Comparative Genome Analysis of Three Eukaryotic Parasites with Differing Abilities To Transform Leukocytes Reveals Key Mediators of Theileria-Induced Leukocyte Transformation

Kyoko Hayashida,¹ Yuichiro Hara,² Takashi Abe,³ Chisato Yamasaki,³ Atsushi Toyoda,⁴ Takehide Kosuge,⁵ Yutaka Suzuki,⁴ Yoshiharu Sato,⁵ Shuichi Kawashima,⁶ Toshiaki Katayama,⁷ Hiroyuki Wakaguri,⁸ Noboru Inoue,⁹ Keiichi Homma,⁶ Masahito Tada-Umezaki,¹ Yukio Yagi,¹ Yasuyuki Fujii,¹ Takuya Habara,¹ Minoru Kaniehisa,²⁶ Hidemi Watanabe,²⁶ Kimihito Ito,²⁶ Takashi Gojobori,¹,² Hideaki Sugawara,¹ Tadashi Imanishi,³ William Weir,² Malcolm Gardner,⁴ Arnab Pain,² Brian Shiels,³ Masahira Hattori,¹ Vishvanath Nene,³ and Chihiro Sugimoto⁴

Division of Collaboration and Education, Research Center for Zoonosis Control, Hokkaido University, Sapporo, Hokkaido, Japan; Biomedical Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan; Information Engineering, Niigata University, Niigata, Japan; Comparative Genomics Laboratory, Center for Genetic Resource Information, National Institute of Genetics, Research Organization of Information and Systems, Mishima, Shizuoka, Japan; Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Mishima, Shizuoka, Japan; Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan; Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan; Human Genome Institute, Institute of Medical Science, University of Tokyo, Tokyo, Japan; National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido, Japan; Institute of Natural Medicine, University of Toyama, Toyama, Japan; Hokkaido Research Station, National Institute of Animal Health, National Agricultural Research Organization, Sapporo, Hokkaido, Japan; Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan; Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan; Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan; Division of Bioinformatics, Research Center for Zoonosis Control, Hokkaido University, Sapporo, Hokkaido, Japan; Institute of Comparative Medicine, Medicine University Veterinary School, Glasgow, United Kingdom; Seattle Biomedical Research Institute, Seattle, Washington, USA; Pathogen Genomics, Computational Bioscience Research Center, Chemical Life Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; and International Livestock Research Institute, Nairobi, Kenya

ABSTRACT We sequenced the genome of Theileria orientalis, a tick-borne apicomplexan protozoan parasite of cattle. The focus of this study was a comparative genome analysis of T. orientalis relative to other highly pathogenic Theileria species, T. parva and T. annulata. T. parva and T. annulata induce transformation of infected cells of lymphocyte or macrophage/monocyte lineages; in contrast, T. orientalis does not induce uncontrolled proliferation of infected leukocytes and multiplies predominantly within infected erythrocytes. While synteny across homologous chromosomes of the three Theileria species was found to be well conserved overall, subtelomeric structures were found to differ substantially, as T. orientalis lacks the large tandemly arrayed subtelomere-encoded variable protein-encoding gene family. Moreover, expansion of particular gene families by gene duplication was found in the genomes of the two transforming Theileria species, most notably, the TashAT/TpHN and Tar/Tpr gene families. Gene families that are present only in T. parva and T. annulata and not in T. orientalis, Babesia bovis, or Plasmodium were also identified. Identification of differences between the genome sequences of Theileria species with different abilities to transform and immortalize bovine leukocytes will provide insight into proteins and mechanisms that have evolved to induce and regulate this process. The T. orientalis genome database is available at http://totdb.czc.hokudai.ac.jp/.

IMPORTANCE Cancer-like growth of leukocytes infected with malignant Theileria parasites is a unique cellular event, as it involves the transformation and immortalization of one eukaryotic cell by another. In this study, we sequenced the whole genome of a nontransforming Theileria species, Theileria orientalis, and compared it to the published sequences representative of two malignant, transforming species, T. parva and T. annulata. The genome-wide comparison of these parasite species highlights significant genetic diversity that may be associated with evolution of the mechanism(s) deployed by an intracellular eukaryotic parasite to transform its host cell.
Theileria spp. are tick-borne intracellular parasites that belong to the phylum Apicomplexa and infect domestic and wild ruminants, including cattle, Asian water buffalos, sheep, goats, and African buffalos. Although infection by some Theileria species is asymptomatic or persists as a chronic infection, Theileria parva and Theileria annulata can be highly pathogenic to cattle and Theileria lestoquardi can cause significant disease in sheep. These three species are among the “transforming Theileria” species because of their ability to transform and induce indefinite proliferation of infected host leukocytes (1–4). The resulting disease syndromes can be described as lymphoproliferative disorders, which often culminate in disorganization and destruction of the host lymphoid system. Although detailed information has been generated for a number of host cell signal transduction pathways that are perturbed during leukocyte transformation, parasite molecules responsible for the initiation or regulation of the host cell transformation event have yet to be identified or fully validated (5, 6).

A comparative analysis of the T. parva and T. annulata genome sequences was reported in 2005 (7, 8). Despite the identification of a number of Theileria genes that could be involved in the transformation process, the selectivity of the approach was compromised by a high number of hypothetical proteins of unknown function and the high number of shared genes that exists across the genomes of these two closely related species. One way in which the discriminatory power of a comparative genomic approach could be increased would be to conduct bi- and trilateral genome comparisons with Theileria and Babesia parasites that lack the ability to transform host leukocytes but otherwise show strong similarity over the rest of their parasitic life cycle (9).

Theileria orientalis, an intraerythrocytic parasite of cattle, is a member of the nontransforming group of Theileria species that proliferate in the bovine host as an intraerythrocytic form and can generate anemia and icterus but rarely cause fatal disease (10). This parasite has frequently been referred to as T. sergenti, but this specific name is now considered invalid (11). Bovine piroplasmosis caused by this species causes enormous economic losses in the livestock industry in Japan (12–14). T. orientalis is often classified into two major genotypes, the Chitose type and the Ikeda type, which are distinguishable on the basis of diversity in the small-subunit rRNA and major piroplasm surface protein (MPSP) gene sequences (15). The T. orientalis Ikeda type is limited to eastern Asian countries, including Japan, South Korea, the northeastern part of China, and Australia (16), and it is present in areas where livestock succumb to severe clinical cases of theileriosis and serious production losses. In contrast, T. orientalis Chitose is found throughout the world and is usually associated with benign infection (15, 17). Thus, even though it is believed to be relatively mild compared to the transforming Theileria species, T. orientalis can be an important pathogen in its own right and many researchers have been looking forward to the derivation of the genomic sequence to provide an important resource for further studies.

