Characterization of the subtelomeric regions of *Giardia lamblia* genome isolate WBC6

Anjali Prabhu[a,d], Hilary G. Morrison[b], Charles R. Martinez III[a,e], Rodney D. Adam[c,*]

[a]*Dept. of Immunobiology, University of Arizona College of Medicine, 1501 N. Campbell, Tucson, AZ 85724-5049 USA*

[b]*Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543-1015 USA*

[c]*Departments of Medicine and Immunobiology, University of Arizona College of Medicine, 1501 N. Campbell, Tucson, AZ 85724-5049 USA*

[d]*Current address, Department of Microbiology and Immunology, 1638 N.W. 10th Ave, Room 727 McKinght Building, Miami, FL 33136 USA*

[e]*Current address, Department of Biological Sciences, Box 1634 Station B, Vanderbilt University, Nashville, TN 37235 USA*

*Corresponding author.
Rodney D. Adam, Department of Immunobiology, University of Arizona, 1501 N. Campbell, Tucson, AZ 85724-5049 USA

Tel.: +1 520 626 6430; fax: +1 520 626 2100.
E-mail address: adamr@u.arizona.edu

Abstract

Giardia trophozoites are polyploid and have five chromosomes. The chromosome homologues demonstrate considerable size heterogeneity due to variation in the subtelomeric regions. We used clones from the genome project with telomeric sequence at one end to identify six subtelomeric regions in addition to previously identified subtelomeric regions, to study the telomeric arrangement of the chromosomes. The subtelomeric regions included two retroposons, one retroposon pseudogene, and two *vsp* genes, in addition to the previously identified subtelomeric regions that include ribosomal DNA repeats. The presence of *vsp* genes in a subtelomeric region suggests that telomeric rearrangements may contribute to the generation of *vsp* diversity. These studies of the subtelomeric regions of *Giardia* may contribute to our understanding of the factors that maintain stability, while allowing diversity in chromosome structure.

Keywords: Giardia lamblia; Telomere; Subtelomeric region; *vsp*; Retroposon

1. Introduction

Giardia lamblia is a flagellated unicellular eukaryotic protozoan parasite that is a common cause of diarrheal disease worldwide. The life cycle of *Giardia* consists of two stages, the infectious cyst and the vegetative trophozoite. Infection occurs when the host ingests the environmentally stable cyst via a contaminated water or food source. After passage through the stomach, the cysts excyst into trophozoites, which multiply in the proximal small intestine. A few trophozoites encyst in the jejunum after exposure to biliary fluid and are excreted via the feces, thus completing the life cycle.

Giardia trophozoites have two nuclei that are both transcriptionally active (Kabnick and Peattie, 1990) and replicate at approximately the same time (Wiesehahn et al., 1984). Trophozoites are polyploid and have five chromosomes ranging in size from approximately 1.6 to 3.8 Mb, although some isolates have size variants of chromosome 1 as small as 1.1 Mb (Adam et al., 1988). This adds up to a haploid genome size of approximately 10-12 Mb, a value which is supported by the results of the genome project (www.mbl.edu/Giardia). Each nucleus has at least one copy of each chromosome (Yu et al., 2002) and the two nuclei have approximately equal quantities of DNA. Therefore, since trophozoites have an estimated ploidy of four (Bernander et al., 2001), it is likely that each nucleus has a ploidy of approximately two. However, the exact number of copies of each chromosome in each nucleus may vary somewhat, as suggested by a cytogenetic evaluation in which the number of chromosomes per nucleus varied between nine and 11 (Tumova et al., 2006), leaving open the possibility of a small degree of aneuploidy.

Previous studies of the *Giardia* chromosomes have demonstrated substantial size variation of chromosome 1 (Adam et al., 1988, 1991; Adam, 1992; Le Blancq et al., 1991a, 1992, 1991b; Hou et al., 1995). These analyses of chromosome 1 demonstrated a core region that was highly similar for all copies of the chromosome and substantial variation in the subtelomeric regions (STRs). Ribosomal

DNA (rDNA) repeats were found adjacent to the TAGGG telomeric repeat at one end of chromosome 1 and accounted for a substantial portion of the size heterogeneity for the different chromosome 1 homologues (Adam, 1992; Hou et al., 1995). Different *Giardia* isolates demonstrate substantial differences in the chromosomal locations of the rDNA repeats (Adam et al., 1991), even for different cloned lines of the same isolate (Adam and Prabhu, unpublished observations). In addition to being present as repeating units, the rRNA genes may be present in subtelomeric regions as rearranged or incomplete genes (Upcroft et al., 2005).

In addition to rDNA, other STRs have been reported. A chromosome 1 STR (p4e) was cloned on the basis of having the telomeric repeat at one end (Hou et al., 1995). Clones have been identified that have telomeric repeats at one end and a retroposon sequence at the other (Arkhipova and Morrison, 2001). Three different retroposons were identified, GilM, GilT and GilD. All three belong to long interspersed nuclear element (LINE)-like families. GilM and GilT have intact reverse transcriptase open reading frames (ORFs) and are candidates for functional genes, but whether they are expressed has not yet been published. On the other hand, GilD has multiple deletions, stop codons and frameshifts, so it is presumably a pseudogene.

A telomeric *vsp* gene was identified in the same study. In addition, previous mapping studies had shown that the *vsp*s, CRP65 and CRP136, hybridized to *Xba*I fragments that also contained the telomeric repeat sequence (Upcroft et al., 1997; Chen et al., 1995, 1996), but in these mapping studies, rDNA and ankyrin were found between the *vsp* and the telomeric repeat. It is of interest to know if the *vsp*s are telomeric, in view of the fact that surface antigen genes of other protists may be expressed from subtelomeric locations.

In the current study, we have used the telomeric clones from the genome project to determine the chromosomal organization of the STRs of the genome isolate WBC6.

2. Materials and methods

2.1. Growth and cloning of Giardia isolates

The genome isolate, WB clone C6 (ATCC #50803), was the major isolate used for this study (McArthur et al., 2000; Gillin et al., 1990). Trophozoites were grown in modified TYI-S-33 medium (Keister, 1983). Cloning was performed by limiting dilution as previously described (Nash et al., 1988).

2.2. Identification and directed sequencing of telomeric clones

Telomeric clones that were sequenced as part of the *Giardia* genome project were chosen for further analysis. The plasmid clones used for this project were obtained from small insert plasmid libraries with inserts generated by partial enzyme digestion or by random shearing of the DNA (*Giardia lamblia* Genome Project website: www.mbl.edu/Giardia). Since chromosome-internal telomeric sequences also occur, we surveyed the assembled sequence for the presence of TAGGG telomeric repeats. The longest chromosome-internal telomeric sequence was 18 bp. Therefore, we limited our analysis to clones with at least four copies of the telomeric repeat at one end of the clone. Previous studies have reported the telomeric location of rDNA. Therefore, we limited our analysis to clones with a sequence other than rDNA at the nontelomeric end.

For each subtelomeric region, we chose at least one clone for complete bidirectional sequencing which was obtained by primer walking. However, for some clones, the read from the telomeric end did not go beyond the telomeric repeat. For those clones, we were only able to sequence

a single direction for a small portion of the telomere transition region adjacent to the telomeric repeats. When more than one clone was available for a subtelomeric region, we obtained a complete single pass sequence for the additional clone(s). The GenBank accession numbers are shown in Table 1.

2.3. Pulsed Field Gel Electrophoresis (PFGE)

Pulsed Field Gel Electrophoresis (PFGE) of total *Giardia* DNA was performed as described (Adam et al., 1988) using OFAGE (orthogonal field alternation gel electrophoresis) with 1% Seakem GTG agarose (Cambrex Bioscience, Rockland, ME) and switching intervals of 11 min at 80 V for 2 days, 25 min at 65 V for 2 days and 60 min at 50 V for 3 days. When Seakem Gold agarose (Cambrex Bioscience, Rockland, ME) was used, 0.8% gels were run with switching intervals of 6 min at 80 V for 1 day, 11 min at 65 V for 1 day and 25 min at 50 V for 3 days. The separated DNA was either transferred to Nytran (Schleicher and Schuell, Keene, NH) by alkaline transfer, or individual chromosomal bands were removed for restriction enzyme digestion. The individual chromosomes were cut into blocks and enzymatically digested following equilibration in the appropriate buffer. After equilibration, 200 µl of fresh buffer along with 20 U of enzyme was added and the blocks were incubated overnight at 4^0C. The tubes were then brought to the appropriate digestion temperature for 2 h, after which 20 U of enzyme was added for an additional 2 h of digestion. The digested blocks were separated by CHEF (Contour-clamped homogeneous electric field) in 1% GTG agarose for 22 h at 120 V with a switching interval time of 20 s and transferred to Nytran by alkaline transfer.

