Giardia lamblia: haploid genome size determined by pulsed field gel electrophoresis is less than 12 Mb

Jian-Bing Fan[1], Stanley H.Korman[2], Charles R.Cantor[1,3] and Cassandra L.Smith[1,3]*
[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA,
[2]Department of Pediatrics, Hadassah University Hospital, Mount Scopus, Jerusalem, Israel and
[3]Human Genome Center, Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

Received November 7, 1990; Revised and Accepted February 21, 1991

ABSTRACT

Previous estimates of the size of the *Giardia lamblia* genome have ranged from 30 to 80 million base pairs (Mb), based on DNA renaturation kinetics. This is much larger than the sum of the sizes of the 4 to 5 chromosomal DNAs seen in typical pulsed field gel electrophoretic analyses. One possible explanation is that each visible chromosomal DNA consists of several unresolved DNA species. To examine this we have performed quantitative densitometry of ethidium stained chromosomal DNAs and *NotI* genomic digests. We have also examined the distribution of rDNA on *NotI* genomic fragments. All of our results suggests that the true genome size is 10.6 to 11.9 Mb. It is conceivable that the previous larger estimates may be distorted by impurities in the DNA preparations used.

INTRODUCTION

Giardia lamblia is the major protozoan agent responsible for diarrhoea throughout the world (1). Infection appears to occur via ingestion of a cyst, usually from contaminated water. Acidic gastric conditions induce excystation (2). The emerging tetranucleated trophozoite either divides and replicates in the small intestine, yielding trophozoites with two nuclei, or it encysts. The trophozoites adhere to the intestinal wall. The cysts are excreted.

We are interested in characterizing the extent of genomic variation in *Giardia* strains isolated from different parts of the world. One way to determine this is to examine the size and number of chromosomes. This was not possible to do until the size limits of pulsed field gel electrophoresis (PFG) were expanded to 7 Mb (3, 4). Chromosomal DNAs were purified from several *Giardia lamblia* isolates and resolved by PFG (5, 6). Usually, four major DNA bands were detected, while sometimes a few minor bands were also seen.

MATERIALS AND METHODS

Manipulation and analysis of genomic DNA

The *Giardia lamblia* Portland DNA was a gift from Dr. M.Singh, National University of Singapore. The *Giardia lamblia* KC DNA samples have been previously described (7). Genomic DNA from *Giardia* cells grown in culture was prepared in agarose and digested with the restriction enzyme *NotI* using enzyme:DNA ratios of 20 Units: 1 μg as previously described (5). DNA samples were analysed by PFG using inhomogeneous fields in a Pulsaphor apparatus (Pharmacia-LKB) as described (8). Essentially a 1% agarose (Seakem LE Agarose, FMC) 20 cm square gel was run at 15°C in modified TBE buffer (100 mM Tris−HCl, pH 8.0−8.4, 100 mM boric acid and 0.2 mM EDTA). The DNA bands were visualized by staining the gels in distilled water containing 1 μg/ml ethidium bromide and destaining in TBE for 60 min. Photographs of ethidium bromide stained agarose gels were scanned on a Pharmacia-LKB Ultrascan laser densitometer (model 2202). Peak intensity was calculated using a Pharmacia-LKB integrator (model 2221). The ethidium bromide intensities serve as measurements of the relative amounts of DNA mass in the scanned DNA bands.

RESULTS

Sizes of the four major *Giardia* DNA bands

The sizes of the four major *Giardia* DNA bands are 1.4, 1.7, 2.7, and 3.4 Mb (Figure 1), based on comparison with the available DNA standards for this size domain, i.e. the *NotI* digested *S. pombe* chromosomal DNAs (9). These size estimates are smaller than those of 1.6, 2.3, 3.0, and 3.8 Mb reported previously (6). The earlier estimations were made at the time when no reliable size standards were available in this size range.