Unlike transforming Theileria species, the macroschizonts of nontransforming Theileria parasites are only transiently found in cells within lymph nodes or the spleen following the invasion of host cells by the infective sporozoite, and no evidence for proliferation of infected cells has been reported in vivo or in vitro. Indeed, in vivo studies indicate that the schizont undergoes continual enlargement over the course of 4 to 8 days before generating multiple merozoites that are released upon host cell destruction. A lack of host proliferation is indicated by a substantial increase in host cell size, but it is unknown whether the parasite manipulates the cell at the molecular level or inhibits an apoptotic response to infection (13, 18). Free merozoites subsequently invade erythrocytes and, unlike the case with transforming species, undergo significant rounds of proliferation in red blood cells, similar to the proliferation observed with Babesia parasites. Clinical signs, when observed, are associated primarily with anemia and icterus. In addition to the schizont stage, the intraerythrocytic stage of T. annulata can also cause anemia.

In this study, we focused primarily on a comparative analysis of the genome of the T. orientalis Ikeda type relative to the genomes of the transforming Theileria species T. parva and T. annulata and a closely related hemoparasite species, Babesia bovis. The main goals of this analysis were to provide supportive data on existing candidate genes and/or identify novel candidate genes that enable the transformation of bovine leukocytes upon infection with T. annulata and T. parva.

RESULTS AND DISCUSSION

Structure of the T. orientalis genome. Whole-genome shotgun sequence data on T. orientalis (Ikeda strain) were assembled, and physical gaps between scaffolds were manually closed, resulting in the complete sequence of all four chromosomes. The derived sequence has been deposited in the DNA Data Bank of Japan (DDBJ) under project accession numbers AP011946 to AP011951. In addition to the nuclear genome, partial sequences of the apicoplast and mitochondrial genomes were also obtained. The complete genome sequence of the mitochondria has already been published (accession number AB499090) (19).

At 9.0 Mb, the genome size of T. orientalis is approximately 8% larger than the reported genome sizes of T. parva, T. annulata, and B. bovis. The number of predicted protein-coding genes identified in T. orientalis is, however, almost the same as that found in T. parva (Table 1). The G+C composition of the T. orientalis genome (41.6%) is higher than those of T. parva and T. annulata (34.1% and 32.5%, respectively) but similar to that of B. bovis (41.8%). The frequencies of the top 50 InterPro entries (see Table S1 in the supplemental material) are similar for the three Theileria species, suggesting that, in general, the three parasite species possess similar sets of gene families and encoded protein domains. For example, the InterPro domain of DUF529, known as the FAINT (frequently associated in Theileria) domain, described later in detail, is found frequently in all of the Theileria species sequenced to date. In contrast, the PEST motif, associated with rapid degradation of (nuclear) proteins, was found to be encoded by several gene families in the genomes of the two transforming Theileria species but was not identified in T. orientalis.

Synteny across all of the chromosomes of all three Theileria species is generally conserved, except for the subtelomeric regions, and several internal inversions were identified for each chromosome (Fig. 1). Most large-scale inversions were found when comparing T. orientalis versus T. annulata (Fig. 1, lower half of each panel) and were not present in the T. annulata-versus-T. parva comparison (Fig. 1, upper half of each panel), suggesting that these structural changes occurred following the speciation of T. orientalis and a common ancestor of T. annulata/T. parva. However, a large inversion of approximately 113,000 bp in chromosome 3 of T. annulata may have occurred after the speciation of T. annulata and T. parva (Fig. 1, upper right panel, indicated by a
Theileria and Phylogenetic analysis has indicated that the two transforming metabolism, nucleotide metabolism, and amino acid degradation. This enzyme is known to thiolase that catalyzes the conversion of acetyl coenzyme A and transforming species had 255 KOs in common, indicating no significant difference in the proliferating capacities due to an increasing host cell metabolite-scavenging ability. Four species could offer a valuable insight into how these parasites have evolved and adapted to their different host environments, including the acquisition of leukocyte transformation capability. To examine the expansion processes of gene families in the Theileria lineages in detail, we constructed gene families composed of sequences representing the three Theileria species, B. bovis, and two Plasmodium species (Plasmodium falciparum and P. vivax) on the basis of the ortholog clustering framework of OrthoMCL (23), as well as additional computational and manual curations. We assigned 3,419 orthologous groups in which at least one Theileria species was included. While 1,740 of these orthologous groups consisted of single-copy genes across all six species, 223 orthologous groups possessed Theileria paralogs (see Data set S1 in the supplemental material). We focused on several family groups in the Theileria lineage that showed evidence of marked expansion that could be associated with acquisition of the ability to generate the proliferating, transformed, infected leukocyte.

### Expansion of gene families in the genomes of transforming Theileria species

Three gene families showed a striking association with the genomes of the two transforming Theileria species. PiroF0100022 (Tar/Tpr family), PiroF0100037 (SVSP family), and PiroF0100038 (TashAT/TpHN family) are all significantly expanded within or unique to the genomes of the host cell-transforming Theileria lineage and are composed of genes predicted to encode proteins possessing FAIN domains. The TashAT family of T. annulata contains 17 tandemly arrayed genes, some of which have been shown to encode proteins that are translocated to the host nucleus, bind DNA, and alter gene expression and protein profiles of transfected bovine cells (24, 25). An orthologous cluster of 20 genes (TpHN) has also been identified in T. parva (25). In sharp contrast, only a single TashAT/TpHN-like gene, TOT0100571, was identified in the genome of T. orientalis. Reciprocal best hits using BLASTP indicate that the T. orientalis gene is likely to be the ortholog of Tash-a (TA03110) and TP01_0621 in the transforming Theileria species. Both of these genes are located at the 3’ ends of their respective clusters in the T. annulata and T. parva genomes (Fig. 3A).

