

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 5, 2009):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/309/5731/134>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/309/5731/134/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/309/5731/134#related-content>

This article **cites 26 articles**, 7 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/309/5731/134#otherarticles>

This article has been **cited by** 86 article(s) on the ISI Web of Science.

This article has been **cited by** 15 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/309/5731/134#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Genome Sequence of *Theileria parva*, a Bovine Pathogen That Transforms Lymphocytes

Malcolm J. Gardner,^{1*} Richard Bishop,² Trushar Shah,²
Etienne P. de Villiers,² Jane M. Carlton,¹ Neil Hall,¹ Qinghu Ren,¹
Ian T. Paulsen,¹ Arnab Pain,³ Matthew Berriman,³
Robert J. M. Wilson,⁴ Shigeharu Sato,⁴ Stuart A. Ralph,⁵
David J. Mann,⁶ Zikai Xiong,³ Shamira J. Shallom,¹
Janice Weidman,¹ Lingxia Jiang,¹ Jeffery Lynn,¹ Bruce Weaver,¹
Azadeh Shoaibi,¹ Alexander R. Domingo,¹ Delia Wasawo,²
Jonathan Crabtree,¹ Jennifer R. Wortman,¹ Brian Haas,¹
Samuel V. Angiuoli,¹ Todd H. Creasy,¹ Charles Lu,^{1†}
Bernard Suh,^{1‡} Joana C. Silva,¹ Teresa R. Utterback,¹
Tamara V. Feldblyum,¹ Mihaela Pertea,¹ Jonathan Allen,¹
William C. Nierman,¹ Evans L. N. Taracha,² Steven L. Salzberg,¹
Owen R. White,¹ Henry A. Fitzhugh,^{2§} Subhash Morzaria,^{2||}
J. Craig Venter,⁷ Claire M. Fraser,¹ Vishvanath Nene¹

We report the genome sequence of *Theileria parva*, an apicomplexan pathogen causing economic losses to smallholder farmers in Africa. The parasite chromosomes exhibit limited conservation of gene synteny with *Plasmodium falciparum*, and its plastid-like genome represents the first example where all apicoplast genes are encoded on one DNA strand. We tentatively identify proteins that facilitate parasite segregation during host cell cytokinesis and contribute to persistent infection of transformed host cells. Several biosynthetic pathways are incomplete or absent, suggesting substantial metabolic dependence on the host cell. One protein family that may generate parasite antigenic diversity is not telomere-associated.

Theileria parva is a tick-borne parasite that causes a fatal disease in cattle known as East Coast fever (ECF). This disease, which kills over 1 million cattle each year in sub-Saharan Africa, results in economic losses exceeding \$200 million annually (1). *Theileria* organisms belong to the phylum Apicomplexa, which is predicted to have originated about 930 million years ago (2). Unlike other apicomplexans,

penetration of host cells by *T. parva* is not orientation-specific. Rhoptries and microspheres discharge after invasion, coincident with dissolution of the surrounding host cell membrane, leaving the parasite free in the host cell cytoplasm. Morbidity and mortality due to ECF are attributed to the ability of the schizont stage to malignantly transform its host cell, the bovine lymphocyte. Parasitosis increases exponentially because the schizont divides in synchrony with the host cell and infected cells infiltrate all tissues; cattle die of this lymphoproliferative disease 3 to 4 weeks after infection. Little pathology is due to the tick infective piroplasm, the red blood cell stage (1).

We sequenced the genome of *T. parva* in order to facilitate research on parasite biology, assist the identification of schizont antigens for vaccine development (3), and extend comparative apicomplexan genomics, in particular with *Plasmodium falciparum*, which causes malaria. Comparison with *T. annulata*, which causes tropical bovine theileriosis and mainly transforms macrophages, is described in an accompanying report (4). (This whole-genome shotgun project has been deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under the project accession AAGK00000000.)