In 1988 Upcroft *et al.*, using the field inversion gel electrophoresis (FIGE) technique, reported that the chromosomes of *Giardia intestinalis* were larger in size than those of *Trypanosoma brucei* (10). Since the *T. brucei* chromosomes range

* To whom correspondence should be addressed

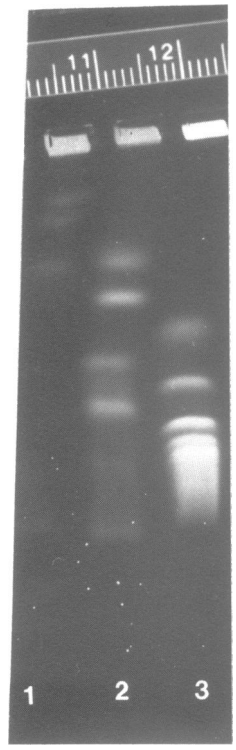

Figure 1. PFG separation of *Giardia lamblia* DNA. Chromosomal DNAs from the Portland strain of *Giardia* are 3.4, 2.7, 1.7, 1.4 Mb in size (lane 2). Chromosomal DNAs from *S. cerevisiae* (18, lane 3), *S. pombe* (9, lane 1) and other DNAs (data not shown) were used as size standards. PFG fractionation was carried out using a pulse-program: 4500 sec pulse times for 45 hrs; 2700 sec pulse times for 50 hrs; 1800 sec pulse times for 65 hrs and 900 sec pulse times for 30 hrs at field strength 3 V/cm.

in size up to at least 6 Mb (5), *G. intestinalis* chromosomes would be much larger than our current size estimates. However, since FIGE mobility of DNA is not a monotonic function of size (11), the *G. intestinalis* estimates are not likely to be correct.

The 1.4 Mb DNA band contains two chromosomal DNAs

If each major PFG band represented a single chromosome, these experiments would indicate a minimum *G. lamblia* genome size of about 9.2 Mb. This size is only twice that of *E. coli.* (12), and it would make *Giardia* a very small eukaryotic genome. However, inspection of Figure 1 clearly reveals that the ethidium bromide staining intensity of the smallest band is too great for its relative size. Densitometric scanning results showed that the band is actually an unresolved doublet containing two chromosomal DNAs (Table 1). In fact, in some isolates studied by Adam *et al.* (6) these two DNAs differed in size sufficiently to be resolved by PFG (6). Since two chromosomes are contained in the 1.4 Mb band, the minimum size of the *Giardia* genome is 10.6 Mb. This estimate of the *Giardia* genome size is 3 to 8 times smaller than estimates of the genomic complexity previously obtained by renaturation kinetics. Those estimates are 30−32 Mb (13) and 80 Mb (14). Adam *et al.* (6) reconciled the PFG results with the renaturation results by suggesting each band seen by PFG actually consists of many similarly sized different chromosomes (6).

Note that the integrated staining intensity of the largest *Giardia* chromosomal DNA band is less than expected from its apparent size (Table 1). One possibility is that some of this material has been lost due to DNA damage or degradation. However a likely alternative explanation is that some of this material never entered the agarose running gel under the PFG conditions used. DNA trapping is known to be a sensitive function of DNA size (15, 16). Note that the largest and smallest *Giardia* chromosomal DNA bands are also broader than the others (Figure 1). A similar situation occurs with the largest *S. cerevisiae* chromosomal DNA band (Figure 1), known to contain tandem repeats of rDNA (17, 18). The width of these bands, representing rDNA containing chromosomes presumably reflects heterogeneity in the number of rDNA repeats. To examine these issues, a blot of separated *Giardia* chromosomal DNAs was hybridized with a cloned *Giardia* rDNA probe. This result showed that rDNA was present in the largest and smallest *Giardia* chromosomal DNA bands, and that a considerable amount of rDNA-containing material remained in the sample gel insert (data not shown).