To gain further insight into the species-conserved Tash-a gene relative to the other members of the TashAT cluster, we obtained microarray data to examine whether gene expression of the different TashAT genes is associated with proliferating, macroschizont-infected leukocytes (26). Analysis of the normalized dataset showed that, in general, TashAT family expression is consistently double-headed arrow). A striking difference between the genomes is that a number of gene families show evidence of expansion and diversification specific to the genomes of the transforming Theileria species, while few instances of T. orientalis-specific gene family expansion were recorded. In addition, several lineage-specific genes were identified at microsynteny breakpoints. Finally, the subtelomeric regions of all four T. orientalis chromosomes are markedly different from those of T. annulata and T. parva because they completely lack the largest subtelomeric gene family reported for T. annulata and T. parva, which encodes subtelomere-encoded variable secreted proteins (SVSPs) (see Fig. 2) of unknown function (20).

### Metabolic pathways

To reconstruct KEGG metabolic pathways of T. orientalis, we assigned 263, 263, 273, and 264 KEGG orthology (KO) identifiers (21) to the predicted proteomes of T. orientalis, T. parva, T. annulata, T. orientalis, and B. bovis, respectively (see Fig. S1 in the supplemental material). These four species had 255 KOs in common, indicating no significant differences in known metabolic pathways between nontransforming and transforming Theileria and Babesia species, despite the known preference to proliferate in different host cell types (leukocytes versus erythrocytes).

K00626 is the only KO common to B. bovis and T. orientalis and not identified in T. parva and T. annulata. It codes for a putative thiolase that catalyzes the conversion of acetyl CoA (acetyl-CoA) into acetoacetyl-CoA. This enzyme is known to function in a variety of metabolic pathways, including fatty acid metabolism, nucleotide metabolism, and amino acid degradation. Phylogenetic analysis has indicated that the two transforming Theileria species diverged from T. orientalis after the speciation of Theileria and Babesia (22). Therefore, the acetoacetyl-CoA thiolase might represent an example of a reduction of metabolic capacity due to an increasing host cell metabolite-scavenging ability/dependence of Theileria species.

### Gene families

Expansion of gene families specific to different Theileria species could offer a valuable insight into how these parasites have evolved and adapted to their different host environments, including the acquisition of leukocyte transformation capability. To examine the expansion processes of gene families in the transforming species, we took the orthology (KO) identifiers (21) to the predicted proteomes of the transforming species. Theileria annulata TashAT family expression is consistently associated with proliferating, transformed, infected leukocytes.

| Nuclear genome feature | T. orientalis | T. annulata | T. parva | B. bovis |
|------------------------|--------------|-------------|----------|----------|
| Size (Mb)              | 9.0          | 8.4         | 8.3      | 8.2      |
| No. of chromosomes     | 4            | 4           | 4        | 4        |
| Total G+C content (%)  | 41.6         | 32.5        | 34.1     | 41.8     |
| No. of protein-coding genes | 3,995      | 3,792        | 4,035     | 3,641     |
| % of genes with introns | 78.3         | 70.6        | 73.6     | 61.5     |
| Mean gene length (bp)  | 1,861        | 1,606       | 1,407    | 1,514    |
| % Coding               | 68.6         | 72.8        | 68.4     | 70.2     |
| Mean intergenic length (bp) | 390         | 396         | 402      | 589      |
| % G+C composition of exons | 44.5        | 37.6        | 35.9     | 44.0     |
| % G+C composition of intergenic regions | 35.2 | 22.5 | 24.9 | 37.0 |
| % G+C composition of introns | 38.1       | 22.2        | 23.6     | 35.9     |
| No. of tRNA genes      | 47           | 47          | 47       | 44       |
| No. of 5S rRNA genes   | 3            | 3           | 3        | NA       |
| No. of 5.8S, 18S, and 28S rRNA units | 2       | 2           | 2        | 3        |
| Mitochondrial genome size (kb) | 2.5    | 6           | 6        | 6        |
| Apicoplast genome size (kb) | 26.5 | NA          | 39.5     | 33       |
| Gene density<sup>a</sup> | 2,249        | 2,202       | 2,059    | 2,228    |

<sup>a</sup> Genome size/number of protein-coding genes.

<sup>b</sup> NA, not available.
downregulated as the macroschizont undergoes differentiation to the merozoite and host cell proliferation subsides, as demonstrated previously for a number of individual family members (25). In marked contrast, transcripts representing Tash-a were found to be significantly upregulated during the differentiation process (see Fig. S2A in the supplemental material). This result may indicate a requirement for synthesis of the protein during merozoite production. This postulation was supported by an indirect fluorescent-antibody test (IFAT) using serum raised against a Tash-a fusion protein (see Fig. S2B) and colocalization of Tash-a staining with a merozoite rhoptry antigen (see Fig. S2C). We conclude that the Tash-a protein performs a function that is required during or following merozoite production and that the temporal expression and location of the protein are distinct from those of other members of the family. Phylogenetic analysis suggests that Tash-a and its orthologs represent ancestral members of the TashAT and TpHN clusters (see Fig. 3B). In addition, we did not find any obvious TashAT orthologs in B. bovis or two Plasmodium species genomes. We propose that Tash-a diverged after the separation of Theileria from a common ancestor of Theileria and Babesia and that gene duplication and functional diversification of the TashAT and TpHN clusters has then occurred as Theileria species of the transforming lineage evolved. Whether expansion of the cluster coincided with acquisition of a transforming capability is unknown.

Polypeptides encoded by the subtelomeric SVSP gene family (PiroF010037) are a major component of the predicted macroschizont secretome of T. annulata and T. parva, and a number of SVSPs have been predicted to translocate to the nucleus of the infected cell. Most SVSP genes are coexpressed in cultures of macroschizont-infected cells, and the SVSP family shows a high level of amino acid sequence diversity (20). Further work is necessary to determine the functions of these proteins and their role in the transforming process.
needed to determine the function of SVSPs, whether they contribute directly to the transformation of the host cell or play a role in subverting the bovine immune response. Some of the SVSPs contain bioinformatically detectable signal peptides, suggesting secretion into the host cell cytoplasm. Though the expression patterns of *T. parva* SVSPs appear complicated and their involvement in phenotypic changes in host leukocytes remains unclear, the fact that some SVSPs encode functional nuclear localization signals (NLSs) in addition to a predicted signal sequence for secretion suggests that they might be transported to the host nucleus and modulate signaling pathways (20). In this context, the absence of SVSP loci in *T. orientalis* is noteworthy. Thus, like the TashAT/TpHN clusters, SVSP gene expansion in *T. annulata* appears to be associated with species of the transforming *Theileria* lineage and may provide an as-yet-unknown function that promotes the establishment or maintenance of proliferating macroschizont-infected leukocytes.

