ELSEVIER

Molecular and Biochemical Parasitology 95 (1998) 147–151

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

Short communication

# Construction and characterisation of a genomic PAC library of the intestinal parasite *Cryptosporidium parvum* [1]

Mike B. Piper, Alan T. Bankier, Paul H. Dear *

*Medical Research Council Laboratory of Molecular Biology, Protein and Nucleic Acid Chemistry Division, Hills Road, Cambridge CB2 2QH, UK*

Received 4 June 1998; received in revised form 22 June 1998; accepted 29 June 1998

*Keywords:* *Cryptosporidium parvum*; PAC; Genomic library; STS; Karyotype; Mapping

*Cryptosporidium parvum* is an obligate-intracellular parasite of the gut mucosa. It infects many mammals, and can be considered an emerging pathogen in man, the first case of human cryptosporidiosis being reported as recently as 1976 [1].

Little is known about the molecular genetics of *Cryptosporidium*, partly due to our inability to continuously culture it in vitro. Oocysts for study are usually obtained by extraction from the faeces of infected calves or lambs [1]. Karyotype analysis suggests that the *Cryptosporidium* genome consists of eight chromosomes of $\approx 1-2$ Mb [2,3], giving a genome size of $\sim 10.4$ Mb [2]. Several partial and complete gene sequences have been reported [4–15], and an expressed sequence tag (EST) sequencing project is currently in progress (http://www.embl-ebi.ac.uk/parasites/news.html). Several small insert *Cryptosporidium* libraries, both genomic [6,9] and cDNA [4,5,12–15], have been reported. However, no large-insert libraries have hitherto been available, impeding mapping and sequencing efforts.

The vector chosen for constructing this library was the P1 artificial chromosome (PAC) vector pCYPAC2 [16], which differs by only a single *Not*I restriction site from pCYPAC1 [17]. PAC vectors are based on the P1 vector system [18,19], but are introduced into the host cell by electroporation rather than viral transformation. A PAC library can be manipulated exactly as one would a P1 library and has the same desirable features, e.g. clonal stability over many generations and

*Abbreviations:* EST, expressed sequence tag; PAC, P1 artificial chromosome; PFGE, pulsed field gel electrophoresis; STS, sequence tagged site.

* Corresponding author. Tel.: +44 1223 402190; fax: +44 1223 412178; e-mail: phd@mrc-lmb.cam.ac.uk

[1] *Note:* Nucleotide sequence data reported in this paper are available in the EMBL, DDJB and GenBank™ databases under the accession numbers G35127–G35275 and G35338–G35349.

0166-6851/98/$ - see front matter © 1998 Elsevier Science B.V. All rights reserved.
*PII* S0166-6851(98)00095-4