Table 1. Densitometric quantitation of *Giardia lamblia* Portland chromosomal DNAs

Band[1]	Size (Mb)	Relative intensity[2]	Number of chromosomes[3]	Number of chromosomes[4]
1	3.4	24.0	0.75	1
2	2.7	29.6	1.16	1
3	1.7	17.0	1.06	1
4	1.4	29.4	2.23	2

[1]In lane 2 of Figure 1.
[2]Percent of ethidium fluorescence in the four major bands in lane 2 of Figure 1. The minor bands are not included because they are clearly present in non-stoichiometric amounts.
[3]Calculated as (relative intensity)×(genome size)/(band size)
[4]Assumed.

Table 2. Densitometric quantitation of *Giardia lamblia* Portland and KC4 genomic DNA digested with the restriction enzyme *Not*I

(A). Portland strain:

Band	Size (kb)[1]	Relative Intensity[2]	Number of Fragments[4]	Genome Size (Mb)[5]
1	475	3.8	1	12.6
2	393	7.0	2	11.2
3	325	2.8	1	11.8
4	257	13.2	6	11.7
5	230	6.4	3	10.8
				Average 11.6

(B). Strain KC4:

Band	Size (kb)[1]	Relative Intensity[3]	Number of Fragments[4]	Genome Size (Mb)[5]
1	563	4.6	1	12.3
2	393	6.3	2	12.4
3	376	3.2	1	11.6
4	330	8.4	3	11.8
5	266	9.0	4	11.8
6	243	4.4	2	11.1
7	194	3.2	2	12.2
				Average 11.9

[1]In Figure 2 by comparison with annealed λ_{cI857} concatemers.
[2]Percent of total ethidium fluorescence in lane 4 of Figure 2.
[3]Percent of total ethidium fluorescence in lane 6 of Figure 2.
[4]Assumed.
[5]Calculated as (band size)×(number of fragments)/(relative intensity).