In addition to the SVSP and TashAT clusters, the Tar/Tpr (PiroF0100022) family of orthologous genes showed evidence of significant expansion in the transforming *Theileria* lineages, as only five genes dispersed over the four chromosomes were detected in *T. orientalis* is noteworthy. Thus, like the TashAT/ TpHN clusters, SVSP gene expansion in *T. annulata*/*T. parva* appears to be associated with species of the transforming *Theileria* lineage and may provide an as-yet-unknown function that promotes the establishment or maintenance of proliferating macroschizont-infected leukocytes.

Transcriptome studies indicate that copies of Tpr genes dispersed throughout the *T. parva* genome are expressed in the macroschizont stage (27), while those organized in a tandem array of 28 genes are expressed by the intraerythrocytic piroplasm (28).

The CD8 T cell response is considered to play a key role in immunity to *T. parva*/*T. annulata* (29). Of the macroschizont antigens that are recognized by CD8 T cells from immune animals (30, 31), one, TA9/TP9 (TA15705/TP02_0895), is encoded by a member of a small orthologous gene family (PiroF0100041) in the genomes of transforming *Theileria* species. The family consists of five and six members in *T. annulata* and *T. parva*, respectively, all of which encode predicted proteins with a signal peptide for secretion by the parasite. Expressed sequence tag (EST) data and microarray data indicate that one of the TA9 family members (TA15705) is expressed in a specific manner by the transforming macroschizont stage (see Fig. S3C in the supplemental material), and it has been reported that the protein can be detected in the host cell cytosol (32). In the *T. orientalis* genome, a single gene (TOT020000921) showing weak homology in the signal peptide region and C-terminal region with the TA9/TP9 family was found in a syntenic region of chromosome 2 (see Fig. S3). The data indicate that the TA9/TP9 gene family has expanded uniquely in the transforming *Theileria* species. A role for TA9-encoded polypeptides in the transformation of the host cell requires further investigation.
Evolution of the FAINT domain superfamily. As observed for *T. annulata* (8) and *T. parva* (7), a large number of genes whose predicted polypeptides encode DUF529 domains (IPR007480 in InterPro), alternatively called FAINT domains, were found in *T. orientalis* (see Table S1 in the supplemental material). Previous analysis revealed that ~900 copies of FAINT domains are present in the genomes of *T. annulata* and *T. parva* (8). With our pipelines for InterPro annotation, 686 FAINT copies were identified in 137 predicted *T. orientalis* proteins, and 913 and 725 copies were identified in 126 *T. annulata* and 142 *T. parva* putative proteins, respectively. This suggests that expansion of FAINT domain-containing polypeptides (FAINT superfamily) is likely to have occurred in the common ancestor of the three *Theileria* species. In addition, ortholog clustering indicated that different FAINT families have been expanded in *T. orientalis* than in *T. parva* and *T. annulata*. For example, the FAINT superfamilies of PiroF0001942 and PiromF0001943 are specifically expanded in *T. orientalis* (see Table S2 in the supplemental material). In contrast, the PiromF0100056 orthologous group of SfiI-related genes showed greater expansion in *T. parva* and *T. annulata* (see Table S2). A protein of the FAINT superfamily was also found in *T. equi* (8), which has been considered to be an outlier species in the genus *Theileria* (33). This indicates that FAINT domain polypeptides were present in early ancestral species of the *Theileria* genus and
have subsequently been subjected to differential expansion or contraction pressures as the different species evolved.

Many of the FAINT superfamily members in T. parva and T. annulata are inferred to be secretory proteins (5). Out of 137 proteins of the FAINT superfamily identified in T. orientalis, signal peptides were found in 103, indicating that members of the FAINT superfamily are significantly enriched for proteins with a predicted signal peptide \( (P = 5.97 \times 10^{-35}, \text{Fisher’s exact test}) \). Thus, the differential expansion and diversification of FAINT domain proteins could be associated with the adaptation of different Theileria species to preferential host niches that require specific host-parasite interactions. Comparison of additional genome sequences derived from both nontransforming and transforming Theileria species may be informative.

**Candidate genes responsible for Theileria-induced host cell transformation.** Comparative genomic analysis of T. orientalis and T. annulata/T. parva provides a tool for identifying candidate genes responsible for Theileria-mediated host cell transformation. This premise is based on the assumption that transformation-related genes are unique to the T. annulata/T. parva lineage, as there is no evidence that T. orientalis can transform leukocytes into proliferating infected cells. It can also be predicted that molecules that regulate the transformation event are likely to be secreted or localized to the macroschizont membrane, since Theileria parasites have direct contact with the host cell cytoplasm (34). In the course of ortholog classification analysis, we applied both of these criteria and identified 97 ortholog groups present in the T. parva and T. annulata lineages that were absent from T. orientalis, B. bovis, P. falciparum, and P. vivax. Of these lineage-specific ortholog groups, 29 are predicted to encode polypeptides with an endoplasmic reticulum signal sequence (several of which also contain a GPI anchor motif), indicating potential interaction with the host cell compartment (Table 2). The majority of these genes encode hypothetical proteins and do not show any similarities to known cancer-related genes, although several domains are predicted in the InterPro entries. We propose that genes placed within these 29 groups, plus the TashAT/TPHN family, can be considered candidates for involvement in the transformation process.

Identification of candidate genes as host cell phenotype manipulators has been reported previously (5, 35). The predicted proteins have signal sequences, protein kinase properties, phosphatase properties, NLSs, or DNA binding motifs, or they show identity with higher eukaryotic proteins that are involved in neoplasia. We searched for these genes in the genome of T. orientalis and found that all of them, with the exception of TashAT and SVSP family genes, are conserved across the three Theileria species (see Table S3 in the supplemental material). However, four T. orientalis genes lack the signal sequence or NLS that is predicted in each of the T. annulata/T. parva orthologs. Thus, it is possible that the function or localization of the encoded polypeptides has diverged between T. orientalis and the transforming Theileria species, and this may be worthy of further investigation.

**Conclusions.** This is the first genome sequence of a nontransforming Theileria species that occupies a phylogenetic position close to that of the transforming Theileria species and thus provides an ideal opportunity to analyze unique features of Theileria parasitism from an evolutionary viewpoint. Genome sequencing of the nontransforming Theileria species T. orientalis and comparison with the transforming Theileria species T. annulata and T. parva highlighted lineage-specific evolutionary features. Several transforming Theileria lineage-specific gene family expansions were identified, including the SVSP, Tash/TPHN, Tpr/Tar, and TP9/TAF families, that may have been coincident with development of the ability to transform host leukocytes. Additional genes identified as specific to the genomes of transforming Theileria species can also be considered transformation candidates. This study provides increased understanding of the evolution of transforming Theileria species at the genomic level and has generated a database that will serve as the foundation for future studies on Theileria pathobiology and parasite-host cell interaction.