The haploid *T. parva* nuclear genome is 8.3×10^6 base pairs (Mbp) in length and consists of four chromosomes (Table 1). We provide a complete sequence, except for a 1- to 2-kbp gap in chromosome 4 and a gap in chromosome 3 (Tpr locus) that contains a 41-kbp and a 13-kbp set of overlapping sequences (contig) (5). The parasite apicoplast and mitochondrial (6) genomes have also been sequenced. Like *P. falciparum*, *T. parva* chromosomes contain one extremely A+T-rich region (>97%) about 3 kbp in length that may be the centromere. The regions between the CCCTA_{3,4} telomeric repeats and the first protein-encoding gene are short, 2.9 kbp on average, and do not contain other repeats. Thus, the structure of the subtelomeric regions in *T. parva* is much less complex than that in *P. falciparum*, where arrays of repeats extend up to 30 kbp (7).

The *T. parva* nuclear genome contains about 4035 protein-encoding genes, 20% fewer than *P. falciparum*, but exhibits higher gene density, a greater proportion of genes with introns, and shorter intergenic regions. There are two identical, unlinked 5.8S-18S-28S rRNA units, suggesting that unlike *P. falciparum* *T. parva* does not possess functionally distinct ribosomes (8). Putative functions were assigned to 38% of the predicted proteins (Table 1).

The complexity of the *T. parva* life cycle is not matched by a large number of recognizable cell cycle regulators. Thus, the parasite is more akin to yeasts than higher eukaryotes, lacking discernable components of both the p53-MDM2-p14ARF-p21 and the Ink4-retinoblastoma-E2F pathways (9). There are four predicted cyclins and five cyclin-dependent kinases (cdks), most of which have close homologs in *P. falciparum*. However, *T. parva* lacks one cyclin and two cdks found in *P. falciparum*. These parasite cyclins are poorly conserved (~25% identity), making cross-species comparisons difficult. The reduced recognizable *T. parva* cell cycle machinery suggests that a number of novel regulatory features remain to be discovered.

A unique aspect of *T. parva* biology is that infection of T and B lymphocytes results in a reversible transformed phenotype with uncontrolled proliferation of host cells that remain persistently infected. Parasite proteins that may modulate host cell phenotype are described in an accompanying report (4). Host cell microtubules that decorate the surface of schizonts are captured by the host cell spindle during mitosis, favoring infection of both daughter cells (1). *T. parva* encodes putative secreted forms of EMAP115- and Tau-like proteins, which are absent from *P. falciparum*; in higher eukaryotes, these proteins interact with microtubules (10). In addition, *T. parva* may modulate host cell mitosis by influencing disassembly of the host cell spindle via a secreted cdc48-like AAA-adenosine triphosphatase

¹Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville, MD 20850, USA. ²International Livestock Research Institute, Post Office Box 30709, Nairobi, Kenya. ³Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ⁴National Institute for Medical Research, Ridgeway, Mill Hill, London NW7 1AA, UK. ⁵Institut Pasteur, 25 Rue du Docteur Roux, 75724 Paris Cedex 15, France. ⁶Department of Biological Sciences, Imperial College, London SW7 2AZ, UK. ⁷Venter Institute, 9708 Medical Center Drive, Rockville, MD, 20850, USA.

*To whom correspondence should be addressed. E-mail: gardner@tigr.org

†Present address: Lewis Thomas Lab, Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA.

‡Present address: Baskin School of Engineering, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA.

§Present address: 3709 Summercrest, Ft. Worth, TX 76109, USA.

||Present address: Food and Agriculture Organization (FAO), 39 Phra Atit Road, Bangkok 10200, Thailand.

(ATPase associated with diverse cellular activities) (11). A likely *P. falciparum* homolog of this protein contains an N-terminal signal anchor sequence, whereas the *T. parva* protein contains a signal peptide and lacks a recognizable endoplasmic reticulum retention signal.

