomic A6.GBE3-4 and cDNA CRP170-L5 clones were identical to the RNA transcript for each of the 5 nucleotides, and each contained coding regions in the correct size range for the 6 kb RNA transcript of the vspA6 gene (16). These data suggested that the A6.GBE3-4 clone contained the expressed form of the vspA6 gene, while the other clones contained nonexpressed genes or pseudogenes.

Identity in the upstream noncoding region

In order to look for sequence differences 5' to the initiation codon that might affect expression of the vspA6 gene, the sequences of three vspA6 clones, one of each containing the 3.7, 3.9, and ~6 kb HindIII fragments (Figure 1A), and two clones from WB1269, one of each containing the 3.7 and 3.9 kb fragments, were analyzed from -372 to the initiation codon (Figure 2B). The sequences were identical, indicating that the absence of a stable transcript did not result from alterations in the sequence of the region immediately upstream from the initiation codon. It is still possible that differences in the downstream region or farther upstream may affect transcription, but the identity of the restriction maps in these regions for expressed and nonexpressed alleles suggests that any sequence differences are small.

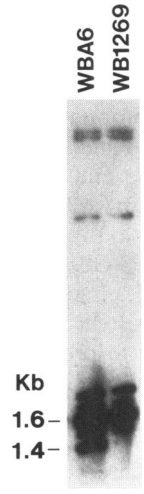


Figure 3. Southern blot of WBA6 and WB1269 DNA double-digested with HindIII and NarI, and probed with Probe 1 (Figure 2A). The 1.4 kb (representing the gene with a NarI site in the first repeat) and 1.6 kb (representing the genes that do not have the NarI site in the first repeat) fragments are noted.

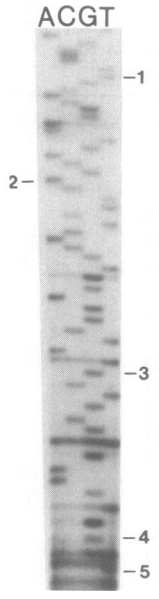


Figure 4. RNA sequence using an antisense primer from 350 to 333 in the downstream region. The locations of the 5 downstream nucleotide substitutions are numbered (see Figure 2C) and in each case the band is unambiguous.

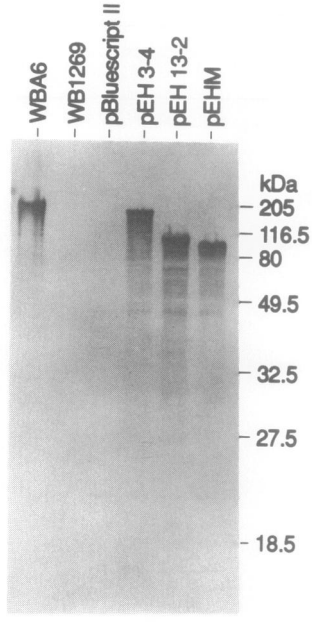


Figure 5. Western blot of *Giardia* trophozoites and recombinant protein reacted with MAb6E7. The first two lanes contain cell lysate from WBA6 and WB1269 *Giardia* trophozoites, respectively. The third lane contains pBluescript II cell lysate, and the following lanes contain recombinant proteins expressed by the EcoRI/HindIII subclones of the expressed clone, A6.GBE3-4 (pEH3-4), and the nonexpressed clones, A6.GBE13-2 (pEH13-2) and 1269.GBEm (pEHm). As expected, WB1269 trophozoites and pBluescript II cell lysate did not react with the MAb. Of note, the EcoRI/HindIII subclone of A6.GBE3-4 in pBluescript II (pEH3-4) contained only 17 copies of the repeat rather than the 23 in the λEMBL3 clone, due to deletion of repeats during subcloning in pBluescript II. The sizes of the recombinant proteins as estimated by gel migration (155, 100, and 93 kD for pEH3-4, pEH13-2, and pEHm, respectively) closely match the sizes calculated from the translated amino acid sequences of the open reading frames (160, 102, and 95 kDa, respectively).

