



A Physical Map of the *Leishmania major* Friedlin Genome

Alasdair C. Ivens, Sharon M. Lewis, Azadeh Bagherzadeh, et al.

Genome Res. 1998 8: 135-145

Access the most recent version at doi:[10.1101/gr.8.2.135](https://doi.org/10.1101/gr.8.2.135)

References This article cites 36 articles, 4 of which can be accessed free at:
<http://genome.cshlp.org/content/8/2/135.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the text "LEARN MORE". On the right, there is a woman wearing a red mask and a red cape, and the logo for CELLECTA, which consists of a green molecular structure and the word "CELLECTA" in white capital letters.

CRISPR and RNAi Genetic Screening.
Your new superpower.

LEARN MORE

CELLECTA

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

LETTER

A Physical Map of the *Leishmania major* Friedlin Genome

Alasdair C. Ivens,¹ Sharon M. Lewis, Azadeh Bagherzadeh,
Linglan Zhang, Ho M. Chan, and Deborah F. Smith

Department of Biochemistry, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK

An extensive physical map of the *Leishmania major* Friedlin genome has been assembled by the combination of fingerprint analysis of a shuttle vector cosmid library and probe hybridization. The integrated data obtained for 9004 fingerprinted clones and 974 probes have placed 91.2% of the 33.58-Mb genome into contigs representing each of the 36 chromosomes. This first-generation map has already provided a suitable framework for both high-throughput DNA sequencing and functional studies of the *L. major* parasite.

Leishmania is a flagellated protozoan parasite belonging to the family Trypanosomatidae. The 13 different species of *Leishmania* are responsible for a wide spectrum of human disease occurring mostly in the tropics and subtropics, although cases in North America have been documented (McHugh et al. 1996). Found in parts of Asia, the Mediterranean region, and South America, there are estimated to be over 2 million new cases of leishmaniasis each year in 88 countries, with 367 million people at risk (WHO 1995). Leishmaniasis thus constitutes a major threat to human health.

The parasite is digenic; the extracellular, flagellated forms of *Leishmania major* differentiate from noninfective promastigotes to infective metacyclics in the alimentary tract of their dipteran (sandfly) vector. After inoculation into the mammalian host, the parasites become intracellular, entering macrophages and differentiating into nonmotile amastigotes. Infective parasites are able to resist the action of resident hydrolytic enzymes, inhibit activation of the oxidative burst, and mitigate the immunological attack of the host (Bard 1989; Killick-Kendrick 1990; Blackwell 1996; Mauel 1996). As a result, there is considerable worldwide interest in determining the factors that contribute to *Leishmania*'s efficiency as a parasite.

As a eukaryote, *Leishmania* is atypical. Genes are often organized in tandem repeats, many of which are transcribed polycistronically. Nonrepeated genes of related function can also occur in long transcription units, akin to operons in prokaryotes.

Extensive post-transcriptional processing is then required to yield mature mRNAs, including the *trans*-splicing of a 39-nucleotide spliced leader (SL, or miniexon-derived) RNA onto the 5' ends of all mRNA molecules (Miller et al. 1986; Ramamoorthy et al. 1996). In contrast, no introns have been detected, removing a requirement for *cis* splicing. Hence, *Leishmania* species are the subject of fundamental interest with respect to the evolution of gene regulatory mechanisms, as well as for the exploitation of these properties for the development of new methods of control.

The application of standard genetic techniques to the study of *Leishmania* has been hampered by two factors: (1) No sexual cycle has been observed; and (2) the chromosomes do not condense at any phase of the cell cycle. However, the refinement of pulsed field gel electrophoresis (PFGE) methods over the last 10 years has enabled a molecular "karyotype" for the various *Leishmania* species to be obtained (Samaras and Spithill 1987; Pagès et al. 1989; Bastien et al. 1992). Typically, the genome of *Leishmania* comprises ~25 visible chromosomal bands, with a total haploid genome size of approximately 35 Mb. A genome of this size is amenable to dissection by complementary mapping technologies. To achieve this goal, the *Leishmania* Genome Network (LGN), supported by United Nations Development Programme/World Bank/World Health Organization (WHO), was set up in 1995.

Until relatively recently, characterization of the *Leishmania* genome relied almost exclusively on the identification of the individual chromosomes (Samaras and Spithill 1987), typically by the use of gene and/or anonymous DNA probes. More recent

¹Corresponding author.
E-MAIL a.iven@ic.ac.uk; FAX 44-171-2250960.

IVENS ET AL.

studies, by the use of 244 probe loci, enabled 36 physical linkage groups to be defined, and a genome size of ~35 Mb to be calculated (Wincker et al. 1996). Importantly, the physical linkage groups exhibited conservation among the Old World species of *Leishmania*, suggesting that overall chromosomal structure and gene order might have been maintained despite speciation and concomitant chromosome size variability (Wincker et al. 1997). This significant observation vindicated the use of a single species of *Leishmania* as a working model of the genus and provided a multispecies method for the chromosomal assignment of DNA markers in *Leishmania*. The LGN adopted the virulent *L. major* Friedlin strain (MHOM/IL/81/Friedlin) as the reference organism for genome mapping (for review, see Ivens and Blackwell 1996; Ivens and Smith 1997).

Eukaryotic organisms with small genomes (e.g., *Caenorhabditis elegans* and *Saccharomyces cerevisiae*) have been studied extensively using clone-based methods (cosmids), leading onto large-scale sequencing projects (Sulston et al. 1992; Waterston and Sulston 1995; Dujon 1996). Adopting similar approaches, we have constructed a physical map of the *L. major* Friedlin genome. A shuttle-vector cosmid library of 9216 clones (ninefold genome coverage) was analyzed by fingerprinting. Assembled into overlapping clone contigs, the clones have been assigned chromosomal map locations by the hybridization of cosmid-derived markers to PFGE-separated chromosomes and high-density gridded colony arrays.

This approach has not only resulted in the joining of contigs but has also facilitated the identification of "tile sets," the minimum number of clones required to cover individual chromosomes. A directed approach to the sequencing of the entire *L. major* Friedlin genome can now be adopted.

The data presented in this paper can be viewed in a number of ways:

1. On the LGN World Wide Web site (<http://www.ebi.ac.uk/parasites/leish.html>).
2. As a fully annotated form, which can be downloaded from the LGN ACeDB-based database LeishDB (accessed via the address given above).
3. Interactively in ACeDB (ACeDB software not required) from the U.K. Medical Research Council-funded Human Genome Mapping Project Resource Centre WWW site (<http://www.hgmp.mrc.ac.uk/Public/genome-db.html>).

Requests for clones and other information should be directed in the first instance to A.I.

RESULTS

A number of different approaches were adopted during the course of these studies. A flow diagram outline of the overall process is shown in Figure 1.