We used the Tribe-MCL algorithm (5) to identify 384 protein families containing 1063 proteins in the *T. parva* proteome (table S1). The largest family, containing 85 proteins, exists primarily in tandem arrays in the subtelomeric regions of all chromosomes. Many members of the family have a similar architecture, consisting of a secretion signal at the N terminus and a low-complexity glutamine- and proline-rich central domain that may be difficult for vertebrate immune systems to recognize (12). These genes are polymorphic between parasite isolates, and specific genes are absent from certain isolates (13). Each telomere has a conserved ~140-bp sequence immediately adjacent to the telomeric repeat (14), and several subtelomeric regions exhibit 70 to 100% sequence similarity (fig. S1). As in other eukaryotic pathogens, these features may facilitate interchromosomal recombination and the generation of antigenic diversity.

Proteins in the most rapidly evolving *T. parva* protein family, the Tpr (*T. parva* repeat) family, contain complex domain structures reminiscent of a system that has evolved to generate diversity (15). Unlike the majority of hypervariable gene families in parasitic protozoa (16), Tpr sequences are not telomere-associated. This family comprises a tandem array of highly conserved open reading frames (ORFs) on chromosome 3, located ~570 kbp from a telomere. The locus, estimated to span 100 kbp, contains at least 28 ORFs, of which 18, ranging in length from 192 to 674 amino acids, lack methionine codons in the first 50 amino acids (fig. S2). Eleven additional dispersed copies of Tpr, also of varying length, contain a 268-amino acid membrane-associated helical domain typical of the Tpr family. Massively parallel signature sequencing (17) and expressed sequence tags suggest that some genes in the locus are only transcribed in the piroplasm stage, whereas at least two of the dispersed genes are transcribed in the schizont stage. In common with the var genes of *P. falciparum* (18), domains within the Tpr genes are isolate-specific (19), and the 3' end of Tpr has been used for genotyping of *T. parva* isolates. Tpr proteins have not yet been detected in piroplasms, and the function of these proteins remains unknown.

The genome sequence provides a global view of the metabolic potential of *T. parva* and allows a comparative analysis with *P. falciparum* metabolism. We predict a reduced functional role for the *T. parva* apicoplast and a greater dependence on the host for many

substrates (fig. S3). *T. parva* lacks many enzymes in the shikimic acid, porphyrin, polyamine, and type II fatty acid biosynthetic pathways, but it retains the ability to produce isoprenoids via a methyl erythritol phosphate pathway in the apicoplast. *T. parva* cannot salvage purines, its ability to interconvert amino acids is very limited, and it lacks enzymes that permit the alternative nonoxidative production of pentoses and tetroses via the pentose phosphate pathway. Analysis of predicted transporters revealed fewer transporters of organic nutrients and inorganic cations than are present in *P. falciparum*. However, *T. parva* has more adenosine 5'-triphosphate-binding cassette (ABC) transporters of unknown substrate specificity. Another difference is that *T. parva* encodes an amino acid-cation symporter that is not present in *P. falciparum* (7) or *C. parvum* (20). In contrast to *P. falciparum*, *T. parva* encodes trehalose-6-phosphate synthase and trehalose phosphatase. Trehalose is a disaccharide that plays a role in desiccation and stress tolerance. It may protect the parasite during its long developmental cycle in the tick.

T. parva genes encode all of the enzymes necessary for glycolysis, glycerol catabolism, and the tricarboxylic acid (TCA) cycle. Unlike *P. falciparum*, *T. parva* does not encode malate dehydrogenase, but this could be functionally replaced by malate-quinone oxidoreductase, an activity also predicted to be present in *P. falciparum*. The origin of mitochondrial acetyl-coenzyme A (CoA) in both parasites presents a problem, because *P. falciparum* encodes a single pyruvate dehydrogenase that is targeted to the apicoplast (21) and *T. parva* does not encode all the subunits of this enzyme. Both parasites are predicted to contain cytoplasmic acetyl-CoA synthetase and a plasma membrane acetyl-CoA-CoA antiporter, but how mitochondrial oxidation of carbon chains is fueled in these two pathogens

remains enigmatic because glycolysis and the tricarboxylic acid cycle do not appear to be linked by a classical route (22). Thus, it is not clear whether the complete TCA cycle is functional. Nitrogen metabolism differs from *P. falciparum* because *T. parva* lacks glutamate-ammonia ligase and only contains a nicotinamide adenine dinucleotide (NAD⁺)-dependent glutamate dehydrogenase, which is usually associated with glutamate catabolism. This suggests that imported glutamate could play a role in supplementing intermediates in the TCA cycle.