Open reading frame in the nonexpressed clones

If the open reading frame was altered in the nonexpressed clones, it is likely that these clones would be incapable of expressing fusion proteins of the appropriate size in *E.coli*. Therefore, the repeat-containing EcoRI/HindIII fragments of the expressed A6.GBE3-4 (pEH3-4;) and nonexpressed A6.GBE13-2 (pEH13-2) and 1269.GBEm (pEHm) clones (see Figure 2A) were subcloned into the expression vector, pBluescript II. Cell lysates were separated by SDS-PAGE and western blots were reacted with MAb 6E7 (Figure 5). Each was the size predicted by the open reading frame, consistent with the fact that the insert was not in the correct frame to produce a fusion protein with β-galactosidase. Initiation from within the insert, rather than from the β-galactosidase initiation codon, has also been noted for another vsp gene (9).

The sequence of the nonrepeat portion of the EcoRI/HindIII subclone (pEHm) of the 1269.GBEm clone was determined to further assess whether the nonexpressed vspA6 genes contained open reading frames. The sequence contained an open reading frame identical to that of the expressed A6.GBE3-4 clone, except that it contained the same 8 nucleotide substitutions identified in the other nonexpressed clones.

The vspA6 gene is present at a single locus on chromosomal band 4

PFGE separations of *G.lamblia* trophozoites reveal at least five chromosomal bands, for which chromosome-specific probes have been defined (18). These are called Bands 1−5 in order of increasing size. The repeat region of the vspA6 gene hybridized to both Bands 4 and 5, but more intensely to Band 4; however, a vspA6-specific probe (Probe 3, Figure 2A) hybridized specifically to Band 4, indicating that the vspA6 gene was located on the chromosome(s) contained in Band 4. In addition, a vspA6-specific probe hybridized specifically to a 250 kb NotI fragment of Band 4 (data not presented). These findings could be explained by the presence of multiple (at least 3) copies of the vspA6 gene(s) as repeating units on a 250 kb region of the chromosome, or by the existence of different alleles at the same locus on homologous chromosomes.

In order to distinguish between the two possibilities, Band 4 DNA was digested with infrequently digesting restriction enzymes, separated by CHEF electrophoresis, and the blots probed with upstream or downstream fragments of the A6.GBE3-4 (expressed) clone. The probes were common to the expressed and nonexpressed vspA6 clones, allowing them to identify the locations of expressed and nonexpressed vspA6 genes. SgrAI was a particularly useful enzyme, since a single 42 kb fragment was detected with a probe from the upstream region (Figure 6A) while a 100 kb fragment was detected with the downstream probe (Figure 6B). For each enzyme, only a single band of hybridization was present, with the exception of enzymes that yielded short (<23 kb) fragments spanning the repeat region (eg. SgrAI/PmeI digest, Figure 6B). The different sizes of these smaller fragments are consistent with the variability in the repeat copy number of the vspA6 clones. A long-range map of the vspA6 gene derived from these data is shown in Figure 7. The data exclude the possibility of a large repeating unit with multiple copies of the vspA6 gene and suggest that multiple alleles of the vspA6 gene are present on homologous chromosomes at a single locus.

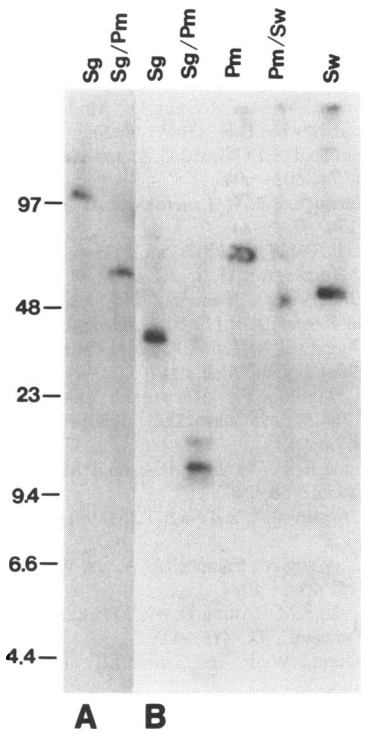


Figure 6. Digests of Band 4 with infrequently-digesting enzymes separated by CHEF electrophoresis and probed with (**A**). probe 3 from the downstream flanking region (see Figure 2A), and (**B**) probe 2 from the upstream flanking region (see Figure 2A). For probe 3, only the digests using SgrAI are shown since the results with probes 2 and 3 were identical for the other enzymes, consistent with lack of sites for these enzymes in the A6.GBE3-4 clone. The blot was stripped between hybridizations. Restriction enzymes shown are SgrAI (Sg), PmeI (Pm), and SwaI (Sw).