Fingerprinting

Of the 9216 clones that constitute the cLHYG cosmid library, fingerprints were obtained for 9004 clones (97.7%). Approximately 100 clones appeared to contain either no insert or had undergone rearrangement, as judged by electrophoretic mobilities of the vector fragments. Contigs, assembled by CONTASP software, were subsequently imported into the FPC package. Contigs containing three or fewer clones were disassembled, yielding 545 contigs (70% of fingerprinted clones placed in contigs; Ivens and Smith 1997). Potential end-end joins were identified by the software package EXTEND for subsequent confirmation by hybridization studies. A significant number of clones (345, ~4%) exhibited

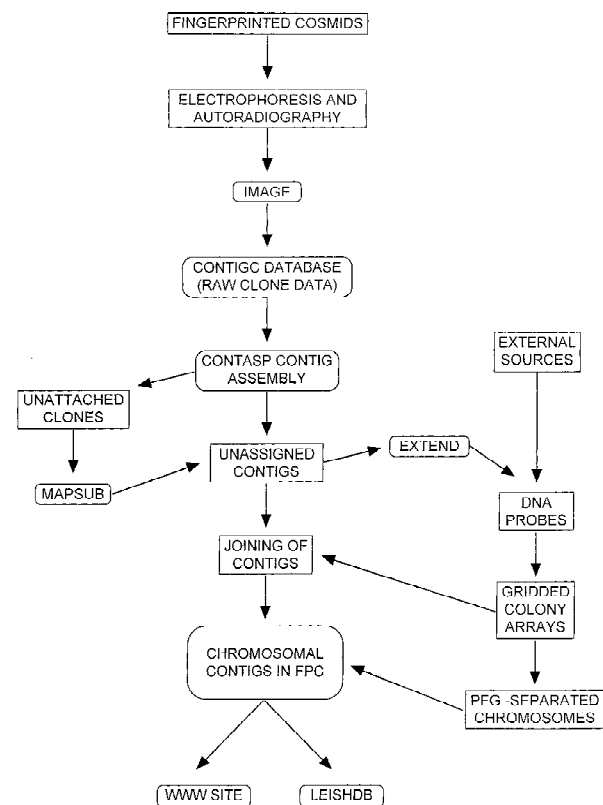


Figure 1 The processes involved in the construction of the genome map. A flow chart detailing the inter-related activities involved in the construction of the physical map of *L. major* Friedlin. Rounded boxes indicate computer-based analyses.

<10 fingerprint bands. These were not assembled into contigs by the CONTASP software and required either manual editing or analysis by hybridization.

PFGE Separation of Chromosomes and Mapping of Markers

PFGE parameters were optimized for the separation of *L. major* Friedlin chromosomes (285 kb–2.8 Mb). Using 24-cm gels, a minimum of two sets of conditions were required. Although many chromosomes could be identified individually, ~51% of the genome by size could not be resolved because of comigration. It is thus likely that the map locations of probes hybridizing to these chromosomes (7–11, 13/14, 16/17, 18/19/20/22, 21/23, 12/24, 27/29, 32/33) can only be ascertained definitively by the application of the multispecies approach described by Wincker et al. (1997). If the number of chromosomes is conserved between the different species of *Leishmania*, the data obtained from PFGE analysis suggest that the size of the *L. major* Friedlin genome is ~33.58 Mb (Fig. 2).

Loci of biological interest, in particular ESTs, had already been mapped to the individual chromosomes of a number of *Leishmania* species (Wincker et al. 1996), although a definitive karyotype for *L. major* Friedlin had not been determined. Gridded colony arrays and PFGE-separated chromosomes were hybridized with ESTs, concentrating initially on those that had been mapped by Wincker et al. (1996). The data were used to allocate numbers to the chromosomes. To date, we have hybridized 102 EST markers to gridded colony arrays and PFGE strips; colony array, and some PFGE hybridization data for an additional 166 ESTs are available.

The probe-based approach was also employed to facilitate joining of contigs assembled by fingerprinting data. Five hundred contig-end probes (CEPs) (see Methods) were isolated from cosmids at the ends of contigs and subsequently used to screen both the gridded colony arrays and PFGE strips. Hybridization data have now been obtained for a total of 974 markers (826 with mapping information), and of these, 755 (718 with mapping information) are from this laboratory. Of the 826 probes that have associated chromosomal mapping information, 114 hybridized to more than one chromosome or could not be mapped definitively (data not shown).

High-density grid hybridization data gave rise to a number of observations. The average number of positives identified by probes (excluding whole-chromosome probes) was 24, although a significant proportion was thought to be false positives. An es-

PHYSICAL MAP OF THE *L. MAJOR* FRIEDLIN GENOME

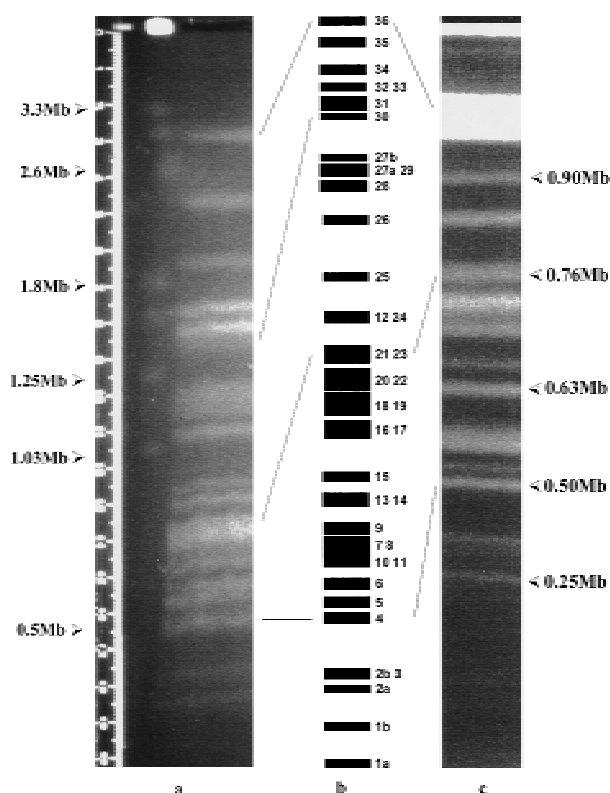


Figure 2 PFGE separation of *L. major* Friedlin chromosomes. (a) Separation of the larger chromosomes: Agarose-embedded *L. major* Friedlin DNA was prepared as described in Methods and subjected to PFGE [0.8% agarose (GIBCO), Bio-Rad DRII apparatus, 100 V, 1000- to 100-sec ramp, 250 hr, 4°C; electrophoresis buffer was changed after 90 hr]. The chromosomes of *S. cerevisiae* YNN295 (Bio-Rad) and *H. wingiae* (Promega) were used as molecular mass markers. Sizes are indicated. The chromosomes were visualized under UV after staining with ethidium bromide. (b) Ideogram of the *L. major* Friedlin karyotype. Numbering of the chromosomes was derived from this study and Wincker et al. (1996). Lines indicate the relative mobilities of selected chromosomes in panels a and c. (c) Separation of the smaller chromosomes: Agarose-embedded *L. major* Friedlin DNA was prepared as above but separated under different PFGE conditions [1.0% agarose (GIBCO), Bio-Rad DRII apparatus, 150 V, 140- to 90-sec ramp, 110 hr, 4°C].