2.4. Nucleic acid isolation and hybridization

Plasmid DNA was prepared using the QIAprep Spin Miniprep Kit (QIAGEN Incorporation, Valencia, CA). DNA was labeled by random priming and excess nucleotide removed by spun G50

column chromatography. Hybridization was performed in 5 X sodium chloride-sodium citrate (SSC) with 1% SDS at 50^0C and washing was performed in 0.1 X SSC at 65^0C. Total RNA was extracted using the RNeasy mini kit Yeast protocol I from QIAGEN (QIAGEN Incorporation, Valencia, CA) with minor modifications.

2.5. RT-PCR

RT-PCR was carried using the Gene Amp kit (Applied Biosystems, Foster city, CA). The reaction was carried out at 60^0C for 30 min, then 94^0C for 1 min, followed by 40 cycles of denaturation (94^0C for 15 s) and annealing/extension (60^0C for 30 s), then a terminal incubation at 60^0C for 7 min. Controls included RNA treated with either RNase or RNase plus DNase. The primers used for GilM were: coding F- TATTTCACAGAGATGGCGG, coding R- TCTCCTGCACTGCGATAG, 3' untranslated region (UTR) F- AATTTGGGCCTGAAGACC, and 3'UTR R- CTGCAGGCTGTGTGTTAG. The primers used for GilT were: coding F- CTGGTGAGGGAGCTGAGG, coding R- AGCGAGAAGATGATCACC, 3' UTR F- TCTCGTTTGGCTGCCTTC, and 3'UTR R- GTGGCGTCTGCATGCTTG.

2.6. Hybridization probes

Probes used for the hybridization studies are indicated in Table1. They were obtained by enzymatic digestion or PCR of the telomeric clones used in this study. The telomere probe consists of five TAGGG repeats.

3. Results

3.1. Characterization of the subtelomeric regions (STRs)

We identified 14 clones with a telomere sequence at one end and a nontelomere sequence other than rDNA at the other end (Table 1). The telomeric clones contained from 30 to > 406 bp of telomeric repeat sequence (Table 1). These 14 clones consist of six STRs that include *vsp*s (STRs 1 and 2) and retroposons (STRs 3, 4 and 5). The portion of STR6 that was used as a probe had no identified BLAST hit, but had an rDNA sequence located between the unknown sequence and the telomere. In addition to the six STRs we identified, two other STRs have been identified in previous reports from other Genotype A-1 isolates by the identification of clones with telomeric repeats at one end and nontelomeric sequence at the other end ((rDNA; STR7 (Adam et al., 1991)) or (p4e; STR8 (Hou et al., 1995)). STRs 7 and 8 co-localized with telomeric fragments on restriction digests of chromosomal DNA and are discussed below. We also evaluated CRP136 and CRP65 to determine whether they were subtelomeric in WBC6, but neither co-localized to telomeric fragments of WBC6 (data not shown), so they were not further evaluated.

STR1 contains an ORF encoding a *vsp* gene that is oriented towards the telomere and is incomplete at the 3' end because of its transition into the telomere repeat region (Fig. 1). The clone containing STR1 was also missing the *vsp* gene start site. The 5' end of STR1 (1-1,195 bp) gave the best match with *vsp*S7 (95% nucleotide identity) (accession no. **AY142128**) (Slavin I, Touz M.C, Lujan H.D.). However, the portion closer to the 3' end (835-2,198 bp) gave the best match to *vsp*M20 (92% identity) (accession no **AF298864**) (Mansouri M, Ey P.L.). Thus, STR1 may represent a new *vsp* gene with regions of similarity to *vsp*M20 and *vsp*S7 or may have been generated from these or similar genes by subtelomeric recombination.

STR2 contains two *vsp* genes arranged in a head–to-tail fashion, oriented towards the telomere (Fig. 1). The *vsp* genes of STR2 have the greatest similarity to *vsp*MM (97% identity) (accession no. **AF235029**) (Mansouri M, Ey P.L.). The clone containing *vsp*1 of STR2 did not include its start site but

included the highly conserved 3' region (Mowatt et al., 1991), which was followed by an intergenic region of 243 bp, then the start site of *vsp*2. The ORF of *vsp*2 was disrupted by the abrupt transition into the telomere repeat region after 183 bp.

STR3 contains the GilM retroposon gene (accession no.**AF433875**), which has a 125 bp 5' UTR, a 3,206 bp reverse transcriptase ORF and a 2,150 bp 3' UTR (Arkhipova and Morrison, 2001). We identified four plasmid clones containing the GilM gene at one end and the telomeric repeat at the other (Fig. 1). The four plasmid clones have three different regions at which the STR region transitions into a telomere repeat region. Two of the transition points occurred in the reverse transcriptase coding region while the third occurred in the 3' UTR. These three different transition types may be from different chromosomes, from different homologues of the same chromosome or may represent chromosome length heterogeneity that has developed since the WBC6 isolate was initially cloned. The GilM sequence hybridizes to telomeric restriction fragments on chromosomes 5 and 2/1b and does not hybridize to any chromosome-internal fragments. These results are consistent with previous reports that all copies of the GilM gene are present as tandem repeats in a telomeric location.

In order to determine whether the GilM gene is transcribed, we used RT-PCR with two different primer sets, one to identify the coding region and another to identify the 3' UTR. The RT-PCR reaction gave products of the expected sizes with the DNAse-treated samples and no product with the DNAse + RNAse-treated samples, in both amplifications (Fig. 2). The sequences of these products were identical to the published GilM sequence, confirming the expression of GilM.

STR4 contains the GilT retroposon gene (accession no. **AF433876**) (Arkhipova and Morrison, 2001), which has a 124 bp 5' UTR, 3,230 bp reverse transcriptase ORF, a 2,644 bp 3' UTR and is also oriented away from the telomere (Fig. 1). The four plasmid clones with GilT at one end and the telomeric repeat at the other, each represent different transition types, one within the 5' UTR and the

other three within the reverse transcriptase ORF, although the transition points of two of the clones (KJ1356 and NJ1197) differ by only 46 bp. The GilT probe hybridized to only one end of chromosome 3, so each of the four transition variants represents size variants of that end of chromosome 3. They could represent different chromosome homologues or size variation that accumulated since the cloning of the WBC6 isolate.

We used RT-PCR with two separate sets of primers to separately amplify portions of the ORF and the 3' UTR from WBC6 total RNA (Fig. 2). The sequences of these products were identical to the published sequences, confirming the expression of the GilT gene.

STR5 contains no ORF on GenBank matches in the initial 790 bp (Fig. 1). The following 60 bp is 95% identical to the 5' upstream region of *vsp* AS3 (accession no **AY142132**) (Slavin I, Touz M.C., Lujan H.D.). The region from 1,679-1,921 bp is 96% identical to the RT/Endonuclease portion of the GilD pseudogene (accession no **AF433877**) (Arkhipova and Morrison, 2001), and the region from 2,295-2,494 bp is identical to the 5' UTR of GilM.

STR6 has no matches to the Genbank nonredundant database in the initial 2,895 bp and has no ORFs greater than 200 bp in length (Fig. 1). The STR6 sequence then transitions into the small subunit rRNA gene, followed by the telomere repeat region. Thus, even though the strategy for identification of telomeric clones was designed to eliminate telomeric clones with rDNA sequences, this clone contained a portion of the rRNA gene located between the telomeric repeat and an additional subtelomeric region.