The ionophores valinomycin and gramicidin D kill *T. parva*, suggesting that a mitochondrial electrochemical gradient is essential for parasite survival (23), but it is not known whether this is coupled to ATP synthesis. All subunits of the F1 catalytic domain of ATP synthase and subunit c of the F0 domain are present, but genes coding for subunits a and b of F0 were not found. The *T. parva* respiratory complexes are similar to those described in *P. falciparum*. Buparvaquone, a hydroxynaphthaquinone drug used in the chemotherapy of ECF, probably inhibits electron transport through complex III (23).

The apicoplast is found in most apicomplexans and plays an essential role in parasite metabolism (24). An A+T-rich, ~35-kbp apicoplast genome encoding 30 proteins, rRNAs, and tRNAs is present in *Plasmodium*, *Toxoplasma*, and *Eimeria*, but not in *Cryptosporidium* (20); the latter lacks an apicoplast. The 39.5-kbp *T. parva* apicoplast genome differs from that of *P. falciparum* in that all of its genes are transcribed in the same direction. In addition, it has one rather than two copies of the rRNA genes, *clpC* is duplicated, the *rpoC2* gene encoding the β'' subunit of RNA polymerase is split into two parts, and it lacks the *sufB* gene (Fig. 1). Twenty-six of the 44 *T. parva* apicoplast genome protein-coding genes share sequence

Table 1. Comparison of *T. parva* nuclear genome coding characteristics with other sequenced apicomplexans. Gene length excludes introns; gene density calculated as genome size/number of protein-encoding genes. Source of data for *P. falciparum* was (7), and, for *C. parvum*, (20).

Features	Apicomplexan organism		
	<i>T. parva</i>	<i>P. falciparum</i>	<i>C. parvum</i>
Size (bp)	8,308,027	22,853,764	9,100,000
Number of chromosomes	4	14	8
Total G+C content (%)	34.1	19.4	30
Number of protein encoding genes	4035	5268	3807
Number of hypothetical proteins	2498	3208	925
Mean gene length (bp)	1407	2283	1795
Gene density (gene frequency in bp)	2057	4338	2382
Percent coding	68.4	52.6	75.3
Genes with introns (%)	73.6	53.9	5
Exons per gene (median)	4	2	1
Mean intergenic length (bp)	405	1694	566
G+C content intergenic regions (%)	26.1	13.6	23.9
Number of tRNA genes	47	43	45
Number of 5S rRNA genes	3	3	6
Number of rRNA units	2	7	5

similarity (27 to 61%) with proteins encoded by the *P. falciparum* apicoplast genome.

Most apicoplast proteins are encoded by nuclear genes and imported into the organelle by means of a bipartite targeting presequence (24). Comparison of the 345 *T. parva* (5) and 551 *P. falciparum* (7) predicted apicoplast-targeted (AT) proteins revealed similarities and differences in apicoplast function. The apicoplasts of *Plasmodium* and *Toxoplasma* participate in heme biosynthesis and are the sites of type II fatty acid and isoprenoid biosynthesis. Apicoplast-derived fatty acids in these parasites might contribute to the establishment and modification of the parasitophorous vacuole membrane (25). It may be notable that both *T. parva* and *T. annulata*, which have only retained isoprenoid biosynthesis, do not exist within a parasitophorous vacuole. About 100 AT proteins were found in both species, but 40% of these were hypothetical proteins, indicating that many core apicoplast functions have yet to be defined.