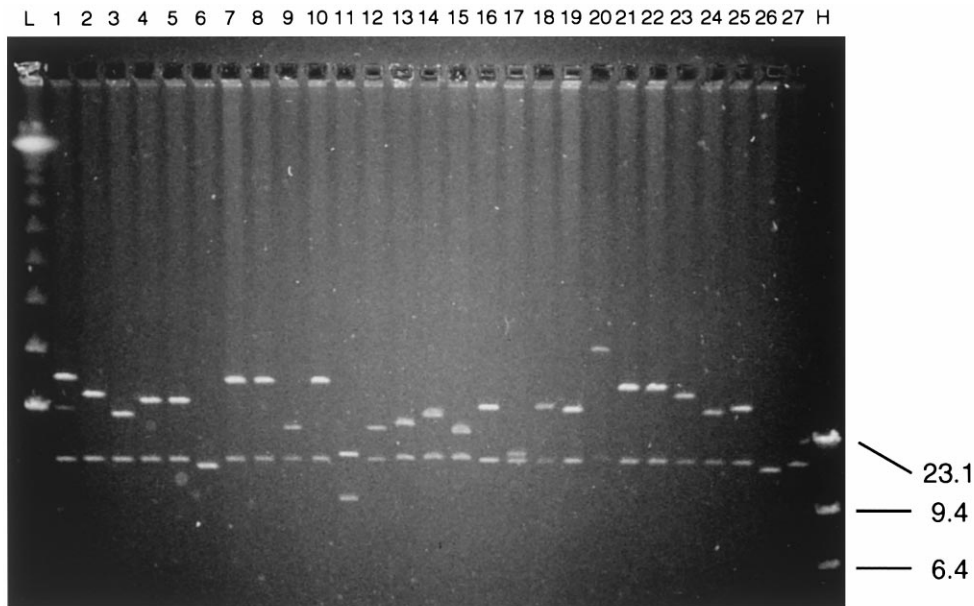


Fig. 1. Library construction and clone size-range. *Cryptosporidium parvum* oocysts (strain Moredun cervine) were obtained from the Moredun Institute (Edinburgh, UK). DNA was prepared by embedding oocysts in agarose strings at a concentration of $1.5–8 \times 10^8$ $ml^{-1}$ and lysing with proteinase K [23]. *E.coli* DH10B cells containing the PAC vector pCYPAC2 were obtained from P. de Jong (Roswell Park Cancer Institute, Buffalo, NY). The vector was prepared for ligation as described previously [16]. Approximately $4 \times 10^7$ agarose-embedded *Cryptosporidium* oocysts were partially digested with 4–8 U of *Mbo*I and size selected by pulsed field gel electrophoresis. The agarose was removed using GELase (Epicentre Technologies, Madison, WI) as described previously [16], and the size-selected DNA (50–100 kb range) ligated to pCYPAC2 (100 ng each, giving a $\sim$1:10 insert:vector molar ratio to prevent insert co-ligation) with 200 NE units of T4 DNA ligase (New England Biolabs, Beverly, MA) in 50 $\mu$l at 16°C overnight. The ligation mixture was heat-inactivated and dialysed [16], then 1 $\mu$l ligation product was mixed with 20 $\mu$l Electromax DH10B cells (Life Technologies, Grand Island, NY) and electroporated at 4°C (Bio Rad Gene Pulser, 12 kV $cm^{-1}$, 200 Ω, 25 $\mu$F, time constant 4–5 ms; 0.2 cm microelectroporation cuvette). Recombinant colonies were isolated as described previously [16], toothpicked into 96-well microtitre plates containing TY [24] plus 7.5% (v/v) glycerol and 25 $\mu$g $ml^{-1}$ kanamycin, grown for 24 h at 37°C, duplicated and stored at −70°C.PAC DNA was isolated by alkaline lysis [16] from 27 random clones, digested with *Not*I, and resolved by PFGE. L, $\lambda$ ladder DNA (multiples of 48.5 kb); H, $\lambda$ *Hin*dIII digest (sizes in kb). The $\sim$16 kb band in all sample lanes is the linear vector. It is visibly shortened in some non-recombinant clones (no inserts) due to *Bam*HI 'star' activity during vector preparation [16], e.g. lanes 6 and 26.

easy isolation of vector DNA. Recombinants are distinguished by kanamycin resistance due to the *kan*$^r$ gene, and by survival on sucrose due to disruption of the *Sac*BII gene by the insert DNA. The *Cryptosporidium* genome seems on average to be $\sim$60–70% A + T (this study; [4,9,15]). Thus, we would expect genomic *Cryptosporidium* DNA to clone stably in bacterial systems.

The complete library consists of 1728 clones with an average insert size of $\approx$38 kb (Fig. 1), although up to 20% of these clones may be non-recombinant (a typical figure for a PAC library). Based on an estimated genome size of 10.4 Mb [2], we estimate that the library represents approximately 5-fold coverage of the genome.

We characterised the library by sequence tagged site (STS)-screening, Southern blotting of individual clones to the *Cryptosporidium* chromosomes, analysis of insert stability upon serial culture and end sequencing.