3.2. Determination of the chromosomal locations of the STRs

We initially hybridized each of the eight STRs to PFGE separations of *Giardia* chromosomes to assign them to individual chromosomes (Fig. 3). Certain STRs were only present on a single

chromosome (STRs 1, 2, 4, 5 and 6), while the others were present on two or more chromosomes (Table 1). Each STR was then hybridized to restriction digests of each chromosome to which it hybridized on the PFGE separation. These studies were performed using chromosome-specific DNA digested with rare cutting enzymes (eight base recognition sites) such as *Not*I, *Pme*I and *Swa*I so that we could assign the STR to a specific chromosome end and determine whether the STR was only telomeric or was also found in a chromosome-internal location. STR4 (GilT) and STR7 (rDNA) have *Not*I enzyme sites within their sequences, so we used *Pme*I and *Swa*I for these STRs.

3.3. Chromosome 5

The telomere probe identified three chromosome 5 *Not*I fragments, 20, 23 and 48 kb, in addition to a broad band about 4-6 kb in size (Fig. 4). The presence of more than two telomeric *Not*I fragments suggested either size heterogeneity among the different homologues of chromosome 5, as already demonstrated for chromosome 1 (Adam, 1992), or that WBC6 had developed population heterogeneity with chromosome size heterogeneity during in vitro passage since its initial cloning. To distinguish between these two possibilities, we cloned the WBC6 line twice by limiting dilution and hybridized the telomeric repeat to *Swa*I and *Pme*I digests of chromosome 5 (and also chromosome 4). The results were the same as those obtained before cloning (data not shown), suggesting that the multiple bands represented size variation of the homologues of chromosome 5 rather than population heterogeneity within the WBC6 trophozoite culture used for constructing the genome library. (We note that an alternative explanation of two co-migrating chromosomes would not fit with the known genome size of 11-12 MB, since the five chromosomes add up to a genome size of about 12 Mb).

We hybridized the telomere probe and STRs 1, 2, 3 and 7 to *Not*I and *Pme*I, digests of chromosome 5. STRs 1 and 2 had nearly identical hybridization patterns, co-localizing to the 23 and 48

kb *Not*I fragments and to larger nontelomeric *Not*I fragments. The hybridization to these nontelomeric fragments may indicate a chromosome-internal copy of this *vsp* gene or a copy(s) similar enough to allow cross-hybridization. STR3 (GilT) hybridized to the 23 and 48 kb telomeric *Not*I fragments recognized by the *vsp*-containing STRs and also to a 20 kb telomeric *Not*I fragment, but to no chromosome-internal sites.

Since STR7 (rDNA) contains two *Not*I sites, we digested chromosome 5 DNA with *Pme*I, an eight-base cutter with no site in rDNA. STR1, 2 and 3 all hybridized to a 290 kb telomeric fragment. STR7 hybridized to 23 and 60 kb telomeric fragments in addition to a 290 kb fragment (Fig. 4). These observations taken along with the *Not*I data suggest that STRs1, 2 and 3 are all found on one end of the chromosome, but that STRs 1 and 2 are deleted from the 20 kb fragment, while STR7 is present on the other end of the chromosome. The additional hybridization of STR7 to a 290 kb *Pme*I fragment may indicate that rDNA is on both chromosome ends, but would not explain the *Not*I data since rDNA limits the size of the telomeric *Not*I fragments because of the internal *Not*I sites. Therefore, we believe the hybridization to a 290 kb *Pme*I fragment (which is in the compression region of the gel) is due either to chromosome-internal rDNA or is the result of incomplete digestion. An excess number of large fragments was obtained with the *Swa*I digest, suggesting that incomplete digestion had also occurred with this enzyme. Therefore, the *Swa*I digest is not shown. A map of the subtelomeric regions of chromosome 5 is shown in Fig. 5 with STR7 on only one end.

3.4. Chromosome 4

STR7 was the only STR that hybridized to chromosome 4 (Fig. 6). The *Not*I digest indicated four telomeric fragments, 290, 48, 23 and a broad 6 kb fragment that most likely represents the rDNA end. The telomere probe identified two *Pme*I fragments, a 340 kb fragment that also hybridized to STR7 and a 15 kb fragment that did not hybridize to an STR. (The 340 kb *Pme*I fragment could

represent a size range of fragments since it is in the compression region of the gel). STR7 hybridized to the 136 and 36 kb *Swa*I telomeric fragments, but not to the 20 kb fragment. The subtelomeric map of chromosome 4 could be constructed a number of ways, but we assumed that the rDNA-containing sequence was found on one end while another unidentified STR(s) was on the other side (Fig. 5).

3.5. Chromosome 3

STR4 hybridized only to chromosome 3 and was the only STR we identified for chromosome 3 (Fig. 6). Although STR7 hybridized to chromosome 3, it did not hybridize to any telomeric fragments of chromosome 3 (data not shown). The telomeric probe identified two *Pme*I fragments (290 and 180 kb) and two *Swa*I fragments (97 and 48 kb). STR4 hybridized to the 290 kb *Pme*I and 97 kb *Swa*I fragments and did not hybridize to any nontelomeric *Pme*I or *Swa*I fragments, suggesting that all copies of STR4 are found within 97 kb of the end of chromosome 3. The telomere probe identified a 48 kb and a broad 6 kb *Not*I fragment. The broad 6 kb band likely includes STR4, which contains a *Not*I site. The ~40 kb *Not*I fragment identified by the STR4 probe is probably adjacent to the telomeric fragment since the STR4-containing *Pme*I and *Swa*I fragments are telomeric and because we know from genomic sequence data that the copies of the GilT gene are in tandem array in a telomeric location. The hybridization data indicate little or no chromosome size heterogeneity for chromosome 3, with only two telomeric fragments identified by the three enzymes. The fact that the four STR4-containing telomere clones had transition points that differed by less than 1.5 kb also supports the proposal that chromosome 3 demonstrates little size heterogeneity.

3.6. Chromosomes 1 and 2

The STR3, STR5, STR6, STR7 and STR8 probes hybridized to band 2/1b (Fig. 3), which

contains chromosome 2 as well as the larger size variant of chromosome 1 (1b). STRs 7 and 8 also hybridized to the smaller size homologue of chromosome which is designated as 1a. The simplest interpretation is that STRs 3, 5 and 6 are on chromosome 2, while STRs 7 and 8 are on chromosome 1. This would also fit with previous reports describing p4E at one end (Hou et al., 1995) and rDNA (Adam, 1992; Hou et al., 1995) at the other end of chromosome 1a on other isolates from the same genotype as WB. The four *Not*I telomeric fragments are the expected results for two co-migrating chromosomes (Fig. 7). Only three *Pme*I and two *Swa*I telomeric fragments were identified, which could indicate co-migrating telomeric fragments or fragments so small that they were not retained on the gel. With the exception of STR8, the STRs hybridized only to telomeric fragments. STR8 hybridized only to telomeric *Pme*I fragments, but hybridized to a 160 kb *Swa*I fragments in addition to the two telomeric *Swa*I fragments. The reason for the hybridization to a nontelomeric fragment is not known, but could be explained if there were a *Swa*I site between STR8 and the telomeric repeats. The patterns were complex enough that we were unable to compile a tentative map for chromosomes 1b or 2.

4. Discussion

The genome of *Giardia* has a number of unique features that make the chromosomal organization and subtelomeric organization especially interesting. The five chromosomes range in size from 1.6 to 3.8 Mb. Trophozoites are tetraploid and have two nuclei with approximately equal amounts of DNA, implying that each nucleus is diploid. However, it remains possible that some degree of aneuploidy exists. Giardia has been assumed to be asexual since sexual reproduction has never been documented and population genetic studies have suggested clonal replication. However, the

assumption of strictly asexual reproduction has been called into question by the identification in the genome of the genes required for meiotic replication (Ramesh et al., 2005). In addition, substantial differences are expected to accumulate among the chromosome homologues in asexual organisms with a ploidy of two or higher, as has been shown for bdelloid rotifers (Welch and Meselson, 2000). The assembled portion of the *Giardia* genome stands in marked contrast to these expectations with its remarkably low level of sequence heterozygosity (< 0.002%). Thus, in the assembled portion of the genome, there is very little difference between chromosome homologues for the assembled portion of the genome. However, the assembled portion of the genome does not include the telomeric regions, and it is these regions that demonstrate substantial variation as demonstrated in this and in previous studies.