Fe-S clusters are required in mitochondria and plastids for the maturation of apoproteins. Fe-S cluster formation in the *T. parva* mitochondrion appears to be similar to that in yeast and *Plasmodium* (26) (table S3). However, of the *sufABCDE* genes involved in the assembly of Fe-S clusters in *Arabidopsis thaliana* (27) and *P. falciparum* plastids (26), only *sufS* was identified in *T. parva*. SufS is a cysteine desulfurase that requires SufE for catalytic activity. The parasite *T. parva* genome encodes a plastid-targeted tRNA thiolation enzyme (MnmA) that has an additional domain similar both in sequence and predicted structure to the sulfur-binding domain of SufE. Thus, a previously unknown complex of SufS/MnmA may catalyze thiolation of tRNA in the *T. parva* apicoplast. The *T. parva* nuclear genome also encodes an AT protein with homology to NFU1, a scaffold protein for Fe-S cluster assembly in *A. thaliana* plastids (28), suggesting that assembly of Fe-S clusters occurs in the *T. parva* apicoplast despite the absence of most Suf proteins.

T. parva and *T. annulata* exhibit near-complete synteny across all chromosomes (4). To examine the extent of conservation of gene synteny between the evolutionarily distant *P. falciparum* and *T. parva*, we applied an iterative syntenic block algorithm and Jaccard-filtered COGs to whole-genome data from *P. falciparum* clone 3D7 (7), *P. y. yoelii* (29), *C. parvum* (20), and *T. parva*. Extensive synteny was found between *P. falciparum* and *P. y. yoelii* but not between *P. falciparum* and *C. parvum* or between *T. parva* and *C. parvum*. A total of 435 microsyntenic regions containing 1279 orthologs were observed between *P. falciparum* and *T. parva*, consisting of groups of 2 to 11 orthologs conserved in position between the two genomes (Fig. 2). This may be an underestimate of the degree of microsynteny as it is possible that, due to its long-term in vitro culture, clone 3D7 may represent an atypical genome. Syntenic clusters were distributed uniformly along each chromosome except for the subtelomeric regions, which contain species-specific gene families.