**MATERIALS AND METHODS**

**Parasite samples.** T. orientalis (Shintoku stock) was used as the starting genomic material in this study. This stock contains two different genotypes, Ikeda and Chitose. Parasites of a single genotype (Ikeda) were selected following syringe passage of the original isolate through calves and then used to infect an animal for parasite isolation. Blood collected from the infected animal was passed through a leukocyte removal filter (Terumo), and the resulting red cells were washed three times with phosphate-buffered saline (PBS). Erythrocytes were resuspended in an equal volume of PBS and disrupted by nitrogen cavitation, and pigoplasms were purified by differential centrifugation as described previously (36). Infection of the cow was conducted in accordance with protocols approved by the National Institute of Animal Health, Japan, Animal Care and Use Committee (approval no. 2000/901). Genomic DNA was purified by proteinase K and SDS treatment, followed by phenol-chloroform extraction. Purified parasite DNA was dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Confirmation that the DNA represented the Ikeda genotype was carried out by PCR targeting genes encoding small-subunit rRNA and the MPSP as previously described (37).

**Genome sequencing.** The complete genome sequence of T. orientalis was determined by a combination of the whole-genome shotgun method and fosmid end sequencing. Genomic DNA was fragmented for plasmid library construction with an average insert size of 2 to 4 kb using a HydroShear DNA Shearing Device (Genemachines). Plasmid DNA was amplified with a TempliPhi DNA amplification kit (GE Healthcare) from the bacterial culture. The fosmid library was constructed by TaKaRa Bio Inc. using a CopyControl pCClFOS vector (Epigen, Madison, WI). Fosmid DNA was extracted with PI-1100 plasmid isolators (Kurabo). Both ends of 40,704 fosmid inserts and 3,840 fosmid clones were sequenced with ABI 3730 sequencers (Applied Biosystems) and MegaBACE 4500 sequencers (GE Healthcare). Contigs were assembled by using 111,945 shotgun reads. Gap closing and resequencing of low-quality regions in the assembled data were performed by shotgun sequencing of fosmid clones that covered the target regions, nested deletion (38), construction of short-insert libraries (39), and primer walking on selected clones and PCR-amplified DNA fragments. The overall accuracy of the finished genome sequence was estimated to have an error rate of less than 1 per 10,000 bases. The sequence is available from DDBJ/GenBank/EMBL under accession numbers AP011946 to AP011951.

**cDNA/ESTs.** Six volumes of Trizol LS was added to 1 volume of parasite-infected erythrocytes and homogenized with a Polytron homogenizer. Total RNA was then isolated according to the manufacturer’s protocol, and full-length cDNA libraries were produced by either the oligo-capping or the vector-capping method (40). Random clones were picked from the oligo capped and vector-capped library, and inserts were amplified by PCR from the single colonies sequenced at the 5’ end or both the 5’ end and the 3’ end. Sequences were aligned with available whole-genome sequences by using the est2genome (41) program. These sequences are available from DDBJ/GenBank/EMBL under accession numbers F557591 to F557853.

**Gene structure prediction and annotation.** All of the repetitive and low-complexity sequences in the T. orientalis genome sequence were masked by using RepeatMasker (http://www.repeatmasker.org) with
| Gene family                          | Product                       | TA ID                                      | TP ID                                | Signal | TMD | GPI |
|-------------------------------------|-------------------------------|--------------------------------------------|--------------------------------------|--------|-----|-----|
| PiroF0100038 TashAT family          | TA ID                         | TA03110, TA03115, TA03120, TA03125, TA03130, TA03135, TA03140, TA03145, TA03150, TA03155, TA03160, TA03165, TA20082, TA20083, TA20085, TA20090, TA20095 | TP01_0602, TP01_0603, TP01_0604, TP01_0605, TP01_0606, TP01_0607, TP01_0608, TP01_0609, TP01_0610, TP01_0611, TP01_0612, TP01_0613, TP01_0614, TP01_0615, TP01_0616, TP01_0617, TP01_0618, TP01_0619, TP01_0620 | 0      | 0   | N   |
| PiroF0100041 Hypothetical protein   | (TA9/TP9 family)              | TA16855, TA17050 (TA9), TA17510, TA17560 | TP02_0890, TP02_0895, TP02_0896, TP02_0891, TP02_0894 | Y      | 0   | N   |
| PiroF0100037 Theileria-specific subtelomeric protein, SVSP family | TA02740, TA04985, TA05540, TA05545, TA05550, TA05555, TA05560, TA05565, TA05570, TA05575, TA05580, TA09420, TA09425, TA09430, TA09435, TA09785, TA09790, TA09795, TA09800, TA09805, TA09810, TA09865, TA11385, TA11390, TA11395, TA11410, TA16025, TA16030, TA16035, TA16040, TA16045, TA17120, TA17125, TA17130, TA17135, TA17140, TA17346, TA17485, TA17485, TA17485, TA17545, TA17550, TA17555, TA18860, TA18865, TA18885, TA18890, TA18895, TA18950, TA19005, TA19060 | TP01_0004, TP01_0005, TP01_0006, TP01_0007, TP01_0008, TP01_0009, TP01_1225, TP01_1226, TP01_1227, TP02_0004, TP02_0005, TP02_0006, TP02_0007, TP02_0008, TP02_0010, TP02_0011, TP02_0953, TP02_0954, TP02_0955, TP02_0956, TP02_0958, TP02_0959, TP02_0960, TP03_0001, TP03_0002, TP03_0003, TP03_0004, TP03_0005, TP03_0049, TP03_0066, TP03_0067, TP03_0068, TP03_0069, TP03_0070, TP03_0071, TP03_0072, TP03_0073, TP03_0074, TP03_0075, TP03_0077, TP03_0078, TP03_0079, TP03_0080, TP03_0081, TP03_0082, TP03_0083, TP03_0084, TP03_0085, TP03_0086, TP03_0087, TP03_0088, TP03_0089, TP03_0090, TP03_0092, TP03_0093, TP03_0090, TP03_0091, TP03_0092, TP04_0002, TP04_0003, TP04_0004, TP04_0005, TP04_0006, TP04_0007, TP04_0008, TP04_0009, TP04_0010, TP04_0013, TP04_0014, TP04_0015, TP04_0016, TP04_0017, TP04_0018, TP04_0019, TP04_0016, TP04_0017, TP04_0018, TP04_0019, TP04_0019, TP04_0019, TP04_0920, TP04_0923, TP04_0927 | 0      | 0   | N   |
| PiroF0003401 Hypothetical protein   | TA09990                       | TA09990                                    | TP01_0378                            | Y      | 0   | N   |
| PiroF0003402 Hypothetical protein   | TA20985                       | TA20985                                    | TP01_0379                            | Y      | 0   | N   |
| PiroF0003403 Hypothetical protein   | TA20781                       | TA20781                                    | TP01_0438                            | Y      | 0   | N   |
| PiroF0003404 Cysteine repeat modular protein homologue, putative | TA20615 | TA20615 | TP01_0487 | Y | 1   | N   |
| PiroF0003405 Hypothetical protein   | TA20325                       | TA20325                                    | TP01_0549                            | Y      | 6   | N   |
| PiroF0003421 Hypothetical protein   | TA18750                       | TA18750                                    | TP03_0632                            | Y      | 1   | N   |
| PiroF0003422 Theileria-specific hypothetical protein | TA18535 | TA18535 | TP03_0582 | Y | 0   | N   |
| PiroF0003423 Theileria-specific hypothetical protein | TA17695 | TA17695 | TP03_0678 | Y | 1   | N   |
| PiroF0003424 Hypothetical protein   | TA17220                       | TA17220                                    | TP04_0030                            | Y      | 1   | Y   |
| PiroF0003425 Hypothetical protein   | TA17215                       | TA17215                                    | TP04_0029                            | Y      | 1   | N   |
| PiroF0003426 Hypothetical protein   | TA16020                       | TA16020                                    | TP02_0952                            | Y      | 0   | N   |
| PiroF0003427 Hypothetical protein   | TA15695                       | TA15695                                    | TP02_0888                            | Y      | 0   | N   |
| PiroF0003428 Hypothetical protein   | TA13955                       | TA13955                                    | TP02_0065                            | Y      | 0   | N   |
| PiroF0003429 Hypothetical protein   | TA11050                       | TA11050                                    | TP04_0896                            | Y      | 1   | N   |