imate for this was obtained by adding the number of contig 0 (unattached clone) positives to those hybridizing fewer than three times to a contig. Twenty-seven percent of all hybridization data was to unattached clones, and each probe identified an average of 5.8 contigs, yielding a “probable” false-positive estimate of 47%. This was achieved despite careful attention being paid to the stringency of hybridization and posthybridization washes. The

IVENS ET AL.

nominally false-positive data were not exclusively instances of weak hybridization; there were numerous examples of strongly positive hybridization to clones that could not be placed in a given contig by the fingerprint data. Caution was required when manually placing or moving clones into contigs on the basis of single hybridization results, as these clones frequently proved to be false positives when DNA sequence data were obtained (P. Myler, pers. comm.). Accordingly, all clones had to be positive for at least two probes before being placed in a contig; these artificial movements have been identified by comments in the FPC clone database.

A similar, though less marked, observation was made for a proportion of the probes hybridized to PFGE-separated chromosomes: The vast majority of the colony array data obtained for multilocus probes would identify a single chromosomal contig. In effect, this suggested that several of the multilocus probes could tentatively be reclassified as single loci, although this has not been done. Taking the above aspects into consideration, hybridization data were deemed valid only if they identified overlapping cosmid clones; adopting this criterion, analysis of the hybridization data identified 175 contig-joining events, 44 of which involved more than one probe.

A simple measure of probe utility “confidence” was obtained. Calculated as the proportion of most-valid positive hybridization, the average confidence value for all probes was 0.44. A higher value was obtained for CEP markers (0.50); the average confidence value for the EST probes was considerably poorer (0.31). Simple repeats in the untranslated regions may account for this, although the hypothesis was not investigated further. It was also observed that EST probes frequently hybridized to cosmid clones known to contain the miniexon gene array, a result consistent with the SL sequence being present in each of the EST clones.

The use of chromosomal contigs has also enabled locations for uncertain (114) or unmapped (146) probes to be tentatively inferred. The high-density grid hybridization data were analyzed; the chromosome contig that contained the highest number of positives was deemed to be the likely site of the locus. A confidence cut-off of 0.25 was chosen arbitrarily, below which point an assignment was not made. When performing the analysis for EST markers, chromosome 2 was excluded, as it is the site of the miniexon gene array. Of the 260 markers with uncertain or unknown map locations, 172 (66.2%) could be assigned a tentative chromosomal assignment (data not shown).

Chromosome Coverage

Contigs that could be assigned map locations on the basis of single-locus probe hybridization data were placed in chromosome contigs (e.g., chromosome 1 into contig 1) for ease of analysis. Chromosome contig sizes were determined as described in Methods. Sub-contigs within the chromosomal contigs that could not be joined satisfactorily by either fingerprint overlap or hybridization data were separated by uniformly sized gaps.

As described above, comigration of *L. major* Friedlin chromosomes precluded the simple assignment of markers; accordingly, contigs representative of these unresolved chromosomes were also generated (e.g., contig 37 for chromosome 32 or 33). Summing the data for all contigs, the total coverage is 30.65 Mb (91.3% of the genome; Table 1). The amount of 30.61 Mb (91.2% of the genome) has been placed in chromosomal contigs (numbered 1–44), with an average subcontig size of 125 kb. Individual chromosome contigs (contigs 1–36) account for 19.81 Mb (59.0%), with the balance (10.80 Mb, 32.2%) assigned to comigrating chromosomes. Only two small contigs (contigs 45 and 46, assigned a minimum size of 35 kb each), which likely contain telomeric and/or subtelomeric sequences, could not be assigned to a single specific chromosome. As telomeric repeat sequences are under-represented in the cosmid library (only 65 positives for 72 loci), it is likely that the genome coverage figures given above are conservative.

Sequencing the *L. major* Friedlin Genome

Having constructed chromosomal contigs, it is possible to determine tile sets for sequencing purposes. This laboratory is particularly interested in chromosomes 30 (1.4 Mb), 23 (760 kb), and 4 (460 kb) (Flinn and Smith 1992; Kelly et al. 1995; McKean et al. 1997). A whole-chromosome probe, isolated from a pulsed-field agarose gel, was used to identify contigs from chromosome 4. Subsequently, the data from 32 other probes were used to confirm the order of the subcontigs identified. SEGMAP software was initially employed to establish the most likely hybridization data-based clone order, but the high level of false positives observed precluded its universal application without extensive data manipulation. The size of the contig was estimated at 395 kb (SEGMAP) or 419 kb (FPC), ~90% of the chromosome. A tile set of 16 cosmid clones was selected. A modified contig display from FPC is shown in Figure 3.

PHYSICAL MAP OF THE *L. MAJOR* FRIEDLIN GENOMETable 1. Contig Coverage of the *L. major* Friedlin Genome