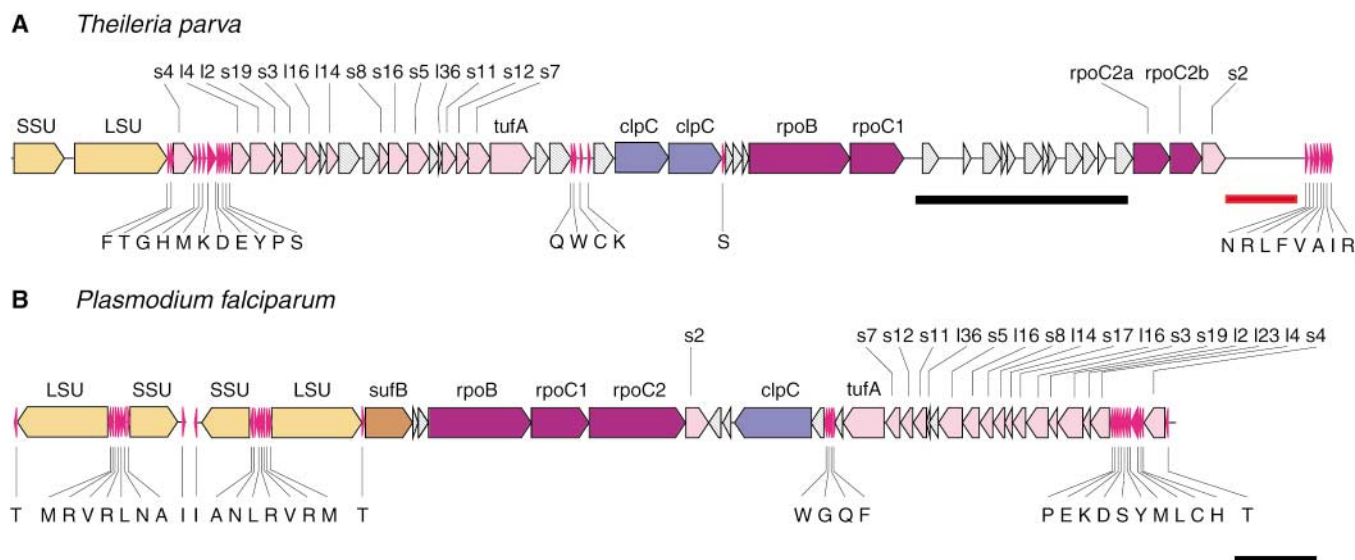


Fig. 1. Comparison of the apicoplast genomes of *T. parva* (A) and *P. falciparum* (B). A circular contig of the *T. parva* apicoplast genome was obtained after assembly of shotgun sequences, but the in vivo conformation has not been determined. The *P. falciparum* apicoplast genome is circular in vivo (30). The genomes are displayed in linear format beginning with the small subunit rRNA genes. Abbreviations and color coding: light orange, small (SSU) and large (LSU) subunit rRNAs; magenta, tRNAs [single-

letter amino acid code (31)]; pink, ribosomal proteins (s and l for small and large subunit ribosomal proteins, respectively) and elongation factor Tu (*tufA*); blue, protein import; stippled gray, hypothetical proteins; purple, transcription; brown, SufB subunit of the SufABCDE Fe-S cluster assembly complex. The black and red bars indicate a region containing repeats and short ORFs and another region containing repeats and potential selenocysteine tRNAs, respectively (5). Scale bar equals 1 kbp.

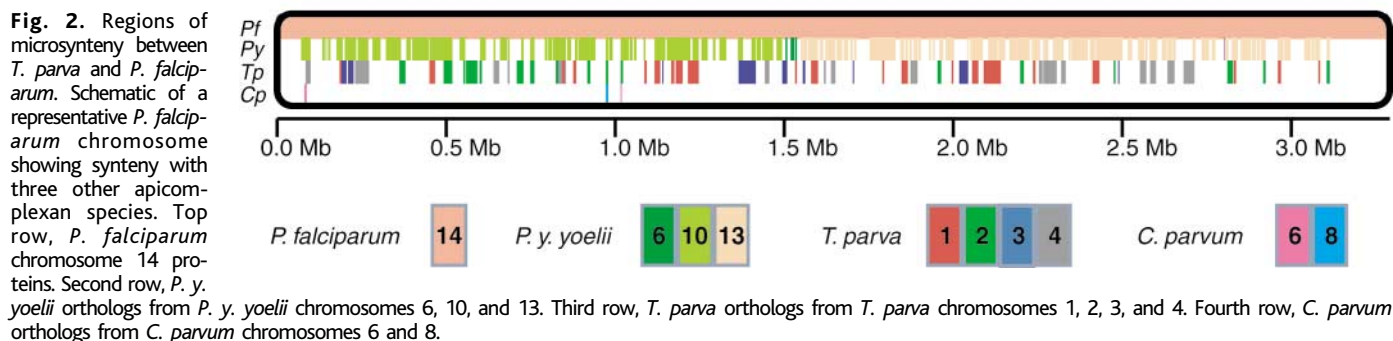


Fig. 2. Regions of microsynteny between *T. parva* and *P. falciparum*. Schematic of a representative *P. falciparum* chromosome showing syntenicity with three other apicomplexan species. Top row, *P. falciparum* chromosome 14 proteins. Second row, *P. y. yoelii* orthologs from *P. y. yoelii* chromosomes 6, 10, and 13. Third row, *T. parva* orthologs from *T. parva* chromosomes 1, 2, 3, and 4. Fourth row, *C. parvum* orthologs from *C. parvum* chromosomes 6 and 8.