Primary pools, representing each of the 18 PAC library plates, were screened by PCR for the presence of 39 STSs (primer sequences and reaction conditions available upon request), including 15 derived from our own M13 libraries (sequences deposited in Genbank, accession numbers

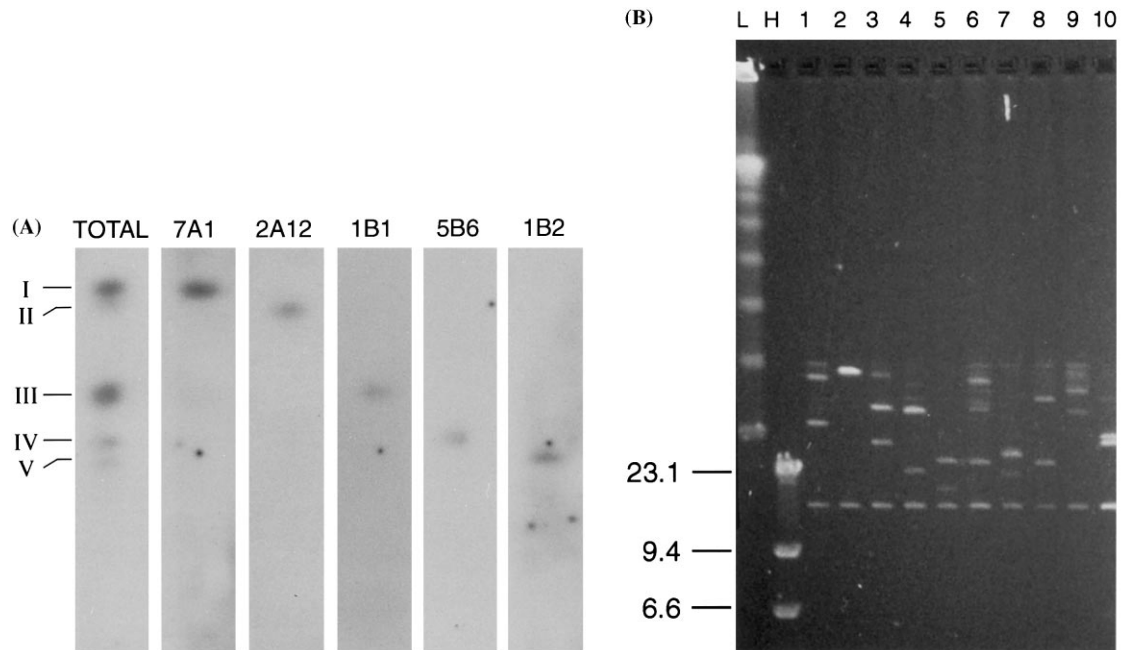


Fig. 2. Characterisation of the library. Panel A. Agarose-embedded oocysts were analysed by PFGE in a CHEF DR II system ( $\sim 10^7$ oocysts per lane, 1% chromosomal grade agarose, 4 V cm $^{-1}$, $0.5\times$ TBE (0.045 M Tris–borate, 0.001 M EDTA), 120° switching angle, 14°C, 120 s pulse for 60 h then 240 s pulse for 60 h). Southern blotting was performed first using total *Cryptosporidium* DNA as the probe. Five chromosomal bands are visible. Band I, chromosomes 1 and 2 ( $\sim$ 1.54 Mb); band II, chromosome 3 ( $\sim$ 1.44 Mb); band III, chromosomes 4, 5 and 6 ( $\sim$ 1.24 Mb); band IV, chromosome 7 ( $\sim$ 1.08 Mb); band V, chromosome 8 ( $\sim$ 1.04 Mb), band sizes are taken from [2]. Filters were stripped and re-probed with individual PAC clones using 500 ng of recombinant, $^{32}$P-labelled PAC DNA as a probe [24]. The blots probed with total DNA were used as guides to accurately assign the chromosomal location of individual PAC clones. PAC clones 7A1, 2A12, 1B1, 5B6 and 1B2 hybridise to chromosomal bands I–V, respectively. Panel B. Clone 1B8 is unstable: *Not*I-digested PAC DNA from the fourth consecutive culture of each of ten colonies of clone 1B8. L, $\lambda$ DNA ladder (multiples of 48.5 kb); H, $\lambda$ *Hin*dIII digest (sizes in kb).

G35127–G35275, G35338–G35349). A total of 8% (3/39) of STSs were not detected in the library. Assuming a Poisson distribution of unique sequences in each plate, mean STS coverage is $\sim$ 8.3 fold.

We probed a Southern blot of the *Cryptosporidium* chromosomes with purified DNA from 16 PACs. Each PAC hybridised to a discrete band, i.e. we detected no co-ligation or chimerisation of insert DNA in this library. The karyotype obtained was similar to that of a previous study [2], with five resolvable bands (Fig. 2A), all of which were positive for at least one of the above clones (Fig. 2A), suggesting that all five chromosomal bands are represented in the library.