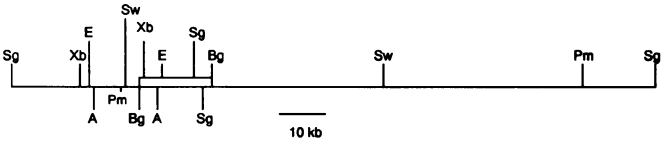


Figure 7. Long range map of the region containing the vspA6 locus. The open rectangle depicts the region spanned by the A6.GBE3-4 clone. In the region flanking the clone, the first site for each of the enzymes is shown. NotI/SgrAI digests were the same as digests with SgrAI alone, placing the NotI sites beyond the SgrAI sites. Of note, the A6.GBE3-4 clone contained 4 SgrAI sites; the two external-most sites are shown. The restriction enzymes shown are AseI (A), BglII (Bg), EcoRI (E), PmeI (Pm), SgrAI (Sg), SwaI (Sw), and XbaI (Xb).

An alternative explanation is that the population of trophozoites from which the DNA was prepared is heterogeneous, containing some trophozoites with approximately 18 to 23 copies of the repeat in the vspA6 gene and some with 8 or 9 copies. However, numerous cloned WB lines all contain one or two HindIII fragments at approximately 6 kb and one or two fragments at approximately 4 kb (see ref (16)). In addition, the location and stoichiometry of the HindIII fragments has remained constant for a cloned population (data not shown). Thus, population heterogeneity is an unlikely etiology of the multiple alleles.

In order to approximate the ratio between nonexpressed and expressed alleles, the stoichiometry was estimated by densitometric comparison of the bands representing the expressed and silent alleles. The probe used in the autoradiograph shown in Figure 1B is present as a single copy in the nonexpressed and expressed alleles; therefore, the hybridization should be in proportion to the copy number represented by each restriction fragment. The 6.9 kb EcoRI and 10 kb BamHI fragments contain the nonexpressed alleles (8 or 9 repeats), while the 9.4 EcoRI and 12.5 kb BamHI fragments contain the expressed allele with the greater number of repeats. A comparison of the intensities yielded ratios of 5:1 and 4:1 for the EcoRI and BamHI digests, respectively.

DISCUSSION

In this report, we have described multiple clones of the vspA6 gene that differ in repeat copy number and the presence or absence of 8 nucleotide substitutions in two portions of the coding region. Despite the differences, all clones hybridize to the same locus on chromosomal Band 4. The existence of multiple forms of the vspA6 gene, all mapping to the same location implies that the *Giardia* genome contains multiple alleles of the vspA6 gene. These alleles can be designated as vspA6.1 (expressed, containing 18 to 23 repeats), vspA6.2 (nonexpressed, containing 9 repeats), and vspA6.3 (nonexpressed, containing 8 repeats). The vspA6.1 allele (represented by the A6.GBE3-4 genomic and CRP170-L5 cDNA clone) is identical to the RNA sequence in the region downstream from the repeats, while the vspA6.2 and vspA6.3 alleles contain 5 nucleotide substitutions in the downstream region. The RNA sequence data as well as the size of the RNA transcript indicate that only the vspA6.1 allele produces a stable transcript.

In previous work, chromosomes from equal numbers of *Giardia* nuclei and the haploid form of Plasmodium falciparum were separated by PFGE and analyzed by densitometry. The results of the comparison suggested that *Giardia* trophozoites contained approximately 6−10 copies of each chromosome per trophozoite (3−5 per nucleus) (18). The evidence for multiple alleles of the same gene provides independent evidence that *G. lamblia* trophozoites are polyploid. When the ratio of 4−5:1 for nonexpressed to expressed alleles is compared with the estimate of 6−10 copies of each chromosomes, it appears likely that each trophozoite contains one or two copies of the expressed allele.

The existence of allele-specific expression of the vsp genes is especially important in understanding the mechanism of antigenic variation. It is difficult to explain how multiple alleles of a gene could be turned off simultaneously, and the expression of only one of several alleles may remove this obstacle to understanding the mechanism of antigenic variation in *Giardia*. It is also interesting to consider whether the vsp genes are transcribed from one or both nuclei. The two nuclei are the same by all parameters that have been tested and both are transcriptionally active (17). If there is only one copy of the expressed allele per trophozoite, transcription must occur from only one nucleus, and could provide a reason for the presence of two nuclei that otherwise appear the same. The answer to this question promises to provide interesting insights into the control of vsp gene expression, and potentially the reason why *Giardia* trophozoites contain two nuclei.