(Continued on following page)
Signal peptides were inferred by SignalP 3.0 (53), TOT-SOUP/G-integra (48, 52). The numbers of manually curated genes was performed by using a custom-made annotation system named 51). Finally, to identify the representative genes, we essentially used annotation procedures described previously (48, 49). For each gene family, putative protein family member, putative fragment-related protein family. We then assigned a standardized functional annotation to each gene as primarily from gene annotation. T. annulata definitions.

| Gene family | Producta | TA ID   | TP ID     | Signalb | TMDc | GPIb |
|------------|----------|---------|-----------|---------|------|------|
| PiroF0003520 | Hypothetical protein | TA11020 | TP04_0585 | Y       | 1    | N    |
| PiroF0003524 | Hypothetical protein | TA10740 | TP04_0642 | Y       | 0    | Y    |
| PiroF0003546 | Sf1 subtelomeric | TA09140 | TP04_0116 | Y       | 0    | N    |
| PiroF0003548 | Hypothetical protein | TA08935 | TP04_0539 | Y       | 2    | N    |
| PiroF0003567 | Hypothetical protein | TA06680 | TP01_0719 | Y       | 0    | N    |
| PiroF0003568 | Hypothetical protein | TA06675 | TP01_0718 | Y       | 1    | N    |
| PiroF0003582 | Hypothetical protein | TA05315 | TP03_0135, TP03_0134 | Y       | 0    | N    |
| PiroF0003592 | Hypothetical protein, conserved | TA04390 | TP03_0410 | Y       | 2    | N    |
| PiroF0003612 | Hypothetical protein | TA02590 | TP03_0038 | Y       | 0    | Y    |
| PiroF0003613 | Hypothetical protein | TA02580 | TP03_0040 | Y       | 0    | N    |

T. orientalis genes were first predicted computationally by using T. orientalis EST pair gene models and several gene prediction programs and then finally identified by genome-wide manual curation. T. orientalis EST sequences, identified from a full-length cDNA library made from parasite-infected erythrocytes, were mapped onto the T. orientalis genome. Based on EST-genome alignments using est2genome (41), EST pair gene models were constructed by merging the exon overlap on the same strand of ESTs of the same clone. We identified 544 T. orientalis EST pair gene models. Genes were predicted by several gene-finding software packages, including GlimmerHMM (44), GeneMark.hmm (45), GeneWise (46), and JIGSAW (47). GlimmerHMM was trained on two sets of full-length gene sequences. The first set consisted of T. orientalis genes (544 EST pair gene models), and the second set consisted of these T. orientalis genes and annotated genes of T. parva and T. annulata that were predicted to be longer than 400 amino acids. GeneWise was trained on all sets of results of genome coordinates provided by GlimmerHMM, GeneMark.hmm, GeneWise, and T. orientalis (544 EST pair gene models) were summarized by using JIGSAW. JIGSAW was also trained on T. orientalis EST pair gene models.

We essentially used annotation procedures described previously (48, 49). For each T. orientalis gene product, we conducted InterProScan (50). We then assigned a standardized functional annotation to each gene as illustrated in Fig. 54 in the supplemental material, based on the results of a BLASTX similarity search against the UniProtKB/Swiss-Prot, UniProtKB/TremBL, and RefSeq protein databases and InterProScan (48, 51). Finally, to identify the representative T. orientalis genes, manual curation was performed by using a custom-made annotation system named TOT-SOUP/G-integra (48, 52). The numbers of manually curated T. orientalis genes were summarized in Table S4 in the supplemental material. Signal peptides were inferred by SignalP 3.0 (53).