A total of 14 clones were grown in serial cultures until saturation for five generations ( $\approx$ 60 cell duplications [17]). All but one of the clones replicated stably over this time. Clone 1B8 showed instability after $\sim$ 36 cell duplications. A further ten colonies of 1B8 were grown for four generations and analysed by pulsed field gel electrophoresis (PFGE), confirming that this clone was indeed unstable (Fig. 2B). However, it should be emphasised that all clones tested showed no instability during at least two serial cultures.

The PAC library described here provides a stable source of *Cryptosporidium* DNA for study. The apparently greater coverage (8.3-fold) calculated from STS content, as opposed to 5-fold based on genome and clone sizes, may reflect a similar bias against certain sequences in both the PAC library and in other bacterial libraries from which the STSs were derived. The pCYPAC2

vector was chosen for its ability to accept inserts with an average size of >100 kb. However, we were unable to clone DNA of this size with acceptable efficiency. We obtained clones with an average insert size of 38 kb with efficiencies up to 30000 recombinants $\mu g^{-1}$ *Cryptosporidium* DNA. This efficiency was still 10–100× lower than has been obtained when cloning human DNA with an average insert size of >100 kb [16]. It has been observed that some batches of agarose-embedded genomic DNA seem incapable of giving high cloning efficiencies [16], and our agarose-embedded *Cryptosporidium* DNA was degraded to a considerable degree (results not shown). However, as a limited number of oocysts ($\sim 5\times10^8$) were available, a low cloning efficiency was considered acceptable. Although the library was made using *Cryptosporidium* DNA with a nominal size range of 50 to 100 kb (larger size ranges did not give acceptable cloning efficiencies), an average insert size of 38 kb is not unexpected as there is always a background of smaller fragments in the size-selected DNA [16] and these tend to clone far more efficiently.

Library coverage appears to be more uneven than expected by chance. Given the estimated 8.3-fold coverage calculated from STS screening data, we would expect >99% of STS sequences to be present. However, ~8% of the STSs used were not present in the library. Some *Cryptosporidium* sequences are unstable in the PAC vector system, e.g. clone 1B8 (Fig. 2B). Thus, it appears that some regions of the *Cryptosporidium* genome may be inherently difficult to clone in the PAC vector system. Bacterial clones can be unstable due to localised regions of high A+T content or long repeats causing unusual secondary structure. It is thought that *Plasmodium falciparum* DNA is unstable in bacterial systems due to its A+T content of ~80% [20]. The overall A+T content of the *Cryptosporidium* sequence obtained from our M13 libraries was only 67.6%, a little higher than that of human. In addition we sequenced 15 PAC ends, totalling ~4.5 kb of unique sequence, in which the A+T content was 64% (results not shown). However, it is possible that discrete genomic regions have a significantly higher A+T content than the remainder of the *Cryptosporidium* genome (the genome of the malarial parasite *Plasmodium vivax* is made up of two components with A+T contents of 70 and 82% [21]). Local regions of long repeats or high A+T content could also limit the maximum clone size, possibly explaining the relatively small average insert size. It is worth noting that an STS containing a $[GGA]_{10}$ repeat (accession number G35348) which we sequenced from our M13 libraries was not present in the PAC library.

The PAC library described here is a powerful new resource for use in genomic studies of *C. parvum*, with >90% of sequences tested present in at least one clone. We have begun mapping the *C. parvum* genome using a HAPPY mapping (an in-vitro linkage technique based on screening *hap*-loid amounts of DNA by the *p*oly*m*erase chain reaction) strategy [22], which along with this library will provide a framework for genomic studies. Library filters and individual clones are available from the UK Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, UK CB10 1SB (http://www.hgmp.mrc.ac.uk/homepage.html).

## References


[1] Tzipori S. Cryptosporidiosis in perspective. Adv Parasitol 1988;27:63–120.

[2] Blunt DS, Khramstov NV, Upton SJ, Montelone BA. Molecular karyotype analysis of *Crytptosporidium parvum*: Evidence for eight chromosomes and a low-molecular-size molecule. Clin Diagn Lab Immunol 1997;4:11–3.