| Ctg. | Chrom. | Chrom. size (kb) | Ctg. size (CB) | Ctg. size (kb) | % chrom. | Num. of clones | Num. of subctg. | Av. subctg. size (kb) | Av. subctg. occup. | Num. of single-locus markers |
|------|-------------|------------------|----------------|----------------|----------|----------------|-----------------|-----------------------|--------------------|------------------------------|
| 1 | 1 | 285 | 161 | 216 | 75 | 131 | 1 | 216 | 131 | 32 |
| 2 | 2 | 370 | 309 | 326 | 88 | 78 | 4 | 94 | 19 | 8 |
| 3 | 3 | 370 | 288 | 238 | 64 | 47 | 6 | 53 | 7 | 8 |
| 4 | 4 | 460 | 311 | 419 | 91 | 115 | 1 | 419 | 115 | 31 |
| 5 | 5 | 500 | 578 | 539 | 107 | 150 | 9 | 74 | 16 | 13 |
| 6 | 6 | 550 | 653 | 520 | 94 | 91 | 13 | 55 | 7 | 19 |
| * | 7 | 580 | * | * | * | * | * | * | * | 0 |
| 8 | 8 | 565 | 58 | 78 | 13 | 11 | 1 | 78 | 11 | 1 |
| * | 9 | 580 | * | * | * | * | * | * | * | 0 |
| 10 | 10 | 580 | 178 | 239 | 41 | 43 | 1 | 239 | 43 | 5 |
| 11 | 11 | 570 | 55 | 74 | 12 | 8 | 1 | 74 | 8 | 1 |
| * | 12 | 830 | * | * | * | * | * | * | * | 0 |
| 13 | 13 | 630 | 393 | 499 | 79 | 94 | 2 | 258 | 47 | 2 |
| * | 14 | 630 | * | * | * | * | * | * | * | 0 |
| 15 | 15 | 650 | 379 | 360 | 55 | 114 | 6 | 73 | 19 | 9 |
| * | 16 | 710 | * | * | * | * | * | * | * | 0 |
| 17 | 17 | 710 | 373 | 412 | 58 | 48 | 4 | 115 | 12 | 4 |
| * | 18 | 710 | * | * | * | * | * | * | * | 0 |
| 19 | 19 | 720 | 154 | 207 | 28 | 55 | 1 | 207 | 55 | 2 |
| * | 20 | 720 | * | * | * | * | * | * | * | 1 |
| 21 | 21 | 760 | 68 | 91 | 11 | 6 | 1 | 91 | 6 | 1 |
| 22 | 22 | 730 | 31 | 35 | 4 | 10 | 2 | 14 | 5 | 3 |
| 23 | 23 | 760 | 786 | 909 | 119 | 291 | 6 | 165 | 48 | 47 |
| 24 | 24 | 830 | 476 | 581 | 70 | 95 | 3 | 204 | 31 | 4 |
| 25 | 25 | 900 | 787 | 760 | 84 | 161 | 11 | 84 | 14 | 17 |
| 26 | 26 | 1100 | 1210 | 1300 | 118 | 272 | 12 | 123 | 22 | 26 |
| 27 | 27 | 1200 | 239 | 202 | 16 | 57 | 5 | 53 | 11 | 7 |
| 28 | 28 | 1125 | 1159 | 1262 | 112 | 223 | 11 | 129 | 20 | 23 |
| 29 | 29 | 1150 | 418 | 473 | 41 | 71 | 4 | 130 | 17 | 11 |
| 30 | 30 | 1400 | 873 | 936 | 66 | 210 | 9 | 118 | 23 | 27 |
| 31 | 31 | 1500 | 2052 | 2285 | 152 | 667 | 17 | 150 | 39 | 39 |
| 32 | 32 | 1600 | 663 | 773 | 48 | 143 | 5 | 167 | 28 | 4 |
| 33 | 33 | 1600 | 765 | 911 | 56 | 161 | 5 | 195 | 32 | 7 |
| 34 | 34 | 2000 | 1641 | 1611 | 80 | 258 | 21 | 92 | 12 | 37 |
| 35 | 35 | 2400 | 1500 | 1631 | 67 | 428 | 14 | 131 | 30 | 29 |
| 36 | 36 | 2800 | 1826 | 1921 | 68 | 299 | 19 | 116 | 15 | 39 |
| 37 | 32/33 | 3200 | 1946 | 1932 | 60 | 340 | 24 | 96 | 14 | 45 |
| 38 | 27/29 | 2350 | 1312 | 1378 | 58 | 223 | 14 | 113 | 15 | 31 |
| 39 | 12/24 | 1660 | 1023 | 988 | 59 | 226 | 14 | 85 | 16 | 26 |
| 40 | 21/23 | 1520 | 803 | 722 | 47 | 175 | 13 | 70 | 13 | 29 |
| 41 | 18/19/20/22 | 2880 | 2821 | 2842 | 98 | 452 | 33 | 102 | 13 | 70 |
| 42 | 16/17 | 1420 | 942 | 939 | 66 | 159 | 12 | 93 | 13 | 17 |
| 43 | 13/14 | 1260 | 1122 | 942 | 74 | 216 | 20 | 62 | 10 | 14 |
| 44 | 7 to 11 | 2875 | 1010 | 1061 | 36 | 180 | 11 | 111 | 16 | 23 |
| 45 | UA | 0 | 8 | 35 | 0 | 7 | 1 | 35 | 7 | 0 |
| 46 | UA | 0 | 22 | 35 | 0 | 4 | 1 | 35 | 4 | 0 |

Subcontigs, mapped to individual chromosomes by DNA marker loci, were placed in representative chromosomal contigs. The sizes of these contigs were estimated from CB coverage (see Methods for calculation details), with the exception of contigs 45 and 46 (8 and 22 CBs, respectively), which were assigned a default minimum size of 35 kb.

(*)No significant colony array hybridization data for single-locus markers assigned to this chromosome, so contigs not assembled (but see contigs 37–44). (CB) Consensus bands; (UA) unable to assign contig to a specific chromosome.

IVENS ET AL.

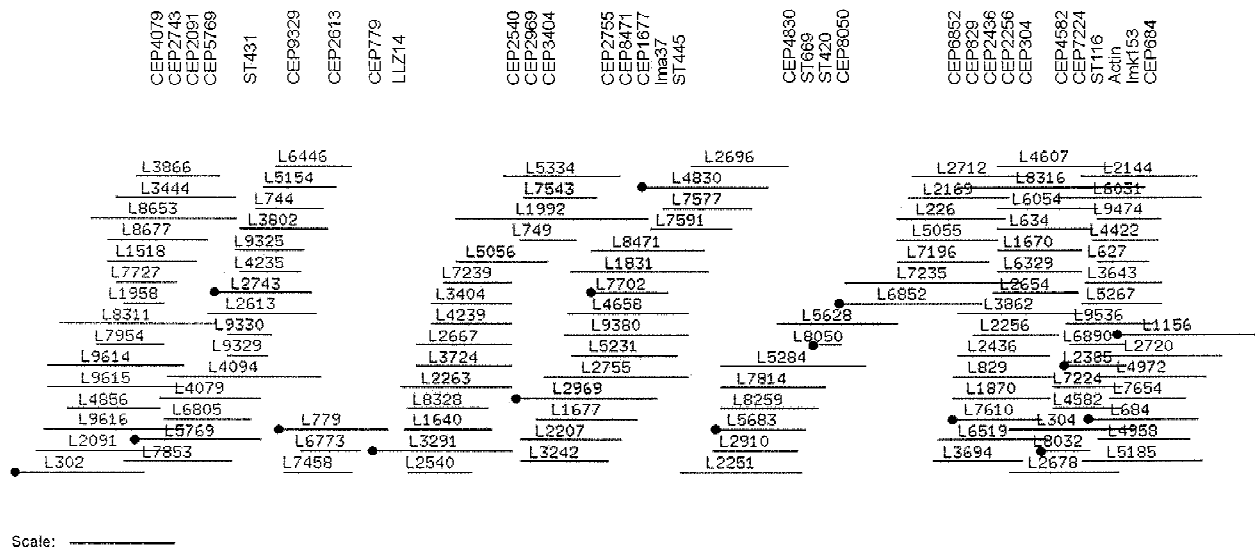


Figure 3 Minimum tile set for chromosome 4. *L. major* Friedlin cosmid clones were fingerprinted, as detailed in Methods, prior to contig assembly by CONTASP and FPC v. 2.8.2 software. Subcontigs shown to contain chromosome 4 marker loci were joined by (1) the application of additional software packages: MAPSUB (identified loose fingerprint matches), EXTEND (identified potential end–end joins), and (2) hybridization data (assigned and joined subcontigs), to yield an integrated map. The contig display from FPC formed the basis of this diagram. All clones that constitute the contig, including those that have been buried, are shown as horizontal lines. DNA hybridization probes are shown along the *top* (see Methods for probe descriptions). A minimum tile set (●) was chosen for DNA sequencing purposes. The scale bar represents 26 consensus bands (~35 kb).