**Ortholog clustering.** Ortholog groups consisted of T. orientalis, T. annulata, T. parva, B. bovis, P. falciparum, and P. vivax proteins derived primarily from gene annotation. T. annulata orthologs were from GeneDB (http://old.genedb.org/genedb/annulata/); T. parva, except for the mitochondrion proteome, and B. bovis orthologs were from RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/); the T. parva mitochondrial proteome was from UniProt (http://www.uniprot.org/); and P. falciparum and P. vivax orthologs were obtained from PlasmoDB (http://plasmodb.org/plasmo/). Ortholog groups were generated by OrthoMCL (23) on the basis of sequence similarity by using an all-versus-all NCBI BLASTP search (54) with a bit score cutoff of <60 and default parameters. Because E values from the BLASTP search were applied for a similarity measure, we recomputed the exact E values between closely related proteins if the E value was approximated at 0.0. We integrated the orthologous groups assumed to be duplicated in the Theileria lineage after separation from Babesia into a single group by using both automatic algorithms/software and manual integration as described below. Ortholog groups A and B were merged if any Theileria-Theileria gene pairs in which two genes belonging to A and B, respectively, had higher bit scores than any Theileria-Babesia/Plasmodium gene pairs within single ortholog group A or B. Several ortholog groups were merged by manual curation based on sequence homology and genomic location if they generated tandem arrays on the chromosomes. We also merged nonclustered genes using OrthoMCL into the ortholog groups with the same procedure. Finally, 3,502 ortholog groups were used for the following analyses; PiroF0100001 to Pi roF0100062 represent the merged ortholog groups, and PiroF0000001 to PiroF0000675 represent the other ortholog groups. The ortholog clustering left 436, 112, and 293 nonclustered genes in T. orientalis, T. annulata, and T. parva, respectively.

**KEGG metabolic pathway reconstruction.** Metabolic pathways in T. orientalis were analyzed by KEGG metabolic pathway reconstruction. First, BLAST searches were performed for protein sequences in each orthologous cluster against the KEGG GENES database. A KO identifier was then assigned to each cluster according to the most similar hit with a KO annotation; the E value threshold was <1.0E-5.

**Molecular phylogenetic analysis.** Amino acid sequences of each ortholog group were multiply aligned with the L-INS-I alignment strategy in MAFFT (55), and gap-rich sequences, such as truncated ones, were removed from the alignments with MaxAlign (56). Ambiguously and/or poorly aligned sites were removed by Gblocks (57), and the rest were subjected to phylogenetic analysis. Phylogenetic trees were inferred by maximum likelihood (ML) (58, 59) with a heuristic ML tree search using RAxML (60) with the WAG-F model (61). Heterogeneity of evolutionary rates among sites was modeled by a discrete gamma distribution, with optimization of gamma shape parameter alpha for each alignment set (62). Bootstrap probability (59) was calculated for each tree node with 1,000 replications.

**Generation of recombinant protein and antiserum.** A 1,788-bp fragment of TA03110 was PCR amplified with the C9 (genome) strain of T. annulata as a template. This corresponds to the full-length encoded protein minus the N-terminal signal peptide sequence and spans nucleotide positions 70 to 1,857 relative to the translation start codon. In addition to gene-specific sequences, the PCR primers incorporated attB adapters to facilitate the use of Gateway Recombination Cloning Technology (Invitrogen); the forward primer was 5’-forward attB adaptor-
GAGGACTTGGACCTAAGCTTCC-3', and the reverse primer was 5'-
reverse attb adaptor-AGGATTTTGATGCTGTTAAATGTC-3'.
The amplicon was cloned into the pDONR221 shuttle vector and subcloned into the expression vector pDEST17, which has a six-histidine (His6-) repeat at the 5' end of the multiple cloning site. After the transformation of chemically competent Escherichia coli BL21 cells (Invitrogen), expression of the His6-tagged fusion protein was induced by adding l-arabinose to a final concentration of 0.2% in LB liquid medium. Recombinant protein was purified by affinity chromatography on nickel agarose columns under denaturing conditions by using the manufacturer’s protocol (Qiagen). Eluted fractions containing the recombinant protein were assessed by using SDS-PAGE before being pooled. To generate polyclonal anti-Ta03110 serum, two rats were immunized a total of four times with 30 µg of recombinant protein per immunization. Immunizations were conducted under a project license issued by the United Kingdom Home Office, i.e., Animals (Scientific Procedures) Act 1986 contract immunization project license PPL 50/3464.

Parasite material and IFAT. The T. annulata-infected cloned cell line Ankara A, D7 (26) was used to provide material for the microarray experiment and for the IFAT. To stimulate differentiation from the macroschizont stage to the merozoite stage, cultures were maintained at 41°C by using a previously described protocol (26). Cytospin preparation of T. annulata-infected parasitaemia, paraformaldehyde fixation, and the IFAT were performed as described previously (62). The anti-Tash-a serum was used at a 1:200 dilution. Alexa 555 (Invitrogen) were used at a 1:200 dilution hybridoma culture medium as previously described (63); anti-rat (sc-65902; Stratagene) was used at 1:200; monoclonal antibodies against a protein and small subunit ribosomal RNA genes. Int. J. Parasitol. 36:9–21.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org./lookup/suppl/doi:10.1128/mBio.00204-12/-/DCSupplemental.

Data set S1, XLSX file, 0.4 MB.
Figure S1, TIF file, 0.5 MB.
Figure S2, TIF file, 0.6 MB.
Figure S3, TIF file, 0.7 MB.
Figure S4, TIF file, 0.4 MB.
Table S1, PDF file, 0.1 MB.
Table S2, PDF file, 0.01 MB.
Table S3, PDF file, 0.01 MB.
Table S4, PDF file, 0.01 MB.

ACKNOWLEDGMENTS

We are grateful to the technical staff of the Sequencing Technology Team at RIKEN GSC and the Department of Medical Genome Sciences, Graduate School of Frontier Sciences, at the University of Tokyo for their assistance.

This research was supported by Grants-in-Aid for Scientific Research (http://www.jsps.go.jp/english/e-grants/grants.html) from the Ministries of Education, Culture, Science, Sport, and Technology (17208026, 21248035) to C.S. and the Program of Founding Research Centers for Emerging and Reemerging Infectious Diseases (http://www.crnid.riken.jp/english/index.html) from the Ministries of Education, Culture, Sports, Science, and Technology. Work performed by B.R.S. and W.W. was funded by a grant from the Wellcome Trust (083488/Z/07/Z).