[3] Hays MP, Mosier DA, Oberst RD. Enhanced karyotype resolution of *Crytptosporidium parvum* by contour-clamped homogeneous electric fields. Vet Parasitol 1995;58:273–80.

[4] Kim K, Gooze L, Petersen C, Gut J, Nelson RG. Isolation, sequence and molecular karyotype analysis of the actin gene of *Crytptosporidium parvum*. Mol Biochem Parasitol 1992;50:105–14.

[5] Steele MI, Kuhls TL, Nida K, et al. A *Crytptosporidium parvum* genomic region encoding hemolytic activity. Infect Immun 1995;63:3840–5.

[6] Khramstov NV, Tilley M, Blunt DS, Montelone BA, Upton SJ. Cloning and analysis of a *Crytptosporidium parvum* gene encoding a protein with homology to cytoplasmic form Hsp 70. J Eukaryot Micro 1995;42:416–22.

[7] Khramstov NV, Blunt DS, Montelone BA, Upton SJ. The putative acetyl-coA synthetase gene of *Crytptosporid-*

*ium parvum* and a new conserved protein motif in acetyl-coA synthetases. J Parasitol 1996;82:423–7.

[8] Ranucci L, Müller H-M, Rossa GL, et al. Characterisation and immunolocalization of a *Crytptosporidium* protein containing repeated amino acid motifs. Infect Immun 1993;61:2347–56.

[9] Dykstra CC, Blagburn BL, Tidwell RR. Construction of genomic libraries of *Crytptosporidium parvum* and identification of antigen-encoding genes. J Protozool 1991;38:76S–8S.

[10] Taghi-Kilani R, Rimacha-Moreno M, Wenman WM. Three tandemly repeated 5S ribosomal RNA-encoding genes identified, cloned and characterized from *Cryptosporidium parvum*. Gene 1994;142:253–8.

[11] Gooze L, Kim K, Petersen C, Gut J, Nelson RG. Amplification of a *Crytptosporidium parvum* gene fragment encoding thymidylate synthase. J Protozool 1991;38:56S–8S.

[12] Jenkins MC, Fayer R, Tilley M, Upton SJ. Cloning and expression of a cDNA encoding epitopes shared by 15 and 60 kDa proteins of *Crytptosporidium parvum* sporozoites. Infect Immun 1993;61:2377–82.

[13] Petersen C, Gut J, Leech JH, Nelson RG. Identification and initial characterisation of five *Crytptosporidium parvum* sporozoite antigen genes. Infect Immun 1992;60:2343–8.

[14] Nelson RG, Kim K, Gooze L, Petersen C, Gut J. Identification and isolation of *Crytptosporidium parvum* genes encoding microtubule and microfilament proteins. J Protozool 1991;38:52S–5S.

[15] Lally NC, Baird GD, McQuay SJ, Wright F, Oliver JJ. A 2359-base pair DNA fragment from *Crytptosporidium parvum* encoding a repetitive oocyst protein. Mol Biochem Parasitol 1992;56:69–78.

[16] Ioannou PA, de Jong PJ. Construction of bacterial artificial chromosome libraries using the modified P1 (PAC) system. Curr Protoc Hum Genet 1996;5:1–24.

[17] Ioannou PA, Amemiya CT, Garnes J, et al. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat Genet 1994;6:86–9.

[18] Pierce JC, Sauer B, Sternberg N. A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficiency. Proc Natl Acad Sci USA 1992;89:2056–60.

[19] Sternberg N. Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kb pairs. Proc Natl Acad Sci USA 1990;87:103–7.

[20] Weber JL. Molecular biology of malaria parasites. Exp Parasitol 1988;66:143–70.

[21] McCutchan TF, Dame JB, Miller LH, Barnwell J. Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. Science 1984;225:808–11.

[22] Dear PH. HAPPY mapping. In: Dear PH, editor. Genome Mapping—A Practical Approach. IRL Press, 1997:95–124.

[23] Mead JR, Arrowood MJ, Current WL, Sterling CR. Field inversion gel electrophoretic separation of *Cryptosporidium* spp. chromosome-sized DNA. J Parasitol 1988;74:366–9.

[24] Sambrook J, Fritsch E, Maniatis T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Laboratory Press, 1989.