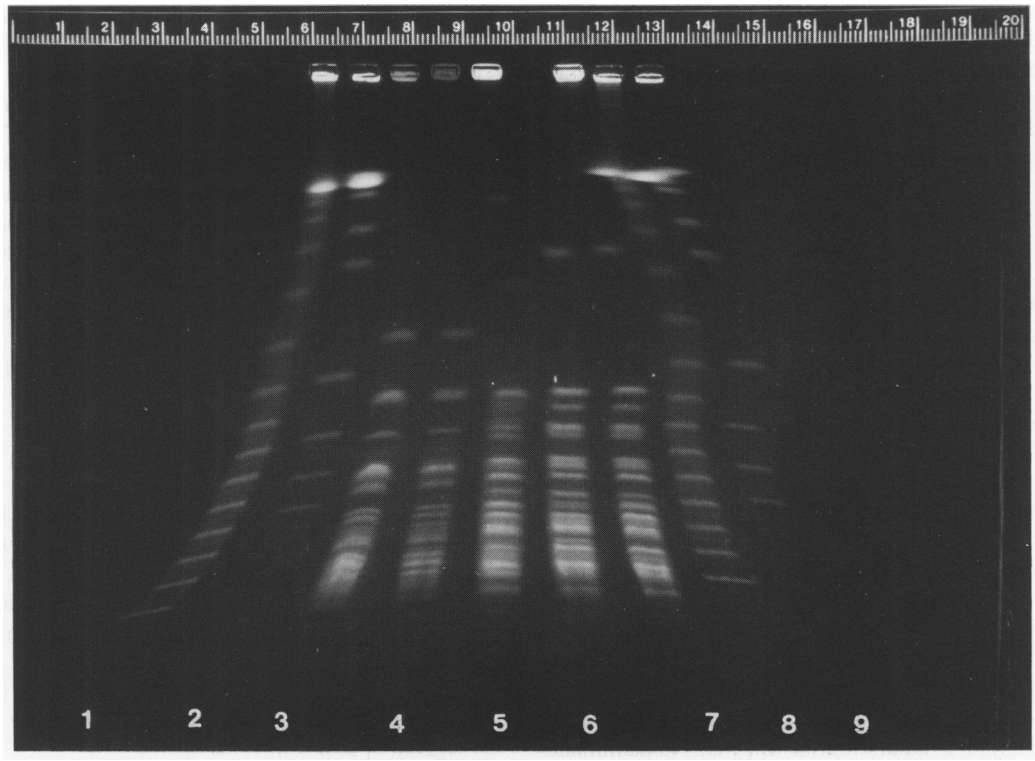

Figure 2. PFG separation of *Giardia lamblia* DNAs digested with the restriction enzyme *Not*I. *Giardia* DNA was purified from the Portland strain (lanes 3 and 4) or the KC 2, 4 and 6 strains (lanes 5–7, respectively). The KC strains are described in (7). Length standards were (lanes 1 and 8) multimers of bacteriophage lambda DNA (48.5 kb monomer) and (lanes 2 and 8) intact *S. cerevisiae* chromosomal DNAs (18). PFG fractionation was carried out at 10 V/cm for 48 hrs using 60 sec pulse times as described previously (8).

The haploid *Giardia lamblia* genome size determined by pulsed field gel electrophoresis and densitometry is less than 12 Mb

To characterize the *Giardia* genome further and to obtain additional evidence about the genome size, chromosomal DNA from several isolates was digested with several restriction enzymes. Results with *Not*I are shown in Figure 2. If each of the five major intact chromosomal DNA bands described above represented many different chromosomes, a large number of different-sized *Not*I fragments would be expected. However the actual results show a fairly simple banding pattern with not much variation among the isolates. This result suggested that the *G. lamblia* genome size is much smaller than 80 Mb and that each major chromosomal DNA band seen by PFG may represent only one or two different chromosomes. Adam *et al.* (1988) suggested that the *Giardia lamblia* trophozoite contains sets of closely related chromosomes (6). If this is true, then the chromosomes within each set should be so similiar to each other that no difference could be detected by *Not*I digestion and PFG.

Ethidium bromide staining intensity is a fairly accurate measure of DNA weight. If the *Not*I fragments in Figure 2 were derived from a small genome, one would expect them to be pure species or unresolved simple multiplets. The relative staining intensity should therefore occur in predictable ratios, dependent on DNA size and number of unresolved species. Densitometric scanning was used to quantitate the amount of DNA in several large *Not*I fragments seen in the *G. lamblia* Portland strain (Table 2-A). By assigning small integral numbers of fragments in each DNA band, the observed ratios of weight to intensity for the DNA bands produces remarkably consistent estimates of the size of the genome. When the scanned raw data was used, an 11.2 Mb genome size was estimated, with a 15% variation. However, when the background was subtracted from the ethidium bromide-stained DNA bands, the genome size was estimated at 11.6 Mb, with a variation in these estimates at less than 5% (Table 2-A). We will use these latter values because of the smaller variation. It is particularly reassuring that two bands are predicted to be singlets. Note that we did not use the data from the smaller DNA bands to calculate the genomic size because of the poor resolution of these bands. It would also be very risky to assume a number of fragments in a band when that band contains more than 6 fragments.

Using the same method an 11.9 Mb genome size was determined for a newly isolated *Giardia* strain, KC4 (7, Table 2-B). As shown in Figure 2, strains KC 2, 4 and 6 all have similar *Not*I fragment patterns. This determination was reassuring since the Portland strain has an insecure history and has been maintained in culture for some time (13, 19, 20). A parallel densitometric experiment of *Schizosaccharomyces pombe* genomic DNA digested with *Not*I yielded a genome size 14 Mb (data not shown). This is the same size as that calculated from restriction map of this organism (9). Thus, we are reasonably confident that densitometry can be used to estimate the size of an unknown genome. Using densitometry, we also made an estimate of the size of chromosome III of *S. pombe*, which was in good agreement with the size determined by other direct methods (9).