A problem was encountered with L8050, the only clone in the cosmid library positive for markers ST445 and ST669. CEP8050.1.2, a probe generated from the cosmid, positioned L8050 within the contig; this was confirmed by additional probe data from clone L4830 (CEP4830.1.7). However, these probe data are inconsistent, as clones positive for both CEP8050.1.2 and CEP4830.1.7 should also be positive for markers ST445 and ST669. It seems likely that this region of the chromosome is unstable and that cosmids positive for the two CEP probes no longer contain the loci identified by the ST probes. For this reason, the region encompassing the ST445/ST669 loci was deliberately over-represented in the tile set. DNA sequence information obtained by shotgun cloning of PFGE-separated chromosome 4 should eradicate this anomaly.

DISCUSSION

Contig Assembly

Variation in contig depth was observed. The data suggest that the library was over-representative for certain (probably coding) regions, and under-represented for others (e.g., telomere repeat se-

quences). It thus appears that the *Leishmania* genome is similar to many other genomes in that it contains sequences not readily maintained in *Escherichia coli* cloning vectors. Coupled with the use of *Sau3AI* for cosmid library construction, these factors may have played a part in preventing the assembly of larger initial contigs by CONTASP software. Despite these impediments, it was always possible to identify at least one cosmid, and usually many more, for all 974 DNA marker probes hybridized to the gridded library.

As noted above, a small but significant proportion of fingerprinted clones exhibited very few fingerprint bands. This had two major effects: (1) These clones were not placed into contigs because of the probability cutoff used; and (2) contigs constructed manually from these clones resulted in an incorrect estimate of contig size. As an example, the miniexon gene locus, mapped to chromosome 2, is known to consist of a tandem array of ~100 identical or nearly identical genes covering ~50 kb (Iovanisci and Beverley 1989; Fernandes et al. 1994). Fingerprints from cosmid clones containing the miniexon gene sequence typically comprised five strongly labeled bands; clones could therefore not be ordered relative to each other. The small number of bands not only made it impossible to join the

miniexon stack to other subcontigs known to map to chromosome 2 but has also led to an underestimate of chromosome 2 contig coverage. Similar problems have been encountered with a number of other tandem arrays, for example, LmcDNA2 and LmcDNA16 (Flinn and Smith 1992; Kelly et al. 1995; McKean et al. 1997), gp63 (Voth et al. 1998), and the histone H1 (sw3) loci (Fasel et al. 1993; Noll et al. 1997). Given this fact, it is possible that many contig size estimations, based solely on the number of consensus bands, under-represent their true sizes and that the physical map presented here is nearer completion than the data suggest.

DNA Probe Hybridization to PFGE-Separated *Leishmania* Chromosomes

The separation of *L. major* Friedlin chromosomes could be achieved using a minimum of two sets of electrophoretic conditions, as shown in Figure 2. Modification of these conditions to selectively expand certain size ranges did not result in the complete separation of all chromosomes. Comigration, as evidenced by both staining intensity and hybridization analyses, remains a feature of the *L. major* Friedlin genome.

A large number of DNA markers, 974, were used in these studies, yielding an average probe density of one per 34.5 kb. Of these markers, 712 (73.1%) mapped to a single locus and were used to assemble the chromosomal contigs. Of the single-locus markers, 255 hybridized as a single band to comigrating chromosomes. To clarify their exact localization, these DNA markers are currently being hybridized to multispecies PFGE blots, thereby taking advantage of interspecies chromosomal size polymorphisms, as described recently by Wincker et al. (1997).

Coverage

The paucity of telomeric sequences in the cosmid library suggested that the coverage figures obtained may be conservative. If telomere and/or subtelomere repeats were to constitute 5–10 kb at the end of each chromosome (totaling 360–720 kb; K. Stuart, pers. comm.), the approximate proportion of nontelomere DNA assigned to chromosomes by this study would be closer to 92.7% of the genome. These figures suggest strongly that despite some variations in contig depth, the coverage of the genome by contigs is close to complete and that the majority of coding sequences have probably been cloned. We do not have sufficient data to determine

PHYSICAL MAP OF THE *L. MAJOR* FRIEDLIN GENOME

whether the distribution of genes in the telomere regions of *Leishmania* is more similar to *C. elegans* than to *Trypanosoma brucei*. Are there regulated expression sites at the telomeres of *Leishmania* as have been observed for the VSG genes in *T. brucei* (Rudenko et al. 1996)? The answer to this question will highlight the relative importance of possessing physical maps that stretch contiguously from telomere to telomere.

Although it is difficult at this stage to generate accurate figures for the physical extent of the presumed gaps, these could be determined hypothetically by restriction enzyme mapping. In addition, the need for gap closure has initiated a number of supplementary approaches. Cosmid clone end sequencing is particularly useful for determining the extent of clone overlap, whereas the construction and integration of PAC library clone data should enable the majority of gaps to be bridged, confirm the order of subcontigs, and also provide an alternative minimum tile set source for DNA sequencing templates.

DNA sequence data obtained from six chromosome 1 cosmids by the Stuart laboratory, also a member of the LGN, have been submitted to GenBank (1997) and can be accessed via the sequencing page of the LGN WWW site (<http://www.ebi.ac.uk/parasites/LGN/leishseq.html>). Analysis of the finished sequence data of five cosmids suggests that although the telomere regions have not been sequenced yet, gene density is relatively high. There are 51 potential ORFs within 176.4 kb of sequence, or one gene per 3.46 kb. As a significant proportion of the putative genes are novel, these data underline the high priority of further sequence data acquisition.

Future Benefits and Activities

Ultimately, the development of prophylactic agents for the control and/or eradication of all types of leishmaniasis remains the primary goal of workers doing research on this parasite. A number of benefits will ensue from the physical map of the *L. major* Friedlin genome and the large quantities of primary DNA sequence data that can be efficiently obtained from this source material. Broad questions as to the behavioral dynamics of the genome can be addressed. Similarity searches of DNA and protein databases have already enabled the identification of orthologs and paralogs of many genes; novel genes have also been identified (see the WWW BLAST analysis pages at <http://www.ebi.ac.uk/parasites/LGN/lshblstn.html> or [lshblstx.html](http://www.ebi.ac.uk/parasites/LGN/lshblstx.html)).