The funder had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

REFERENCES

1. Brown CG, Stagg DA, Purnell RE, Kanhai GK, Payne RC. 1973. Letter: infection and transformation of bovine lymphoid cells in vitro by infective particles of Theileria parva. Nature 243:101–103.
2. Irvin AD, Brown CG, Kanhai GK, Stagg DA. 1975. Comparative growth of bovine lymphosarcoma cells and lymphoid cells infected with Theileria parva in athymic (nude) mice. Nature 255:713–714.
3. Brown CG. 1990. Control of tropical theileriosis (Theileria annulata infection) of cattle. Parasitology 103:23–31.
4. Hooshmand-Rad P, Hawa NJ. 1973. Malignant theileriosis of sheep and goats. Trop. Anim. Health Prod. 5:97–102.
5. Shids B, et al. 2006. Alteration of host cell phenotype by Theileria annulata and Theileria parva: mining for manipulators in the parasite genomes. Int. J. Parasitol. 36:9–21.
6. Dobbelrae D, Baumgartner M. 2009. Theileria parasite infection in East Asia and control of the disease. Comp. Immunol. Microbiol. Infect. Dis. 21:165–177.
7. Uilenberg G. 2011. Theileria sergenti. Vet. Parasitol. 175:386.
8. Shimizu S, Yoshiura N, Mizomoto T, Kondou Y. 1992. Theileria sergenti infection in dairy cattle. Science 309:134–137.
9. Pain A, et al. 2005. Genome of the host-cell transforming parasite Theileria annulata compared with T. parva. Science 309:131–133.
10. Bishop RP, et al. 2009. Theileria, p. 191–224. In Nene V, Kole C (ed), Genome mapping and genomics in animal-associated microbes. Springer-Verlag, Berlin, Germany.
11. Onuma M, Kakuda T, Sugimoto C. 1998. Theileria parasite infection in East Asia and control of the disease. Comp. Immunol. Microbiol. Infect. Dis. 21:165–177.
12. Uilenberg G. 2011. Theileria sergenti. Vet. Parasitol. 175:386.
13. Uilenberg G. 2011. Theileria orientalis, a cosmopolitan blood parasite of cattle: demonstration of the schizont stage. Res. Vet. Sci. 38:352–360.
14. Fujisaki K, Kawazu S, Kamio T. 1994. The taxonomy of the bovine Theileria spp. Parasitol. Today 10:31–33.
15. Kubota S, Sugimoto C, Onuma M. 1996. Population dynamics of Theileria sergenti in persistently infected cattle and vector ticks analysed by a polymerase chain reaction. Parasitology 112(PT 5):437–442.
16. Kamau J, et al. 2011. Emergence of new types of Theileria orientalis in Australian cattle and possible cause of theileriosis outbreaks. Parasit. Vectors 4:22.
17. Kakuda T, et al. 1998. Phylogeny of benign Theileria species from cattle in Thailand, China and the U.S.A. based on the major piroplasm surface protein and small subunit ribosomal RNA genes. Int. J. Parasitol. 28:1261–1267.
18. Sato M, et al. 1993. Histological observations on the schizonts in cattle infected with Japanese Theileria sergenti. J. Vet. Med. Sci. 55:571–574.
19. Hikosaka K, et al. 2010. Divergence of the mitochondrial genome structure in the apicomplexan parasites, Babesia and Theileria. Mol. Biol. Evol. 27:1107–1116.
20. Schmuckli-Maurer J, et al. 2009. Expression analysis of the Theileria parva subtelomere-encoded variable secreted protein gene family. PLoS One 4:e8439.
21. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 38:D355–D360.
22. Chae JS, et al. 1999. A study of the systematics of Theileria spp. Based on small-subunit ribosomal RNA gene sequences. Parasitol. Res. 85:877–883.
23. Li L, Stoeckert CJ, Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.
24. Swan DG, et al. 2003. Characterisation of a cluster of genes encoding Theileria annulata AT hook DNA-banding proteins and evidence for localisation to the host cell nucleus. J. Cell Sci. 114:2747–2754.
25. Swan DG, Phillips K, Tait A, Shids BR. 1999. Evidence for localisation of a Theileria parasite AT hook DNA-binding protein to the nucleus of immortalised bovine host cells. Mol. Biochem. Parasitol. 101:117–129.
26. Shids B, et al. 1992. Disruption of synchrony between parasite growth
and host cell division is a determinant of differentiation to the merozoite in *Theileria annulata*. J. Cell Sci. 101(Pt 1):99–107.

27. Baylis HA, Sohal SK, Carrington M, Bishop RP, Allsopp BA. 1991. An unusual repetitive gene family in *Theileria parva* which is stage-specifically transcribed. Mol. Biochem. Parasitol. 49:133–142.

28. Skilton RA, et al. 2000. A 32 kDa surface antigen of *Theileria parva*: characterization and immunization studies. Parasitology 120(Pt 6):553–564.

29. McKeever DJ, et al. 1994. Adoptive transfer of immunity to *Theileria parva* in the CDB+ fraction of responding efferent lymph. Proc. Natl. Acad. Sci. U. S. A. 91:1959–1963.

30. Graham SP, et al. 2006. *Theileria parva* candidate vaccine antigens recognized by immune bovine cytotoxic T lymphocytes. Proc. Natl. Acad. Sci. U. S. A. 103:3286–3291.

31. Graham SP, et al. 2008. Characterization of the fine specificity of bovine CDB T-cell responses to defined antigens from the protozoan parasite *Theileria parva*. Infect. Immun. 76:685–694.

32. Hattori M, et al. 2004. Oligo-capping: a simple method to replace oligo-d(T) in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2:953–971.

33. Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

34. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:518–529.

35. Tomomasa K, Sugimoto C, Onuma O. 1995. A genetic analysis of mixed population in *Theileria sergenti* stocks and isolates using allele-specific polymerase chain reaction. J. Vet. Med. Sci. 57:279–282.

36. Hattori M, et al. 1997. A novel method for making nested deletions and its application for sequencing of a 300 kb region of human APP locus. Nucleic Acids Res. 25:1802–212.

37. Kubota S, Sugimoto C, Onuma O. 1995. A genetic analysis of mixed population in *Theileria sergenti* stocks and isolates using allele-specific polymerase chain reaction. J. Vet. Med. Sci. 57:279–282.

38. Shiels BR, et al. 2001. A general empirical model of protein database search programs. Nucleic Acids Res. 29:3422–3423.

39. Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18:691–699.

40. Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.

41. Shields BR, et al. 2004. A *Theileria annulata* DNA binding protein localized to the host cell nucleus alters the phenotype of a bovine macrophage cell line. Eur. J. Cell Biol. 79:495–505.

42. Carver TJ, et al. 2005. ACT: the Artemis Comparison Tool. Bioinformatics 21:3422–3423.