In principle, densitometry might underestimate the size of a genome if it contained many small *Not*I fragments that would run off the gel and not be detected. However, under the

electrophoretic conditions that we used, DNA fragments as small as 2 kb still remain in the gel (data not shown). The *Giardia lamblia* genome has an overall G:C content of 46.8% (14). If we assume a random distribution of the nucleotides in the *Giardia* genome, the average size of *Not*I fragments should be 110 kb. Thus, *Not*I fragments with sizes lower than 2 kb should be very rare. An exception might be rDNA since it is G:C rich. Healey *et al.*, reported a 5,566 base-pair sequence for the *G. intestinalis* rDNA (21). Two *Not*I sites are found in this sequence. In contrast, no *Not*I site is present in the *G. lamblia* rDNA sequence (14). In fact, the *G. lamblia* rDNA is found on only four *Not*I fragments, with sizes of 40, 73, 121, and 146 kb (data not shown). The sum of these fragment sizes is 380 kb. This is not much more than the minimum length of DNA needed to contain the estimated 63 copies of the 5.4 kb rDNA repeat, 340 kb (14). While we cannot exclude the possibility that one or more of the rDNA-containing bands contains multiple fragments, the simplicity of the pattern is further evidence of a small genome size.

One possible source for the size discrepancy between our work and previous size estimates for *Giardia* genome would be the presence of significant amounts of DNA which remained in the PFG well. These could be circular molecules, and if they lacked cleavage sites for *Not*I, they might remain in the well even after *Not*I digestion. To examine this possibility, agarose insert samples were subjected to gamma irradiation prior to PFG. No additional species which might arise from once-broken circular DNAs were seen in these experiments, while parallel experiments produced linear *E. coli* genomic DNA (4.5 Mb) which migrated into the gel as a new distinct DNA band (data not shown). Thus we are still left with an unexplained discrepancy between our genome size estimate and previous estimates of the genome complexity.

DISCUSSION

Inspection of the results in Figure 4 of Boothroyd *et al.* (14) provides a possible explanation for the large discrepancy. It is apparent that their sample of *Giardia* DNA displayed substantial DNA premelting and a hypochromicity of only about 0.08. This is a third of the hypochromicity they report for cloned DNA samples or for a sample of *E. coli* B DNA. If one discounts the premelting contribution to the hypochromicity, then the change in absorbance seen upon melting in *Giardia* is only a sixth of that of the other samples. The simplest explanation for this would be that the DNA sample is contaminated by a great excess of a UV absorbing material, perhaps RNA. This contaminant would affect concentrations determined by absorbance measurements, and as a result, the concentrations used for analysis of renaturation kinetics could have been overestimated by a factor of six. If this were true, the resulting estimate of the genome complexity would be reduced from 80 Mb to 13 Mb, which is in reasonable agreement with the results reported here. The same contaminant may conceivably have plagued the earlier studies of Nash *et al.* (13).

Comparative densitometry was employed by Adam *et al.* (6) to estimate the DNA content of *Giardia lamblia*, relative to that of *Plasmodium falciparum*. The result was 80 to 120 Mb per cell or 40–60 Mb per nucleus. These values may be very uncertain because these authors assume 100% recovery of DNA from the two organisms. However, if one tentatively accepts these DNA contents then this determination combined with our genome size estimates suggests that each *Giardia* nucleus may contain 3–5 copies of identical or nearly identical chromosomes.

In this report, we show that the haploid *G. lamblia* genome may contain only five different major chromosomes. *G. lamblia* isolates may have a genome complexity even smaller than the genome size of yeasts like *S. cerevisiae* (18) and *S. pombe* (9) which are both estimated at about 14 Mb. Thus, this organism may have one of the smallest possible eukaryotic genomes which would suggest a very primitive biology consistent with the very early evolutionary separation of *Giardia* from the rest of the eukaryotes (22). Boothroyd *et al.* reported that *G. lamblia* had the smallest rDNA repeat unit for any eukaryote and suggested that there may be a selective pressure on the overall genome size of this organism (14).

ACKNOWLEDGEMENTS

This work was supported by grants from DOE, DE-FG02-87ER-GD852 and the MacArthur Foundation. The authors would like to thank M.Singh, for providing Portland DNA, C.C.Wang, for *G. lamblia* rDNA, D.Spira, for encouragement, and providing the KC DNA samples, and R.Deckelbaum and G.Jasper for expediting this work.