The genome sequence of *T. parva* shows remarkable differences from the other apicomplexan genomes sequenced to date. It provides significant improvements in our understanding of the metabolic capabilities of *T. parva* and a foundation for studying parasite-induced host cell transformation and constitutes a critical knowledge base for a pathogen of significance to agriculture in Africa. Mining of sequence data has already proved useful in the search for candidate vaccine antigens (3).

References and Notes

- R. A. I. Norval, B. D. Perry, A. S. Young, *The Epidemiology of Theileriosis in Africa* (Academic Press, London, 1992), p. 481.
- A. A. Escalante, F. J. Ayala, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5793 (1995).
- S. P. Graham *et al.*, in preparation.
- A. Pain *et al.*, *Science* **309**, 131 (2005).
- Materials and methods are available as supporting material on Science Online.
- A. Kairo, A. H. Fairlamb, E. Gobright, V. Nene, *EMBO J.* **13**, 898 (1994).
- M. J. Gardner *et al.*, *Nature* **419**, 498 (2002).
- J. H. Gunderson *et al.*, *Science* **238**, 933 (1987).
- B. Vogelstein, K. W. Kinzler, *Nat. Med.* **10**, 789 (2004).
- M. Goedert, *Semin. Cell Dev. Biol.* **15**, 45 (2004).
- I. M. Cheeseman, A. Desai, *Curr. Biol.* **14**, R70 (2004).
- R. F. Anders, *Parasite Immunol.* **8**, 529 (1986).
- R. Bishop *et al.*, *Mol. Biochem. Parasitol.* **110**, 359 (2000).
- B. Sohanpal, D. Wasawo, R. Bishop, *Gene* **255**, 401 (2000).
- H. A. Baylis, S. K. Sohal, M. Carrington, R. P. Bishop, B. A. Allsopp, *Mol. Biochem. Parasitol.* **49**, 133 (1991).
- J. D. Barry, M. L. Ginger, P. Burton, R. McCulloch, *Int. J. Parasitol.* **33**, 29 (2003).
- R. Bishop *et al.*, in preparation.
- Z. Su *et al.*, *Cell* **82**, 89 (1995).
- R. Bishop, A. Musoke, S. Morzaria, B. Sohanpal, E. Gobright, *Mol. Cell. Biol.* **17**, 1666 (1997).
- M. S. Abrahamsen *et al.*, *Science* **304**, 441 (2004); published online 25 March 2004 (10.1126/science.1094786).
- B. J. Foth *et al.*, *Mol. Microbiol.* **55**, 39 (2005).
- S. A. Ralph, *Mol. Microbiol.* **55**, 1 (2005).
- A. A. McColm, N. McHardy, *Ann. Trop. Med. Parasitol.* **78**, 345 (1984).
- R. F. Waller, G. I. McFadden, *Curr. Issues Mol. Biol.* **7**, 57 (2005).
- R. F. Waller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12352 (1998).
- F. Seeber, *Int. J. Parasitol.* **32**, 1207 (2002).
- X. M. Xu, S. G. Moller, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9143 (2004).

- S. Leon, B. Touraine, C. Ribot, J. F. Briat, S. Lobreaux, *Biochem. J.* **371**, 823 (2003).
- J. M. Carlton *et al.*, *Nature* **419**, 512 (2002).
- R. J. Wilson *et al.*, *J. Mol. Biol.* **261**, 155 (1996).
- Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
- We thank T. Irvin, O. Ole-MoiYoi, T. Musoke, C. Sugimoto, H. Leitch, R. von Kaufmann, S. MacMillan, R. Koenig, M. Brown, R. Ndegwa, L. Thairo, B. Anyona, T. Akinyemi, the TIGR conferences staff, the Secretariat of the Consultative Group for International Agricultural Research, and the research staff of the International Livestock Research Institute (ILRI). Supported by the TIGR Board of Trustees, ILRI, J. C. Venter, the Rockefeller Foundation, the U.S. Agency for International Development, and the UK Department for International Development.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5731/134/DC1