IVENS ET AL.

The data obtained for these and other genes will provide additional entry points to those studying the functional aspects of parasite virulence that have a direct bearing on the identification of new targets for chemotherapy. Several recently developed techniques are facilitating these investigations. Targeted gene disruption by homologous recombination (Cruz and Beverley 1990; Cruz et al. 1991; Cruz et al. 1993) can generate defined mutants for complementation using shuttle vectors (e.g., cLHYG, the cosmid vector used in these studies; Ryan et al. 1993); random gene inactivation and gene trapping can be achieved by transposon mutagenesis (Gueiros-Filho and Beverley 1997); and the temporal and spatial trafficking of proteins in vivo can be monitored via green fluorescent protein (GFP) constructs (Ha et al. 1996). In the mammalian host, vaccine candidates can be assayed following the inoculation of either recombinant protein or by DNA vaccination (Gardner et al. 1996; Whalen 1996; Ulmer et al. 1996).

All of these approaches will contribute to our understanding of the mechanisms underlying *Leishmania* infectivity. The physical and biological information gained from the genome mapping and sequencing initiatives will be of value not only to those interested in *Leishmania* and other trypanosomes but also to those exploring some of the more unusual aspects of eukaryotic gene expression.

METHODS

Fingerprinting

A *L. major* Friedlin cosmid library of 9216 clones, constructed by ligation of *Sau3AI* partially digested DNA into the *Bam*HI site of the cLHYG shuttle vector (Ryan et al. 1993), was picked into microtiter plates. The entire library was analyzed by fingerprinting, as described previously (Coulson and Sulston 1988; Ivens and Little 1995). DNA samples were digested with *Hinf*I, the termini end-filled in the presence of [³²P]dCTP, and the resulting products separated by electrophoresis (35 W, for ~90 min) on 4% nondenaturing polyacrylamide gels prior to autoradiography. Bacteriophage λ , digested with *Hinf*I and radiolabeled by end-filling in the presence of ³⁵S-labeled dATP, acted as a normalizing internal standard. The cosmid library was also gridded onto nylon membranes (GeneScreen Plus, DuPont) using a Biomek 1000 robotic workstation (Bentley et al. 1992). Membranes were prepared for hybridization by successive treatments with 0.5 M NaOH/1.5 M NaCl, and 1 M sodium phosphate buffer (pH 6.8), followed by baking at 80°C and UV cross-linking (Sambrook et al. 1989).

Image and Contig Analysis

Fingerprint autoradiographs were scanned using an Amer sham scanner. The resulting image files were analyzed by the IMAGE 2.1 software package to generate a clone fingerprint

database. Overlapping fingerprints within this database were initially identified by CONTASP software, using a band match tolerance of 1 mm and a probability cutoff of $1e^{-07}$, and contigs viewed using CONTIGC (Sulston et al. 1988, 1989).

Joining of these contigs initially involved MAPSUB analysis, followed by EXTEND (see Fig. 1). The MAPSUB software package matches fingerprints at user-specified tolerance and probability cutoff values. Match likelihood of an unassembled clone database indicated that for an average 26-band clone, the probability cutoff used for contig assembly would only identify clones exhibiting 54% (tolerance, 0.7 mm) or 62.5% (tolerance, 1.0 mm) or greater overlap with another 26-band clone. The application of higher probabilities (e.g., $1e^{-05}$, for 42% and 50% overlap, respectively) yielded an unacceptable level of multiply inconsistent overlaps and was disregarded. Although exhaustive match analysis with a probability cutoff of $1e^{-02}$ indicated that each clone had, on average, 35 or 44 (tolerance, 0.7 and 1.0 mm, respectively) other matching clones, only a small proportion (8.8% and 7.7%, respectively) matched with a probability below $1e^{-07}$. These figures were considerably lower for clones placed at the ends of the contigs (data not shown).

More recently, the FPC (v 2.8.2) software package has been used (C. Soderlund, pers. comm.). The contig data were imported into the FPC software, and contigs were refined by analysis at tolerance 0.7 mm. Subcontigs that could be joined by hybridization data were placed next to each other, separated by 1 consensus band (CB) calculations, whereas unjoined subcontigs have been separated by 10 CB gaps. All subcontig manipulations have been annotated in the publicly available database, LeishDB.

The number of nonvector fragments resolved by the electrophoresis conditions was $26 [\pm 8.5 (1 \text{ s.d.})]$ bands per clone. Assuming that each cosmid clone contains, on average, 35 kb of inserted DNA, each band thus approximates to 1.347 kb; this value was subsequently applied to all CB calculations of contig size and chromosome coverage. The lowest limit of overlap used by the contig assembly software (~50% cosmid overlap, or 17.5 kb) was used as the minimum gap size. Contig sizes were calculated as follows: [(CB bands) \times 1.347] - [30 \times (no. of gaps)] = size (kb). Although this calculation tends to underestimate the true coverage of the individual subcontigs, it compensates for the "false" separation of hybridization-based joining events.

PFGE

Clamped homogeneous electric field PFGE was used to separate the chromosomes of *L. major* Friedlin (MHOM/IL/81 Friedlin). Parasites were grown at 24°C for 96 hr in vitro [*Drosophila* Schneider's medium (GIBCO)/15% heat-inactivated bovine calf serum (GIBCO)/100 μ g/ml of gentamycin (Sigma)], harvested by centrifugation, and washed at room temperature in 20 ml of PBS (Coulson and Smith 1990). Following centrifugation, the parasites were thoroughly resuspended in 6 ml of PBS, and an equal volume of prewarmed (50°C) 1% LGT [low gelling temperature (Seaplaque, Flowgen)] agarose/PBS was added. The parasite-agarose mixture was aliquoted (100 μ l) into precooled (4°C) block formers (Pharmacia) and allowed to set. Parasites were lysed by extruding the blocks into 40 ml of LiDS mix [1% (wt/vol) lithium dodecylsulfate, 100 mM EDTA, 20 mM Tris-HCl at pH 8.0], followed by incubation at 50°C for 24 hr. Blocks were washed extensively in TE buffer (10 mM Tris-HCl, 1 mM EDTA

PHYSICAL MAP OF THE *L. MAJOR* FRIEDLIN GENOME

at pH 8.0) prior to electrophoresis in $1 \times$ TB, $0.1 \times$ E, using either modified Gene Navigator (Pharmacia) or DRII (Bio-Rad) equipment. Pulse times, voltages, and/or agarose concentrations were varied to achieve different size separation ranges; all gels were transferred to GeneScreen Plus nylon membranes for hybridization studies (Sambrook et al. 1989). When chromosomal probes required isolation, LGT agarose was used. Size standards were *S. cerevisiae* YNN295 (Bio-Rad) or *Hansenula wingiae* (Promega) chromosomes.