REFERENCES

1. Wolfe, M.S. (1984) In Erlander, S.L. and Meyer, E.A. (eds.), *Giardia and Giardiasis: Biology, Pathogenesis, and Epidemiology*. Plenum Press, New York, pp. 147–161.
2. Bingham,A.K. and Meyer,E.A. (1979) *Nature* **277**, 301–302.
3. Vollrath,D. and Davis,R.W. (1987) *Nucl. Acids Res.* **15**, 7865–7876.
4. Smith,C.L., Tomohiro,M., Niwa,O., Klco,S., Fan,J-B., Yanagida,M. and Cantor,C.R. (1987) *Nucl. Acids Res.* **15**, 4481–4489.
5. Smith,C.L. (1990) In Van der Ploeg, L.H.T., Cantor, C.R. and Vogel, H.J. (eds.), P&S Biomedical Sciences Symposia: *Immune Recognition and Evasion: Molecular Aspects of Host-Parasite Interaction*. Academic Press, New York, pp. 237–243.
6. Adam,R.D., Nash,T.E. and Wellems,T.E. (1988) *Nucl. Acids Res.* **16**, 4555–4567.
7. Korman,S.H., Le Blancq,S.M., Spira,D.T., El On,J., Reifen,R.M. and Deckelbaum,C. (1986) *Z. Parasitenkd.* **72**, 173–180.
8. Smith,C.L., Klco,S.R. and Cantor,C.R. (1988) In Davies, K. (ed.), *Genome Analysis: A Practical Approach*. IRL Press, Oxford, pp. 41–72.
9. Fan,J-B., Chikashige,Y., Smith,C.L., Niwa,O., Yanagida,M. and Cantor,C.R. (1989) *Nucl. Acids Res.* **17**, 2801–2817.
10. Upcroft,P., Boreham,P.F.L. and Upcroft,J.A. (1988) In Wallis, P.M. and Hammond, B.R. (eds.), *Advances in Giardia Research*. University of Calgary Press, Calgary, pp. 147–152.
11. Ellis,T.H.N., Cleary,W.G., Burchan,K.W.G. and Bowen,B.A. (1987) *Nucl. Acids Res.* **15**, 5489.
12. Smith,C.L., Econome,J.E., Schutt,A., Klco,S. and Cantor, C.R. (1987) *Science* **236**, 1448–1453.
13. Nash,T.E., McCutchan,T., Keister,D., Dame,J.B., Conrad,J.D. and Gillin,F.D. (1985) *J. Infect. Dis.* **152**, 64–73.
14. Boothroyd,J.C., Wang,A., Campbell,D.A. and Wang,C.C. (1987) *Nucl. Acids Res.* **15**, 4065–4084.
15. Ulanovsky,L., Drouin,G. and Gilbert,W. (1990) *Nature* **343**, 190–192.
16. Turmel,C., Brassard,E., Slater,G.W. and Noolandi,J. (1990) *Nucl. Acids Res.* **18**, 569–575.
17. Schwartz,D.C. and Cantor,C.R. (1984) *Cell* **37**, 67–75.
18. Mathew,M.K., Smith,C.L. and Cantor,C.R. (1988) *Biochemistry* **27**, 9210–9216.
19. Bertram,M.A., Meyer,E.A., Lile,J.D. and Morse,S.A. (1983) *J. Parasitol.* **69**, 793–801.
20. Miller,R.L., Wang,A.L. and Wang,C.C. (1989) *Exp. Parasitol.* **66**, 118–123.
21. Healey,A., Mitchell,R., Upcroft,J.A., Boreham,P.F.L. and Upcroft,P. (1990) *Nucl. Acids Res.* **18**, 4006.
22. Sogin,M.L., Gunderson,J.H., Elwood,H.J., Alonso,R.A. and Peattie,D.A. (1989) *Science* **243**, 75–77.