Materials and Methods

Figs. S1 to S3

Tables S1 to S3

31 January 2005; accepted 5 May 2005

10.1126/science.1110439

Long-Term Monitoring of Bacteria Undergoing Programmed Population Control in a Microchemostat

Frederick K. Balagaddé,^{1*} Lingchong You,^{2,†‡} Carl L. Hansen,^{1,§} Frances H. Arnold,² Stephen R. Quake^{1*||}

Using an active approach to preventing biofilm formation, we implemented a microfluidic bioreactor that enables long-term culture and monitoring of extremely small populations of bacteria with single-cell resolution. We used this device to observe the dynamics of *Escherichia coli* carrying a synthetic "population control" circuit that regulates cell density through a feedback mechanism based on quorum sensing. The microfluidic bioreactor enabled long-term monitoring of unnatural behavior programmed by the synthetic circuit, which included sustained oscillations in cell density and associated morphological changes, over hundreds of hours.

By continually substituting a fraction of a bacterial culture with sterile nutrients, the chemostat (1, 2) presents a near-constant environment that is ideal for controlled studies of microbes and microbial communities (3–6). The considerable challenges of maintaining and operating continuous bioreactors, includ-

ing the requirement for large quantities of growth media and reagents, have pushed the move toward miniaturization and chip-based control (7–10), although efforts have been limited to batch-format operation. Microbial biofilms, which exist in virtually all nutrient-sufficient ecosystems (11), interfere with continuous bioreactor operation (12). Phenotypically distinct from their planktonic counterparts (11), biofilm cells shed their progeny into the bulk culture and create mixed cultures. At high dilution rates, the biofilm, which is not subject to wash-out, supplies most of the bulk-culture cells (13). The increase in surface area-to-volume ratio as the working volume is decreased aggravates these wall-growth effects (13).

We created a chip-based bioreactor that uses microfluidic plumbing networks to actively prevent biofilm formation. This device allows semicontinuous, planktonic growth in six independent 16-nanoliter reactors with no

observable wall growth (Fig. 1A). The cultures can be monitored in situ by optical microscopy to provide automated, real-time, noninvasive measurement of cell density and morphology with single-cell resolution.

Each reactor, or "microchemostat," consists of a growth chamber, which is a fluidic loop 10 μm high, 140 μm wide, and 11.5 mm in circumference, with an integrated peristaltic pump and a series of micromechanical valves to add medium, remove waste, and recover cells (Fig. 1B). The growth loop is itself composed of 16 individually addressable segments. The microchemostat is operated in one of two alternating states: (i) continuous circulation, and (ii) cleaning and dilution. During continuous circulation, the peristaltic pump moves the microculture around the growth loop at a linear velocity of $\sim 250 \mu\text{m s}^{-1}$ (Fig. 1C). During cleaning and dilution, the mixing is halted and a segment is isolated from the rest of the reactor with micromechanical valves. A lysis buffer is flushed through the isolated segment for 50 s to expel the cells it contains, including any wall-adhering cells (Fig. 1D). Next, the segment is flushed with sterile growth medium to completely rinse out the lysis buffer. This segment, filled with sterile medium, is then reunited with the rest of the growth chamber, at which point continuous circulation resumes. This process is repeated sequentially on different growth chamber segments, thus eliminating biofilm formation and enabling pseudocontinuous operation. In comparison, passive treatment of the microfluidic surfaces with nonadhesive surface coatings [such as poly (ethylene glycol), ethylenediaminetetraacetic acid, polyoxyethylene sorbitan monolaurate, and bovine serum albumin] proved ineffective in preventing biofilm forma-

¹Department of Applied Physics, ²Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

*Present address: Department of Bioengineering, Stanford University, Stanford, CA 94305, USA.

†These authors contributed equally to this work.

‡Present address: Department of Biomedical Engineering and Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708, USA.

§Present address: Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

||To whom correspondence should be addressed. E-mail: quake@stanford.edu