The current study has identified at least one STR for each of the five chromosomes. The STRs with identified function fall into three categories; *vsp*, retroposon and rDNA. Previous studies have dealt primarily with rDNA, which demonstrates great variability both in terms of chromosomal location and in the degree of associated chromosome size polymorphism (Adam et al., 1991, 1992). The most remarkable example is the size variation of chromosome 1 in which single homologues from a single cloned isolate of the ISR isolate (same genotype as WB) varied from 1.1 to 1.9 Mb (Adam, 1992). About 30% of this difference could be attributed to differences in the number of copies of the rDNA repeat. The current study identified marked STR variability for chromosomes 4 and 5, although at a lesser level than that previously reported for chromosome 1.

Variation in the STRs also accounts for much of the chromosome size heterogeneity found in other eukaryotic microbes. For example, size variation of chromosome 1 homologs in *Leishmania major* Friedlin is due to variation in the content of STRs (Sunkin et al., 2000). A large extent of the size polymorphism seen in the diploid chromosome homologs of *Trypanosoma brucei* is due to

expansions and contractions of the STRs, although some of the polymorphism is also due to variation in the chromosome internal regions (Melville et al., 1999). Restriction maps of chromosomes 1 and 2 of six cloned isolates of *Plasmodium falciparum* show a conserved central region with length and sequence polymorphism at the subtelomeric region (Corcoran et al., 1988). In *Plasmodium berghei*, a 400 kb variation of chromosome 7 homologues is due to a difference in the copy number of a 2.3 kb tandem repeat family of the subtelomeric region (Ponzi et al., 1990).

In contrast with the other chromosomes, chromosome 3 demonstrates relatively little size variation. This is demonstrated by the finding that only two telomeric restriction fragments were identified by each of three rare-cutting restriction enzymes. The four chromosome 3 telomeric clones all had transitions from GilT to telomeric sequence within a range of less than 1.5 kb. It is possible that this 1.5 kb region represents the extent of size variation for the GilT end of chromosome 3. GilM had a somewhat more variable position, being found at the chromosome 5, and 1 and/or 2 STRs. In addition, GilM sequence was found interspersed with *vsp* and GilD sequence in a clone that hybridized to chromosomes 1 and/or 2. The retroposons found in most organisms are dispersed throughout the genome, although the telomeric location of retroposons in *Drosophila* provides an important exception. GilT and GilM are in tandem arrays as demonstrated by sequence linkage (Arkhipova and Morrison, 2001), a finding that is supported by our hybridization studies demonstrating hybridization only to a telomeric restriction fragment. GilM and GilT have intact reverse transcriptase ORFs and long 3' UTRs, 2,150 bp and 2,644 bp in length, respectively. These long 3' UTRs are in marked contrast to most *Giardia* genes, which have very short 3' UTRs. Our studies using RT-PCR have shown that the ORFs as well as the 3' UTR are expressed. We would note that the telomeric clones of GilM and GilT that we used in the current studies are pseudogenes in that their 5' regions including part of the coding

region of the RT are absent, so the expression presumably comes from the adjacent copies in the genome.

The telomeric location of *vsp* genes in *Giardia* is especially interesting because of examples in other eukaryotic microbes where variant surface antigen genes are located in telomeric regions. The VSG genes of African trypanosomes may be located in telomeric or chromosome-internal positions. However, a telomeric location is required for VSG expression, so a chromosome-internal VSG gene can be expressed only when a copy of the gene is duplicated to a telomeric expression site (Borst and Ulbert, 2001). On the other hand, the *var* genes of *P. falciparum* are found in subtelomeric or chromosome-internal sites and may be activated in situ from either location (Deitsch, 2005). The chromosomal locations have been correlated with expression status for only two *Giardia vsp* genes (Yang and Adam, 1994; Yang et al., 1994). In both cases, the genes were activated or inactivated in situ from chromosome-internal sites. It is not yet known how many *vsp* genes are located near telomeres or whether expression can occur from telomeric regions. The two *vsp*-containing STRs both hybridized to chromosome 5. Our data suggest that they are both found on the same end of the chromosome. If this is true, they are likely to be from different chromosome 5 homologues. The *vsp* genes were oriented toward the telomere and the telomere transition occurred before the highly conserved C-terminus. Therefore, these genes are incomplete. However, for STR2, a second *vsp* gene terminated 243 bp upstream of the telomeric *vsp* gene. How the expression of this second *vsp* gene is affected by its telomeric location is unknown.

An increased understanding of the subtelomeric regions will facilitate determination of the roles of the associated genes in maintaining chromosome stability. It will be instructive to determine whether the telomeric role of certain *vsp* genes plays a role in expression or in generation of the *vsp* gene repertoire. It will also be interesting to know whether the GilT retroposon at the end of

chromosome 3 contributes to its very low size heterogeneity in comparison with the other chromosomes. The complexity of these questions is increased by the two nuclei and polyploid nature of the genome in addition to the newly opened question of sexuality in *Giardia*.

Acknowledgements

This work was supported in part by NIH grant AI43273 to Mitchell L. Sogin. Additional support was provided by the G. Unger Vetlesen Foundation and LI-COR Biotechnology. Sequences that are part of the genome project can be accessed at the web site: www.mbl.edu/*Giardia*.

Refernces

Adam, R. D., 1992. Chromosome-size variation in *Giardia lamblia*: the role of rDNA repeats. Nucleic Acids Res., 20. 3057-61

Adam, R. D., Nash, T. E., Wellems, T. E., 1988. The *Giardia lamblia* trophozoite contains sets of closely related chromosomes. Nucleic Acids Res., 16. 4555-67

Adam, R. D., Nash, T. E., Wellems, T. E., 1991. Telomeric location of *Giardia* rDNA genes. Mol. Cell. Biol., 11. 3326-30

Arkhipova, I. R., Morrison, H. G., 2001. Three retrotransposon families in the genome of *Giardia lamblia*: two telomeric, one dead. Proc. Natl. Acad. Sci. USA, 98. 14497-502

Bernander, R., Palm, J. E., Svärd, S. G., 2001. Genome ploidy in different stages of the *Giardia lamblia* life cycle. Cell. Microbiol., 3. 55-62

Boothroyd, J. C., Wang, A., Campbell, D. A., Wang, C. C., 1987. An unusually compact ribosomal DNA repeat in the protozoan *Giardia lamblia*. Nucleic Acids Res., 15. 4065-84

Borst, P., Ulbert, S., 2001. Control of VSG gene expression sites. Mol. Biochem. Parasitol., 114. 17-27

Chen, N., Upcroft, J. A., Upcroft, P., 1995. A *Giardia duodenalis* gene encoding a protein with multiple repeats of a toxin homologue. Parasitology., 111. 423-31

Chen, N., Upcroft, J. A., Upcroft, P., 1996. A New cysteine-rich protein-encoding gene family in *Giardia duodenalis*. Gene., 169. 33-8

Corcoran, L. M., Thompson, J. K., Walliker, D., Kemp, D. J., 1988. Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. Cell., 53. 807-13

Deitsch, K. W., 2005. - Malaria virulence genes controlling expression through chromatin modification. - Cell, 121. 1-2

Edlind, T. D., Chakraborty, P. R., 1987. Unusual ribosomal RNA of the intestinal parasite *Giardia lamblia*. Nucleic Acids Res., 15. 7889-901

Gillin, F. D., Hagblom, P., Harwood, J., Aley, S. B., Reiner, D. S., McCaffery, M., So, M., Guiney, D. G., 1990. Isolation and expression of the gene for a major surface protein of *Giardia lamblia*. Proc. Natl. Acad. Sci. USA, 87. 4463-7

Hou, G., Le Blancq, S. M., Yaping, E., Zhu, H., Lee, M. G., 1995. Structure of a frequently rearranged rRNA-encoding chromosome in *Giardia lamblia*. Nucleic Acids Res., 23. 3310-7

Kabnick, K. S., Peattie, D. A., 1990. In situ analyses reveal that the two nuclei of *Giardia lamblia* are equivalent. J. Cell Sci., 95. 353-60

Keister, D. B., 1983. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. Trans. R. Soc. Trop. Med Hyg., 77. 487-8

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

Le Blancq, S. M., Korman, S. H., Van der Ploeg, L. H., 1991b. Frequent rearrangements of rRNA-encoding chromosomes in *Giardia lamblia*. Nucleic Acids Res., 19. 4405-12