In some organisms, antigenic variation occurs by DNA recombination which creates an open reading frame in the gene being activated. This occurs for the variant surface glycoprotein

(VSG) genes of African trypanosomes, the pilus protein genes of *Neisseria gonorrhea* and the variable major protein (VMP) genes of *Borrelia hermsii*, (30).The coexistence of expressed and nonexpressed alleles in *G.lamblia* that contain open reading frames and are capable of being expressed in *E. coli* indicates that antigenic variation in *Giardia* does not occur through the creation of an open reading frame by DNA recombination.

In mammalian cells, imprinting and X-chromosome inactivation are forms of allele-specific expression and provide for variable expression of protein-coding genes. These phenomena may occur through differential methylation of the involved genes or chromosomes (31). In African trypanosomes, *in situ* activation of the VSG genes can also occur without any obvious change in the sequence of the DNA of the gene and its flanking regions. The mechanisms have not yet been elucidated, but changes in higher chromosome structure or position, changes in more distant sequences, or the presence of altered nucleotides have been proposed as possibilities (32–34). Which if any of these mechanisms are used in *Giardia* is not known.

G.lamblia trophozoites reproduce by mitosis, and a sexual cycle with meiotic recombination has not been identified. The accumulation of allelic differences in *G.lamblia* is not surprising, since mitotic recombination is 1000 fold less effective than meiotic recombination for maintenance of allelic similarity in *Saccharomyces cerevisiae* (35). However, it is possible that mitotic recombination in *G.lamblia* limits the accumulation of allelic variability. The mobility of clusters of telomere-associated rDNA genes suggests that the frequency of mitotic chromosome recombination may be very high in *G.lamblia* (36–38). In fact, studies using cloned lines of *G.lamblia* have yielded an estimate of 0.03 recombinations per generation for chromosomes containing the rDNA genes (37). If gene conversion occurs during the mitotic recombination, the accumulation of allelic variability may be smaller than otherwise expected. Allelic differences were not been identified for the ADP-ribosylating factor, the one other *Giardia* gene where multiple clones have been sequenced (39). The difference of the expressed vspA6 allele (vspA6.1) from two nonexpressed alleles that at eight nucleotides suggests the possibility that the accumulation of nucleotide substitutions is more common in the expressed allele of a vsp gene. The variability in repeat copy number for vspA6.1 (18–23), as demonstrated by the ~6 kb HindIII ladder of fragments (Figure 1A) is most likely due heterogeneity of the population of trophozoites from which the DNA was prepared. The occurrence of the variability in repeat copy number for the expressed allele, but not for the nonexpressed alleles, suggests that the expressed allele may be more vulnerable to recombination among the repeats with resultant change in copy number. Alternatively, the variability in repeat copy for the expressed allele may simply be a reflection of the greater number of repeats in the expressed allele. It is interesting to speculate whether active transcription increases the vulnerability of a vsp gene to point mutations and to DNA recombination. Determination of the occurrence and role of allelic variability will require the analysis of further expressed and nonexpressed vsp genes.

ACKNOWLEDGEMENTS

We thank Lynda Schurig for invaluable technical assistance and Theodore Nash, Michael Mowatt, and William Birky for helpful comments. This work was supported by NIH grant R29-AI28551-01A1 to R.D.A.