Hybridization Markers

Assembled contigs were analyzed by EXTEND software to identify potential end-end joins, with the resulting data forming the initial basis of the hybridization approach. CEPs were isolated from cosmids located at the ends of contigs. DNAs were isolated using standard miniprep procedures (Ivens and Little 1995) prior to restriction endonuclease digestion (typically *Pst*I) and fragment separation by electrophoresis. Fragments (0.5–3.0 kb), isolated from LGT agarose gels, were subsequently radiolabeled as described below and used as probes against the gridded cosmid colony arrays and PFGE-separated *L. major* chromosomes.

Several other classes of hybridization probes were employed. ESTs, some previously assigned by Wincker et al. (1996), were provided as PCR products and are prefixed by Im or Lm (Levick et al. 1996; H. Schneider and A. Cruz, pers. comm.). Chromosomally mapped probes, prefixed by ST or ISA, were provided as described (Wincker et al. 1996). "Walking" probes from *L. major* LV39 are prefixed by LV. Whole-chromosome probes, isolated from PFGE gels as described above, are prefixed by Chr. Probes prefixed by CEP have been described, whereas those suffixed by AB were isolated from anonymous cosmids that gave few (less than six) fingerprint bands. Probes generated by in vitro transcription of cosmid clones (P. Myler, pers. comm.) are either suffixed by T3, T7, and/or PCR or prefixed by 10.

Hybridization

DNA probes were radiolabeled with [32 P]dCTP by random priming (Feinberg and Vogelstein 1984) prior to purification by G-50 Sephadex chromatography. Hybridizations were carried out at 65°C in Church buffer (0.5 M NaPi, 7% SDS), and washed at high stringency ($0.1 \times$ SSC, 0.1% SDS, at 65°C) prior to autoradiography with intensifying screens at –70°C. The locations of positively hybridizing cosmid clones on high-density colony arrays were identified using a template (Ivens and Little 1995).

ACKNOWLEDGMENTS

We thank the following: Jon Warner for the *L. major* cosmid library (constructed using genomic DNA isolated by D.F.S.), Steve Beverley for the cLHYG cosmid shuttle vector, Jennie Blackwell, Horacio Schneider, and Ira Sampaio for EST markers, Angela Cruz for EST hybridization data, Patrick Bastien and Patrick Wincker for ST/ISA probes, Peter Myler for chromosome 1 probe and SEGMAP data, Nicolas Fasel for sw3, Michal Shapira for Hsp83, Paul McKean for pgmr51 and pgmd1, Liz Bates for LmPABPH3, and Arve Osland for clone 7 and clone 9. We also thank Peter Little and Holger Hummer-

ich for access to the Biomek Workstation and for many useful discussions. We acknowledge the Sanger Centre (Cambridge, U.K.) for MAPSUB, CONTASP, CONTIGC, and IMAGE software packages (available by anonymous FTP), Cari Soderlund for FPC v 2.8.2, and Chuck Magness (Washington University, Seattle) for SEGMAP. This research has been funded by the Medical Research Council (MRC) (to A.C.I. and D.F.S., grants G9402457 and G9503213PB, respectively; L.Z. was funded for 10 months from MRC grant G9402457; S.M.L. from grant G9503213PB), The Royal Society (A.C.I., grant 15836), the Nuffield Foundation (A.C.I., grant AT/100/95/0008; 2 months support for H.M.C.), and the Wellcome Trust (A.C.I., grants VS/94/IMP/003 and VS/95/IMP/002). Research in the laboratory of D.F.S. is supported by the Wellcome Trust (no. 039006/Z/93/Z/1.4A/PMC/MW and no. 045493/Z/95/Z/JRS/JH). This investigation also received financial support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) (D.F.S., grant 950522; support for A.B.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Bard, E. 1989. Molecular biology of *Leishmania*. *Biochem. Cell. Biol.* 67: 516–524.
- Bastien, P., C. Blaineau, and M. Pagès. 1992. Molecular karyotype analysis in *Leishmania*. In *Subcellular biochemistry* (ed. J.L. Avila and J.R. Harris), Vol. 18, pp. 131–187. Plenum Press, New York, NY.
- Bentley, D.R., C. Todd, J. Collins, J. Holland, I. Dunham, S. Hassock, A. Bankier, and F. Gianelli. 1992. The development and application of automated gridding for efficient screening of yeast and bacterial ordered libraries. *Genomics* 12: 534–541.
- Blackwell, J.M. 1996. Genetic susceptibility to leishmanial infections: Studies in mice and man. *Parasitology* (Suppl.) 112: S67–S74.
- Coulson, A. and J. Sulston. 1988. Genome mapping by restriction fingerprinting. In *Genome mapping: A practical approach* (ed. K.E. Davies), pp. 19–39. IRL Press, Oxford, UK.
- Coulson, R.M. and D.F. Smith. 1990. Isolation of genes showing increased or unique expression in the infective promastigotes of *Leishmania major*. *Mol. Biochem. Parasitol.* 40: 63–75.
- Cruz, A.K. and S.M. Beverley. 1990. Gene replacement in parasitic protozoa. *Nature* 348: 171–174.
- Cruz, A.K., C.M. Coburn, and S.M. Beverley. 1991. Double targeted gene replacement for creating null mutants. *Proc. Natl. Acad. Sci.* 88: 7170–7174.
- Cruz, A.K., R. Titus, and S.M. Beverley. 1993. Plasticity in chromosome number and testing of essential genes in

IVENS ET AL.