Le Blancq, S. M., Korman, S. H., Van der Ploeg, L. H., 1992. Spontaneous chromosome rearrangements in the protozoan *Giardia lamblia*: Estimation of mutation rates. Nucleic Acids Res., 20. 4539-45

McArthur, A. G., Morrison, H. G., Nixon, J. E. J., Passamaneck, N. Q. E., Kim, U., Hinkle, G., Crocker, M. K., Holder, M. E., Farr, R., Reich, C. I., Olsen, G. E., Aley, S. B., Adam, R. D., Gillin, F. D., Sogin, M. L., 2000. The *Giardia* genome project database. FEMS Microbiol. Lett., 189. 271-3

Melville, S. E., Gerrard, C. S., Blackwell, J. M., 1999. Multiple causes of size variation in the diploid megabase chromosomes of African tyrpanosomes. Chromosome Res., 7. 191-203

Mowatt, M. R., Aggarwal, A., Nash, T. E., 1991. Carboxy-terminal sequence conservation among variant-specific surface proteins of *Giardia lamblia*. Mol. Biochem. Parasitol., 49. 215-27

Nash, T. E., Aggarwal, A., Adam, R. D., Conrad, J. T., Merritt, J. W., Jr., 1988. Antigenic variation in *Giardia lamblia*. J. Immunol., 141. 636-41

Ponzi, M., Janse, C. J., Dore, E., Scotti, R., Pace, T., Reterink, T. J., van der Berg, F. M., Mons, B., 1990. Generation of chromosome size polymorphism during in vivo mitotic multiplication of *Plasmodium berghei* involves both loss and addition of subtelomeric repeat sequences. Mol. Biochem. Parasitol., 41. 73-82

Ramesh, M. A., Malik, S. B., Logsdon, J. M., Jr., 2005. - A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. - Current Biology, 15. 185-91

Sunkin, S. M., Kiser, P., Myler, P. J., Stuart, K., 2000. The size difference between leishmania major friedlin chromosome one homologues is localized to sub-telomeric repeats at one chromosomal end. Mol. Biochem. Parasitol., 109. 1-15

Tumova, P., Hofstretrova, K., Nohýnková, E., Hovorka, O., Kral, J., 2006. Cytogenetic evidence for diversity of two nuclei within a single diplomonad cell of *Giardia*. Chromosoma, in press.

Upcroft, J. A., Abedinia, M., Upcroft, P., 2005. Rearranged subtelomeric rRNA genes in *Giardia duodenalis*. Euk. Cell, 4. 484-6

Upcroft, P., Chen, N., Upcroft, J. A., 1997. Telomeric organization of a variable and inducible toxin gene family in the ancient eukaryote *Giardia duodenalis*. Genome Res., 7. 37-46

Welch, D. M., Meselson, M., 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science., 288. 1211-5

Wiesehahn, G. P., Jarroll, E. L., Lindmark, D. G., Meyer, E. A., Hallick, L. M., 1984. *Giardia lamblia*: Autoradiographic analysis of nuclear replication. Exp. Parasitol., 58. 94-100

Yang, Y. M., Adam, R. D., 1994. Allele-specific expression of a variant-specific surface protein (VSP) of *Giardia lamblia*. Nucleic Acids Res., 22. 2102-8

Yang, Y. M., Ortega, Y., Sterling, C., Adam, R. D., 1994. *Giardia lamblia* trophozoites contain multiple alleles of a variant-specific surface protein gene with 105-base pair tandem repeats. Mol. Biochem. Parasitol., 68. 267-76

Yu, L. Z., Birky, C. W., Jr., Adam, R. D., 2002. The two nuclei of *Giardia* each have complete copies of the genome and are partitioned equationally at cytokinesis. Euk. Cell, 1. 191-9

Figure legends

Fig. 1. Subtelomeric regions (STR) 1 through 6 are demonstrated as alignments of the individual STR clones with each other and with the relevant *Giardia* genes. The individual clone names are on the left. The bold black lines represent the clone/clones that make up each STR and the dashed lines represent telomere regions. STRs 7 (rDNA; (Edlind and Chakraborty, 1987; Boothroyd et al., 1987; Adam et al., 1991; Le Blancq et al., 1991b)) and 8 pE4 (Hou et al., 1995) have previously been described and are not demonstrated in this figure. The hashed line indicates the orientation of gene/genes or portions of genes found on the STR. The probes used in the hybridization experiments are also indicated.

Fig. 2. RT- PCR performed on *Giardia* total RNA using primers that amplify the open reading frame (ORF) and 3' untranslated region (UTR) of GilT and GilM genes. The ethidium-stained gel shows the RT-PCR products of the ORF and 3' UTR for GilT and GilM. For each amplification, the DNAse + RNAse-treated sample (negative control) yielded no detectable product, indicating that the RT-PCR product was derived from RNA and not from contaminating DNA. The identities of the RT-PCR products were verified by DNA sequencing.

Fig. 3. Hybridization of probes for subtelomeric regions (STRs) 1-8 to Pulse Field Gel Electrophoresis separations of WBC6 chromosomes. The chromosome numbers of hybridizing chromosomes are indicated on the right. Chromosomes 2 and the larger homologue of chromosome 1 co-migrate, so that band is labeled as 2/1b. The smaller homologue of chromosome 1 is labeled as 1a. The type of sequence found in the STR is inserted below, except for STRs 5 and 6. STR 5 included a portion with no GenBank match, and small portions with vsp, GilD and GilM sequence. STR 6 had a small region

of rDNA sequence and a larger region with no match. The probes used for hybridization are demonstrated in Fig 1.The probes for STRs 5 and 6 were from the regions with no match.

Fig. 4. Hybridization of subtelomeric regions (STRs) 1, 2, 3 and 7 to chromosome 5 DNA. Chromosome 5 DNA was digested with *Not*I or *Pme*I, separated by Pulse Field Gel Electrophoresis, and blots were hybridized with the telomeric repeat (tel) and the four STRs that hybridized to chromosome 5 in Fig 3. The broad telomeric band at 6 kb on the *Not*I digest is characteristic of smaller telomeric restriction fragments and results from the variability of telomere length. The large nontelomeric fragments recognized by STRs 1 and 2 (*vsp*s) in all three digests most likely represents cross-hybridization to other similar *vsp*s. The single 5 kb *Not*I fragment recognized by STR7 (rDNA) is consistent with the presence of two *Not*I sites in rDNA, since the smaller fragment would have migrated off the gel. For each set of hybridizations from a single gel, the telomere hybridization is shown on the left. Numbers indicate the sizes of the telomere fragments in kb. Numbers indicate the sizes of the telomere fragments in kb.

Fig. 5. Chromosome 5, 4 and 3 maps showing the relevant telomeric restriction sites. For chromosome 5, we have shown subtelomeric region 7 (STR7) (rDNA) on only one end, because the data fit that interpretation better than placing it on both sides. For chromosome 4, the 290 kb *Not*I fragment is also shown since our data suggest that is on the opposite end of the chromosome from STR7. The smaller *Not*I fragments cannot be reliably placed.

Fig. 6. The telomere and subtelomeric region (STR) probes were hybridized to *Not*I, *Pme*I and *Swa*I digests of chromosomes 4 (left) and 3 (right). The lack of STR7 hybridization to *Not*I telomeric

fragments of chromosome 4 is because rDNA contains two *Not*I fragments. The STR7 band at the top of the Chromosome 4 *Swa*I digest may represent chromosome-internal copies of rDNA. The lack of telomeric hybridization to that band suggests that it is not due to partial digestion. The STR4 *Not*I fragment of chromosome 3 is similar in size to the telomeric fragment, but does not completely co-localize. This result can be explained by the presence of a *Not*I site in GilT. For each set of hybridizations from a single gel, the telomere hybridization is shown on the left. Numbers indicate the sizes of the telomere fragments in kb.

Fig. 7. Hybridization of telomere and subtelomeric region (STR) probes to *Not*I, *Pme*I and *Swa*I digests of chromosome 2/1b. Chromosomes 2 and 1b are combined since they co-migrate by Pulse Field Gel Electrophoresis. Each STR that hybridized to chromosomes 2, 1b, and/or 1a was included. For each set of hybridizations from a single gel, the telomere hybridization is shown on the left. Numbers indicate the sizes of the telomere fragments in kb.