REFERENCES

1. Smith,P.D., Gillin,F.D., Spira,W.M. and Nash,T.E. (1982) Gastroenterology, **83**, 797–803.
2. Moore,G.T., Cross,W.M., McGuire,D., Mollohan,C.S., Gleason,N.N., Healy,G.R. and Newton,L.H. (1969) N. Engl. J. Med., **281**, 402–407.
3. Istre,G.R., Dunlop,T.S., Gaspard,G.B. and Hopkins,R.S. (1984) Am. J. Public Health, **74**, 602–604.
4. Nash,T.E., Herrington,D.A., Losonsky,G.A. and Levine,M.M. (1987) J. Infect. Dis., **156**, 974–984.
5. Aggarwal,A. and Nash,T.E. (1987) Am. J. Trop. Med. Hyg., **36**, 325–332.
6. Nash,T.E., Herrington,D.A., Levine,M.M., Conrad,J.T. and Merritt,J.W.,Jr. (1990) J. Immunol., **144**, 4362–4369.
7. Nash,T.E. and Keister,D.B. (1985) J. Infect. Dis., **152**, 1166–1171.
8. Adam,R.D., Aggarwal,A., Lal,A.A., de la Cruz,V.F., McCutchan,T. and Nash,T.E. (1988) J. Exp. Med., **167**, 109–118.
9. Gillin,F.D., Hagblom,P., Harwood,J., Aley,S.B., Reiner,D.S., McCaffery,M., So,M. and Guiney,D.G. (1990) Proc. Natl. Acad. Sci. USA, **87**, 4463–4467.
10. Ey,P.L., Khanna,K.K., Manning,P.A. and Mayrhofer,G. (1993) Mol. Biochem. Parasitol., **58**, 247–257.
11. Mowatt,M.R., Aggarwal,A. and Nash,T.E. (1991) Mol. Biochem. Parasitol., **49**, 215–227.
12. Boothroyd,J.C., Wang,A., Campbell,D.A. and Wang,C.C. (1987) Nucleic. Acids. Res., **15**, 4065–4084.
13. Nash,T.E., Banks,S.M., Alling,D.W., Merritt,J.W.,Jr. and Conrad,J.T. (1990) Exp. Parasitol., **71**, 415–421.
14. Nash,T.E., Merritt,J.W.,Jr. and Conrad,J.T. (1991) Infect. Immun., **59**, 1334–1340.
15. Nash,T.E. and Aggarwal,A. (1986) J. Immunol., **136**, 2628–2632.
16. Adam,R.D., Yang,Y.M. and Nash,T.E. (1992) Mol. Cell Biol., **12**, 1194–1201.
17. Kabnick,K.S. and Peattie,D.A. (1990) J. Cell Sci., **95**, 353–360.
18. Adam,R.D., Nash,T.E. and Wellems,T.E. (1988) Nucleic. Acids. Res., **16**, 4555–4567.
19. Adam,R.D. (1992) Nucleic. Acids. Res., **20**, 3057–3061.
20. Nash,T.E., Aggarwal,A., Adam,R.D., Conrad,J.T. and Merritt,J.W.,Jr. (1988) J. Immunol., **141**, 636–641.
21. Keister,D.B. (1983) Trans. R. Soc. Trop. Med. Hyg., **77**, 487–488.
22. Nash,T.E., McCutchan,T., Keister,D., Dame,J.B., Conrad,J.D. and Gillin,F.D. (1985) J. Infect. Dis., **152**, 64–73.
23. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
24. MacDonald,R.J., Swift,G.H., Przybyla,A.E. and Chirgwin,J.M. (1987) Methods Enzymol., **152**, 219–227.
25. Sternberg,N.L. (1990) Genetic Analysis: Techniques and Applications, **7**, 126–132.
26. Sanger,F.S., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA, **80**, 5463–5467.
27. Geliebter,J. (1987) Focus, **9**, 5–8.
28. Devereux,J., Haeberli,P. and Smithies,O. (1984) Nucleic. Acids. Res., **12**, 387–395.
29. Feinberg,A.P. and Vogelstein,B. (1983) Anal. Biochem., **132**, 6–13.
30. Borst,P. and Greaves,D.R. (1987) Science, **235**, 658–667.
31. Lyon,M.F. (1993) Trends Genet., **9**, 123–128.
32. Gommers-Ampt,J., Lutgerink,J. and Borst,P. (1991) Nucleic. Acids. Res., **19**, 1745–1751.
33. Van der Ploeg,L.H., Gottesdiener,K. and Lee,M.G. (1992) Trends Genet., **8**, 452–457.
34. Zomerdijk,J.C., Kieft,R., Duyndam,M., Shiels,P.G. and Borst,P. (1991) Nucleic. Acids. Res., **19**, 1359–1368.
35. Orr-Weaver,T.L. and Szostak,J.W. (1985) Microbiol. Rev., **49**, 33–58.
36. Adam,R.D., Nash,T.E. and Wellems,T.E. (1991) Mol. Cell Biol., **11**, 3326–3330.
37. Le Blancq,S.M., Korman,S.H. and Van der Ploeg,L.H. (1991) Nucleic. Acids. Res., **19**, 4405–4412.
38. Le Blancq,S.M., Korman,S.H. and Van der Ploeg,L.H. (1992) Nucleic. Acids. Res., **20**, 4539–4545.
39. Murtagh,J.J.,Jr., Mowatt,M.R., Lee,C.M., Lee,F.J., Mishima,K., Nash,T.E., Moss,J. and Vaughan,M. (1992) J. Biol. Chem., **267**, 9654–9662.