- Leishmania* by targeting. *Proc. Natl. Acad. Sci.* 90: 1599–1603.
- Dujon, B. 1996. The yeast genome project: What did we learn? *Trends Genet.* 12: 263–270.
- Fasel, N.J., D.C. Robyr, J. Mauel, and T.A. Glaser. 1993. Identification of a histone H1-like gene expressed in *Leishmania major*. *Mol. Biochem. Parasitol.* 62: 321–323.
- Feinberg, A.P. and B. Vogelstein. 1984. A technique for radiolabeling DNA restriction fragments to high specific activity. *Anal. Biochem.* 137: 266–267.
- Fernandes, O., V.K. Murthy, U. Kurath, W.M. Degraeve, and D.A. Campbell. 1994. Mini-exon gene variation in human pathogenic *Leishmania* species. *Mol. Biochem. Parasitol.* 66: 261–271.
- Flinn, H.M. and D.F. Smith. 1992. Genomic organization and expression of a differentially regulated gene family from *Leishmania major*. *Nucleic Acids Res.* 20: 755–762.
- Gardner, M.J., D.L. Doolan, R.C. Hedstrom, R. Wang, M. Sedegah, R.A. Gramzinski, J.C. Aguiar, H. Wang, M. Margalith, P. Hobart, and S.L. Hoffman. 1996. DNA vaccines against malaria: Immunogenicity and protection in a rodent model. *J. Pharm. Sci.* 85: 1294–1300.
- Gueiros-Filho, F.J. and S.M. Beverley. 1997. Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan *Leishmania*. *Science* 276: 1716–1719.
- Ha, D.S., J.K. Schwarz, S.J. Turco, and S.M. Beverley. 1996. Use of the green fluorescent protein as a marker in transfected *Leishmania*. *Mol. Biochem. Parasitol.* 77: 57–64.
- Iovannisci, D.M. and S.M. Beverley. 1989. Structural alterations of chromosome 2 in *Leishmania major* as evidence for diploidy, including spontaneous amplification of the mini-exon array. *Mol. Biochem. Parasitol.* 34: 177–188.
- Ivens, A.C. and J.M. Blackwell. 1996. Unravelling the *Leishmania* genome. *Curr. Opin. Genet. Dev.* 6: 704–710.
- Ivens, A.C. and P.F.R. Little. 1995. Cosmid clones and their application to genome studies. In *DNA cloning 3: Complex genomes* (ed. D.M. Glover and B.D. Hames), pp. 1–47. IRL Press, Oxford, UK.
- Ivens, A.C. and D.F. Smith. 1997. A global map of the *Leishmania major* genome: Prelude to genome sequencing. *Trans. R. Soc. Trop. Med. Hyg.* 91: 111–115.
- Kelly, B.L., S.D. Dyal, J. Warner, J. Tang, and D.F. Smith. 1995. Chromosomal organization of a repeated gene cluster expressed in mammalian stages of *Leishmania*. *Gene* 163: 145–149.
- Killick-Kendrick, R. 1990. The life-cycle of *Leishmania* in the sandfly with special reference to the form infective to the vertebrate host. *Ann. Parasitol. Hum. Comp.* (Suppl. 1) 65: 37–42.
- Levick, M.P., J.M. Blackwell, V. Connor, R.M.R. Coulson, A. Miles, H.E. Smith, K.-L. Wan, and J.W. Ajioka. 1996. An expressed sequence tag analysis of full-length, spliced-leader cDNA libraries from *Leishmania major* promastigotes. *Mol. Biochem. Parasitol.* 76: 345–348.
- Mauel, J. 1996. Intracellular survival of protozoan parasites with special reference to *Leishmania* spp., *Toxoplasma gondii* and *Trypanosoma cruzi*. *Adv. Parasitol.* 38: 1–51.
- McHugh, C.P., P.C. Melby, and S.G. LaFon. 1996. Leishmaniasis in Texas: Epidemiology and clinical aspects of human cases. *Am. J. Trop. Med. Hyg.* 55: 547–555.
- McKean, P.G., R. Delahay, P.F.P. Pimenta, and D.F. Smith. 1997. Characterization of a second protein encoded by the differentially regulated LmcDNA16 gene family of *Leishmania major*. *Mol. Biochem. Parasitol.* 85: 221–231.
- Miller, S.L., S.M. Landfear, and D.F. Wirth. 1986. Cloning and characterization of a *Leishmania* gene encoding a RNA spliced leader sequence. *Nucleic Acids Res.* 14: 7341–7360.
- Noll, T.M., C. Desponds, S.T. Belli, T.A. Glaser, and N.J. Fasel. 1997. Histone H1 expression varies during the *Leishmania major* life cycle. *Mol. Biochem. Parasitol.* 84: 215–227.
- Pages, M., P. Bastien, F. Veas, V. Rossi, M. Bellis, P. Wincker, J.A. Rioux, and G. Roizes. 1989. Chromosome size and number polymorphism in *Leishmania infantum* suggest amplification/deletion and possible genetic exchange. *Mol. Biochem. Parasitol.* 36: 161–168.
- Ramamoorthy, R., Donelson, J.E. and Wilson, M.E. 1996. 5' sequences essential for trans-splicing of msp (gp63) RNAs in *Leishmania chagasi*. *Mol. Biochem. Parasitol.* 77: 65–76.
- Rudenko, G., R. McCulloch, A. Dirks-Mulder, and P. Borst. 1996. Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 80: 65–75.
- Ryan, K.A., S. Dasgupta, and S.M. Beverley. 1993. Shuttle cosmid vectors for the trypanosomatid parasite *Leishmania*. *Gene* 131: 145–150.
- Samaras, N. and T.W. Spithill. 1987. Molecular karyotype of five species of *Leishmania* and analysis of gene locations and chromosomal rearrangements. *Mol. Biochem. Parasitol.* 25: 279–291.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sulston, J., F. Mallet, R. Staden, R. Durbin, T. Horsnell, and A. Coulson. 1988. Software for genome mapping by fingerprinting techniques. *Comput. Appl. Biosci.* 4: 125–132.
- Sulston, J., F. Mallet, R. Durbin, and T. Horsnell. 1989. Image analysis of restriction enzyme fingerprint autoradiograms. *Comput. Appl. Biosci.* 5: 101–106.

PHYSICAL MAP OF THE *L. MAJOR* FRIEDLIN GENOME

- Sulston, J., Z. Du, K. Thomas, R. Wilson, L. Hillier, R. Staden, N. Halloran, P. Green, J. Thierry-Mieg, L. Qiu, S. Dear, A. Coulson, M. Craxton, R. Durbin, M. Berks, M. Metzstein, T. Hawkins, R. Ainscough, and R. Waterston. 1992. The *C. elegans* genome sequencing project: A beginning. *Nature* 356: 37–41.
- Ulmer, J.B., J.C. Sadoff, and M.A. Liu. 1996. DNA vaccines. *Curr. Opin. Immunol.* 8: 531–536.
- Voth, B.R., B.L. Kelly, P.B. Joshi, A.C. Ivens, and W.R. McMaster. 1998. Developmentally expressed *Leishmania major* gp63 genes encode cell surface leishmanolysin containing distinct signals for GPI attachment. *Mol. Biochem. Parasitol.* (in press).
- Waterston, R. and J. Sulston. 1995. The genome of *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* 92: 10836–10840.
- Whalen, R.G. 1996. DNA vaccines for emerging infectious diseases: What if? *Emerg. Infect. Dis.* 2: 168–175.
- Wincker, P., C. Ravel, C. Blaineau, M. Pagès, Y. Jauffret, J.-P. Dedet, and P. Bastien. 1996. The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Res.* 24: 1688–1694.
- Wincker, P., C. Ravel, C. Britto, P. Dubessay, P. Bastien, M. Pagès, and C. Blaineau. 1997. A direct method for the chromosomal assignment of DNA markers in *Leishmania*. *Gene* 194: 77–80.
- World Health Organization (WHO). 1995. WHO Leishmaniasis World Wide Web Home Page. <http://www.who.ch/programmes/ctd/diseases/leis/leismain.htm>.

Received September 23, 1997; accepted in revised form January 14, 1998.