

METHODS AND RESOURCES

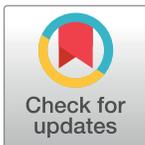
Comparative single-cell transcriptional atlases of *Babesia* species reveal conserved and species-specific expression profiles

Yasaman Rezvani¹✉, Caroline D. Keroack²✉, Brendan Elsworth², Argenis Arriojas^{1,3}, Marc-Jan Gubbels⁴, Manoj T. Duraisingh^{2*}, Kourosh Zarringhalam^{1,5*}

1 Department of Mathematics, University of Massachusetts Boston, Boston, Massachusetts, United States of America, **2** Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, United States of America, **3** Department of Physics, University of Massachusetts Boston, Boston, Massachusetts, United States of America, **4** Department of Biology, Boston College, Chestnut Hill, Massachusetts, United States of America, **5** Center for Personalized Cancer Therapy, University of Massachusetts Boston, Boston, Massachusetts, United States of America

✉ These authors contributed equally to this work.

* mduraisi@hsph.harvard.edu (MTD); kourosh.zarringhalam@umb.edu (KZ)



OPEN ACCESS

Citation: Rezvani Y, Keroack CD, Elsworth B, Arriojas A, Gubbels M-J, Duraisingh MT, et al. (2022) Comparative single-cell transcriptional atlases of *Babesia* species reveal conserved and species-specific expression profiles. *PLoS Biol* 20(9): e3001816. <https://doi.org/10.1371/journal.pbio.3001816>

Academic Editor: Kami Kim, University of South Florida, UNITED STATES

Received: February 21, 2022

Accepted: September 5, 2022

Published: September 22, 2022

Copyright: © 2022 Rezvani et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: An interactive web-application for visualization and exploration of our datasets can be accessed here: <https://umbibio.math.umb.edu/babesiasc/>. Source code available on GitHub: https://github.com/umbibio/babesia_dash_app (DOI: [10.5281/zenodo.7038292](https://doi.org/10.5281/zenodo.7038292)). The analysis R code is available on GitHub at: <https://github.com/umbibio/scBabesiaAtlases> (DOI: [10.5281/zenodo.7036500](https://doi.org/10.5281/zenodo.7036500)). scRNAseq data (fastq) have been deposited to the Sequence Read Archive (SRA) under the accession number PRJNA803312.

Abstract

Babesia is a genus of apicomplexan parasites that infect red blood cells in vertebrate hosts. Pathology occurs during rapid replication cycles in the asexual blood stage of infection. Current knowledge of *Babesia* replication cycle progression and regulation is limited and relies mostly on comparative studies with related parasites. Due to limitations in synchronizing *Babesia* parasites, fine-scale time-course transcriptomic resources are not readily available. Single-cell transcriptomics provides a powerful unbiased alternative for profiling asynchronous cell populations. Here, we applied single-cell RNA sequencing to 3 *Babesia* species (*B. divergens*, *B. bovis*, and *B. bigemina*). We used analytical approaches and algorithms to map the replication cycle and construct pseudo-synchronized time-course gene expression profiles. We identify clusters of co-expressed genes showing “just-in-time” expression profiles, with gradually cascading peaks throughout asexual development. Moreover, clustering analysis of reconstructed gene curves reveals coordinated timing of peak expression in epigenetic markers and transcription factors. Using a regularized Gaussian graphical model, we reconstructed co-expression networks and identified conserved and species-specific nodes. Motif analysis of a co-expression interactome of AP2 transcription factors identified specific motifs previously reported to play a role in DNA replication in *Plasmodium* species. Finally, we present an interactive web application to visualize and interactively explore the datasets.

Introduction

Apicomplexan parasites of the genus *Babesia* are some of the most widespread blood parasites of vertebrates, second only to the trypanosomes [1]. Babesiosis has long been recognized as a

Funding: This work was supported by grants AI150090 (KZ, MJG) and 1R21AI153945 (MD) from the National Institutes of Health. CDK was supported by an AHA pre-doctoral fellowship (#19PRE34380106). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: FC, fold change; DEG, differentially expressed gene; DTW, dynamic time warping; GAM, generalized additive model; GO, Gene Ontology; GOEA, Gene Ontology enrichment analysis; IDC, intraerythrocytic developmental cycle; PCA, principal component analysis; RBC, red blood cell; RNA-seq, RNA sequencing; scRNA-seq, single-cell RNA sequencing; TF, transcription factor; UMAP, Uniform Manifold Approximation and Projection; VESA, variable erythrocyte surface antigen.

disease of tremendous veterinary and agriculture importance, causing hundreds of millions of dollars of economic losses every year [2,3]. Since the first reported case of human babesiosis, caused by *B. divergens*, reported in 1956, and compounded by the emergence of *B. microti* in the US, babesiosis has steadily been gaining recognition as an important human parasitic disease [4–7]. The disease can range from mild febrile illness to severe, life-threatening disease, particularly in immunocompromised patients [6,8]. *Babesia* is usually transmitted through the bite of an infected tick [9], but can also be transmitted congenitally and via blood transfusion [10–12]. Indeed, *Babesia* is listed as a top priority pathogen in the blood supply [13]. Beyond human pathogens, there are at least 100 species of *Babesia* described that cause disease in a variety of hosts [6,14]. Bovine babesiosis, predominantly caused by *B. bovis*, *B. bigemina*, and *B. divergens*, is of significant concern, often resulting in fulminating infection and high mortality, leading to significant economic and agricultural losses [15,16]. With *Babesia* representing such a wide diversity of disease-causing parasites, identifying both conserved and divergent biology is essential to developing therapeutic and vaccine interventions.