Table 1. Subtelomeric regions (STRs)

STR #	STR type	Clone	Length (bp)[b]	Tel rep length (bp)[c]	Probe	GenBank #	Chromosome location
1	*vsp*	AJ1354[a]	2039	>406	*Pvu*II/*Pvu*II (1279-1482)	**DQ098920**	5
		MJ3348	2206	247		**DQ098921**	
2	*vsp*	EJ6107[a]	1869	>87	*Pst*I/*Eco*RV (834-1314)	**DQ098922**	5
3	GilM	KI1170[a]	2129	>213	*Kpn*I/*Hind*III (753-1836)	**DQ098923**	1b/2, 5
		EJ1336[a]	2338	30		**DQ098924**	
		EJ2414	1864	>220		**DQ098925**	
		LJ0347	2163	271		**DQ098926**	
4	GilT	KJ1356[a]	1977	>139	*Hind*III/*Hind*III (1-1846)	**DQ098927**	3
		NJ1197[a]	1587	125		**DQ098930**	
		NJ3761[a]	1844	>227		**DQ100077**	
		KJ2306[a]	1905	>171		**DQ100078**	
5	*vsp*, GilD, GilM, and no match	NF0311[a]	2500	96	*Pst*I/*Pst*I (733-1776)[d]	**DQ098928**	1b/2
		NG0253	2300	275		**DQ098929**	
6	No match and rDNA	EJ7728[a]	4789	116	PCR (16-705)[d]	**DQ098931**	1b/2
7	rDNA	NA				Ref (Adam et al., 1991)	1a, 1b/2, 4
8	Unknown	NA				p4E Ref (Hou et al., 1995)	1a, 1b/2

[a] Bidirectional sequence was obtained
[b] To telomere transition
[c] Length of the telomeric repeat region. The > sign indicates that the telomeric repeats extended as far as the initial sequence. For these clones, the transition region was determined by primer walking from the opposite end.
[d] The probes for STRs 5 and 6 were obtained from the regions with no Genbank match.

Table 2. Chromosome characteristics

Chromosome	Size (Mb)	STR[a]		SizeVariability of homologues (kb)
		Left	Right	
5	3.8	GilM ?rDNA	rDNA	~ 100
4	3.0	Unknown	rDNA	~100
3	2.3	Unknown	GilT	1.5
2	1.6	GilM STR5 STR6		< 100
1	1.6	P4E	rDNA	500[b]

[a] The left/right distinction is arbitrary and is not intended to correlate with other published data. STR, subtelomeric region

[b] The degree of variability of the chromosome 1 homologues for the E11 clone of the ISR isolate was 800 kb (Adam et al., 1988).

Fig 1

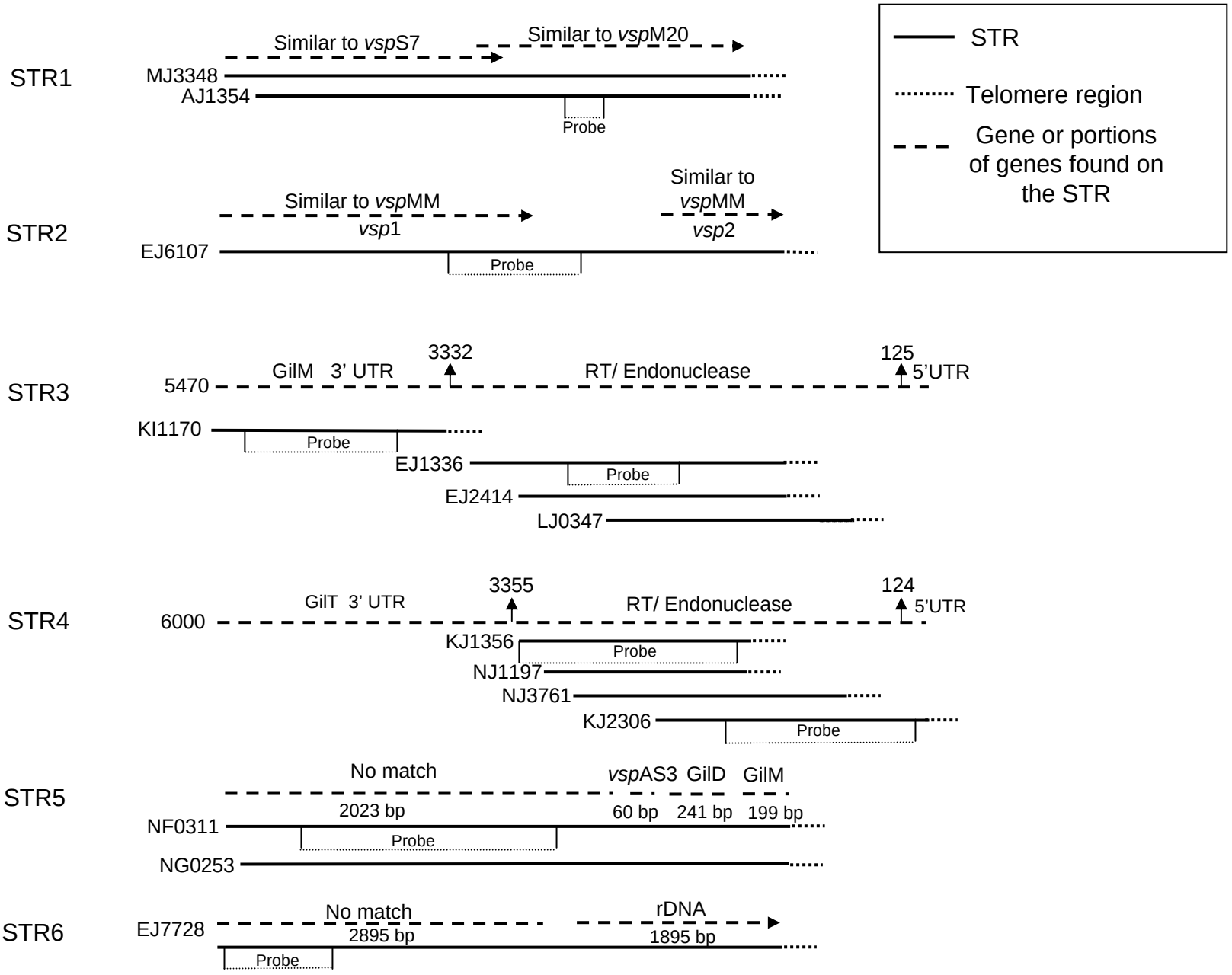

STR
Telomere region
Gene or portions of genes found on the STR
STR1
Similar to vspS7
Similar to vspM20
MJ3348
AJ1354
Probe
STR2
Similar to vspMM
vsp1
Similar to vspMM
vsp2
EJ6107
Probe
STR3
3332
125
GilM 3' UTR
RT/ Endonuclease
5'UTR
5470
KI1170
Probe
EJ1336
Probe
EJ2414
LJ0347
STR4
3355
124
GilT 3' UTR
RT/ Endonuclease
5'UTR
6000
KJ1356
Probe
NJ1197
NJ3761
KJ2306
Probe
STR5
No match
vspAS3 GilD GilM
2023 bp
60 bp 241 bp 199 bp
NF0311
Probe
NG0253
STR6
No match
rDNA
EJ7728
2895 bp
1895 bp
Probe

Fig 2

RT-PCR of Retroposon Genes

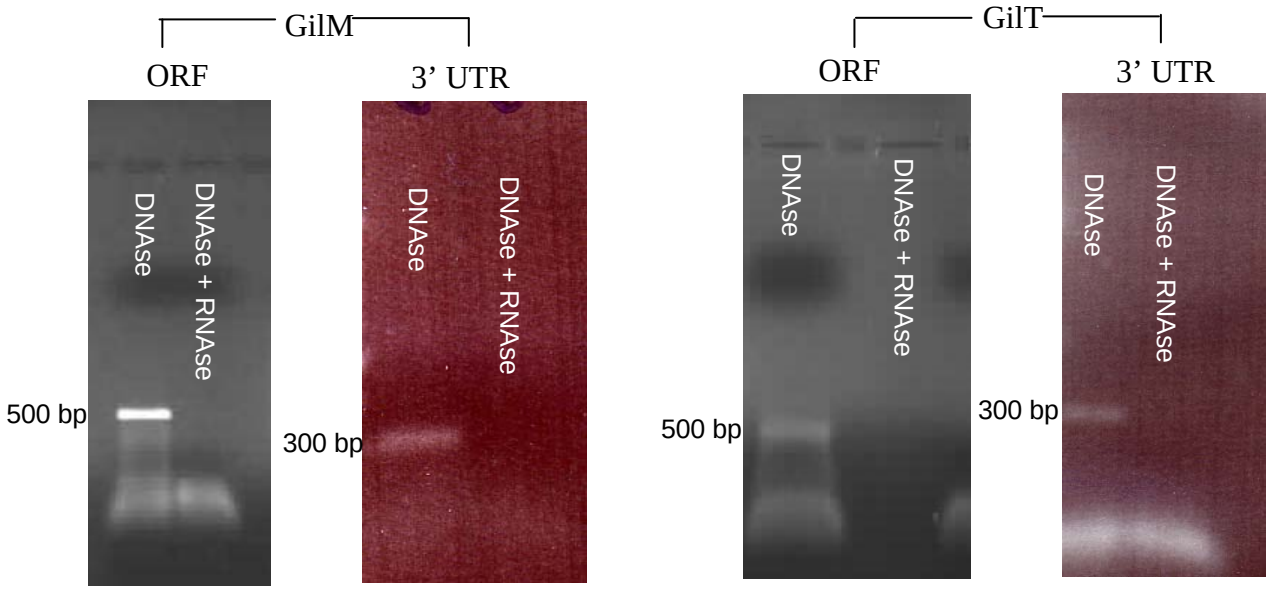


Fig 3

Chromosomal Locations of STRs

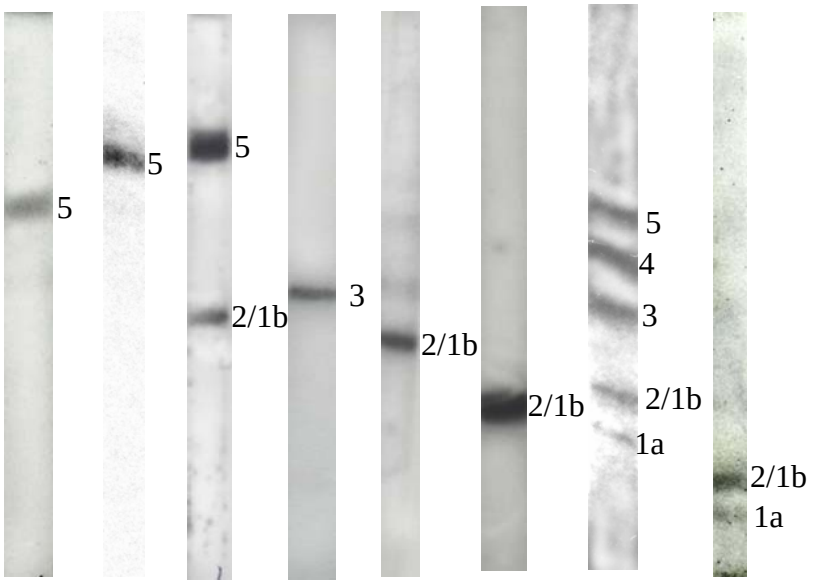


Fig 4

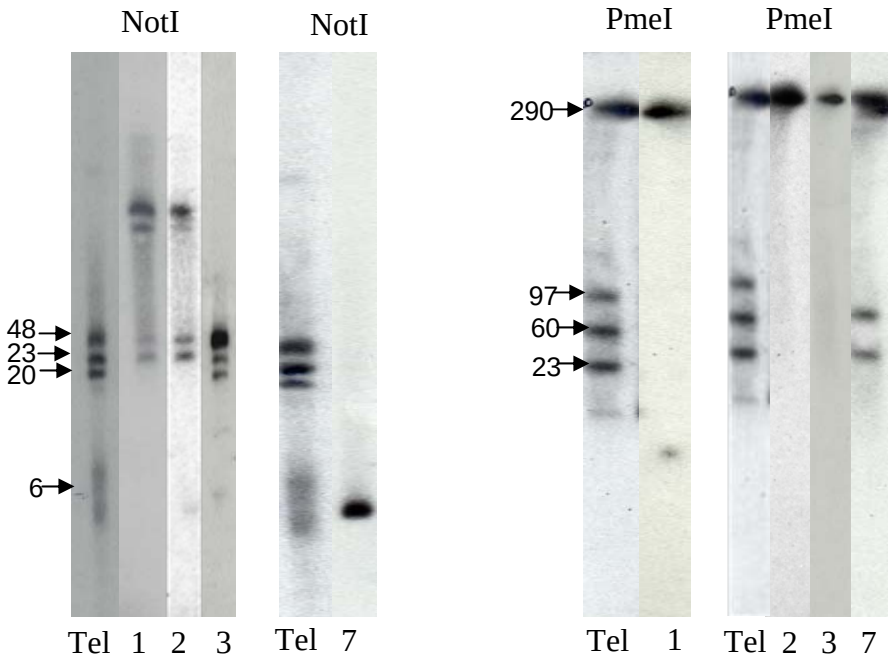

Chromosome 5
NotI
NotI
PmeI
PmeI
48
23
20
6
290
97
60
23
Tel 1 2 3
Tel 7
Tel 1
Tel 2 3 7

Fig 5

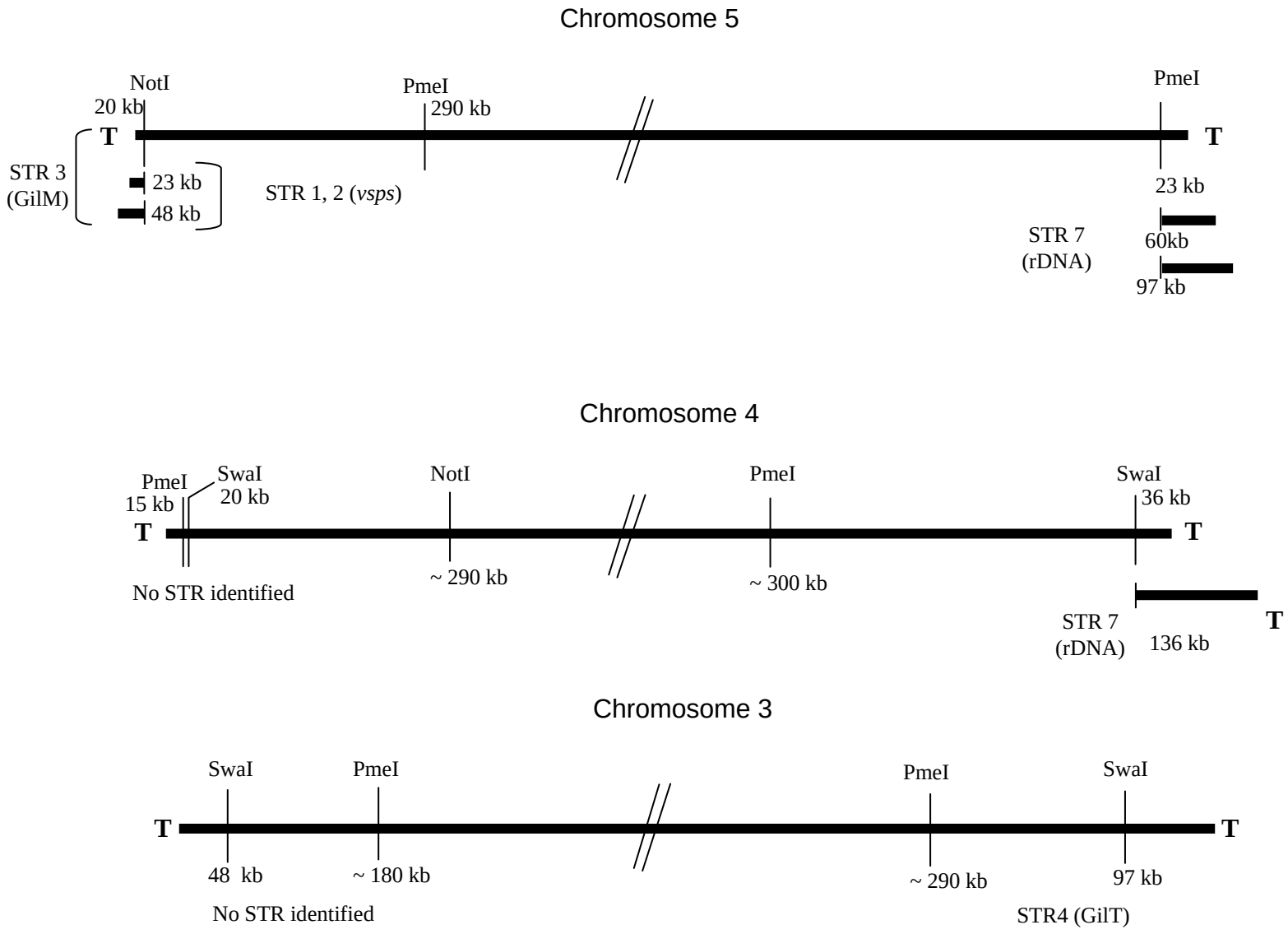
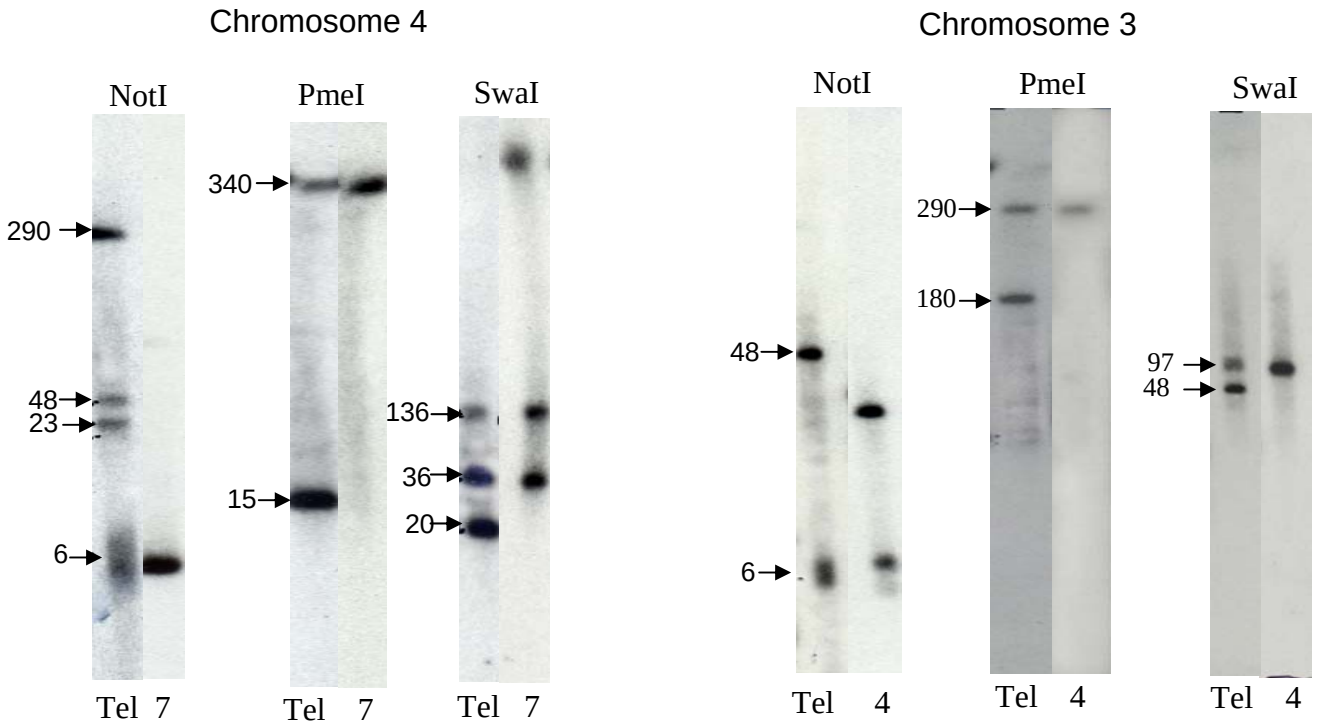

Chromosome 5
NotI
20 kb
PmeI
290 kb
PmeI
T
T
STR 3
(GilM)
23 kb
48 kb
STR 1, 2 (vsps)
23 kb
STR 7
(rDNA)
60kb
97 kb
Chromosome 4
PmeI
15 kb
SwaI
20 kb
NotI
PmeI
SwaI
36 kb
T
T
No STR identified
~ 290 kb
~ 300 kb
STR 7
(rDNA)
136 kb
T
Chromosome 3
SwaI
PmeI
PmeI
SwaI
T
T
48 kb
~ 180 kb
~ 290 kb
97 kb
No STR identified
STR4 (GilT)

Fig 6


Chromosome 4
NotI
PmeI
SwaI
290
48
23
6
340
15
136
36
20
Tel 7
Tel 7
Tel 7
Chromosome 3
NotI
PmeI
SwaI
48
6
290
180
97
48
Tel 4
Tel 4
Tel 4

Chromosome 1b and 2

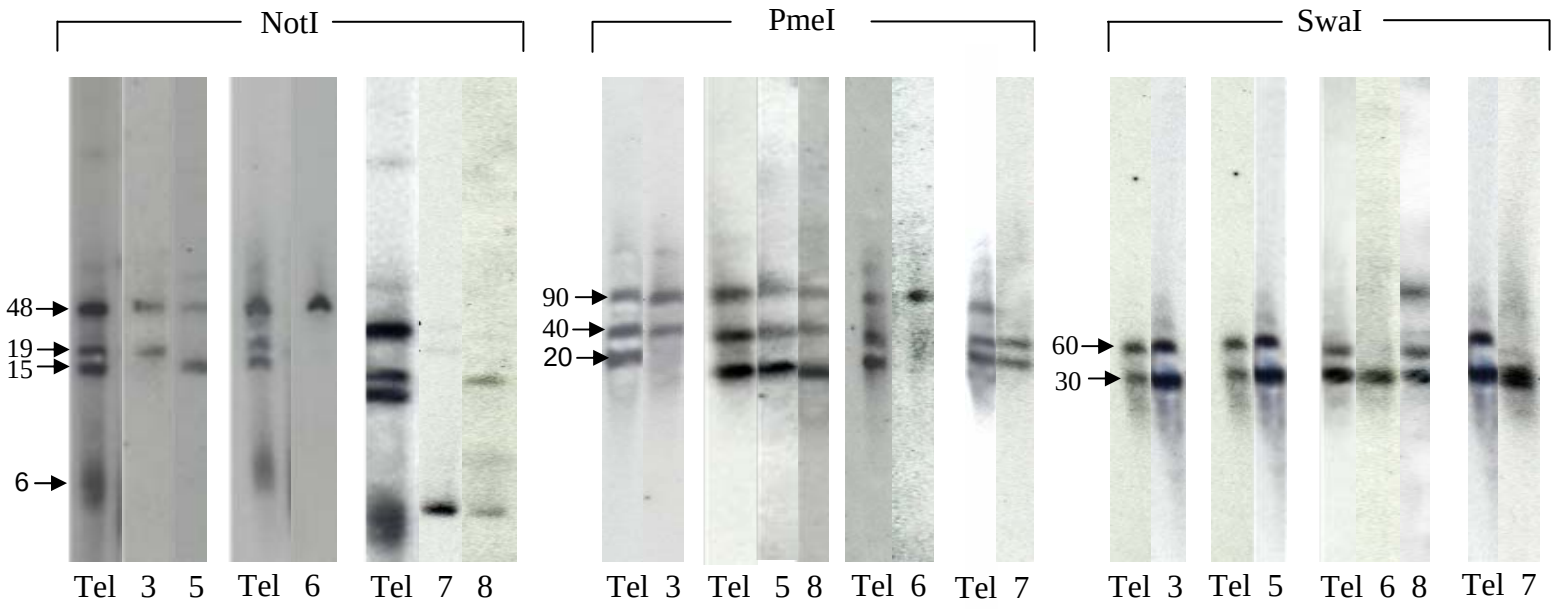



Fig 7