In the asexual replicative cycle, *Babesia* species are obligate intracellular parasites that infect red blood cells (RBCs). While knowledge about the morphology of *Babesia* parasites during these division cycles exists, molecular details of the asexual replication cycle are limited. Transcriptomic studies have been done on various *Babesia* spp. populations to profile life stage, egress and invasion, and virulence [17–22]. While these studies provide rich data resources, no transcriptomic data have been generated to comprehensively describe the asexual replication cycle, as has been done in other related parasites [23–27]. Most knowledge of the molecular mediators of the *Babesia* spp. replication cycle has been gleaned through comparative approaches with *Plasmodium* spp. and *Toxoplasma gondii* [28], and through the analysis of genomic sequences [17,18,29–36]. Currently, synchronization of *B. divergens* and *B. bovis* relies on mechanical release of parasites from the RBC, with other species still unable to be synchronized [37,38]. The current methods of synchronization are variable and do not yet allow for tightly synchronized populations of parasites. Consequently, to the best of our knowledge, only a single synchronous transcriptomic dataset exists to date [39]. While single-cell RNA sequencing (scRNA-seq) can overcome the difficulties of synchronization, no such data have yet been generated for *Babesia* species.

Due to these gaps in knowledge, and the limitations of synchronization, scRNA-seq presents a promising method for delineating the *Babesia* intraerythrocytic replication cycle at a fine scale. This approach offers a powerful, unbiased method of profiling heterogeneous and asynchronous cell populations. scRNA-seq has been used successfully in a range of other apicomplexan parasites and has provided key insights into parasite biology not previously detectable using bulk RNA sequencing (RNA-seq) methods. In *Plasmodium* species, scRNA-seq has been used to describe cell populations through the entire life cycle, from the intraerythrocytic developmental cycle (IDC) through the mosquito [40–47]. This has led to the development of comprehensive cell atlases for both *P. berghei* [41] and *P. falciparum* [48]. Additionally, scRNA-seq in *T. gondii* was used to reveal novel regulators of life cycle stage progression [49,50]. Combined with the ability to culture many diverse species of *Babesia* in vitro [51,52], comparative scRNA-seq offers a unique opportunity in *Babesia* to identify core, conserved regulators of progression through the replicative cycle.

Here, we performed scRNA-seq on 3 bovine *Babesia* species that can be readily cultured in vitro: *B. bovis*, *B. bigemina*, and *B. divergens*. Additionally, to investigate a possible role of the host RBC on the replication cycle, we performed scRNA-seq on *B. divergens* adapted to either human RBCs or bovine RBCs in in vitro culture. Using these 4 datasets, we reconstructed the transcriptome of the asexual replication cycle for *Babesia* spp. and mapped the transition points of the developmental phases. We show that despite the evolutionary divergence of these

Babesia species, the underlying expression patterns of their replication cycles are highly similar. Using these data, we were also able to reconstruct gene co-expression networks and quantify the interactomes of gene families important in each phase of development. Analysis of the interactomes reveals a core conserved set of replication cycle regulators. Taken together, these data represent, to our knowledge, the first asexual replication cycle atlases for any *Babesia* species, as well as the first single-cell transcriptomic datasets. To facilitate usage, we provide an interactive web application for visualization and exploration of our datasets. The web app provides functionality to profile gene expression, perform comparative transcriptomics analysis, assess and visualize gene expression timing across the species, and explore co-expression networks. The web app is available at <https://umbibio.math.umb.edu/babesiasec/>.

Results

scRNA-seq analysis reveals the characteristic cyclical pattern of expression in replicating *Babesia* blood-stage parasites

To characterize replication cycle progression in *Babesia* spp., we performed scRNA-seq on 3 *Babesia* species: *B. bigemina* and *B. bovis* of bovine origin, and *B. divergens* of human origin. To interrogate potential host-cell-specific differences, we collected *B. divergens* parasites growing in vitro in both bovine and human RBCs. The alignment rates were over 90% in all species. Downstream processing and alignment detected between 8,719 and 12,910 cells across the 4 species. The median number of genes detected per cell varied from 417 to 629. Additional filters were applied using the Seurat R package [53]. The number of cells and genes that passed the Seurat cutoffs varied from 3,200 to 4,000 and from 3,563 to 3,810, respectively, across the 3 species. All alignment and data quality metrics are available in [S1 Table](#). Statistics on *Babesia* genomes regarding the percentage of exons, introns, and intergenic regions, as well as the proportions of overlapping genes, are provided in [S2 Table](#). Expression datasets were combined using the expression of 2,548 orthologous genes obtained from reciprocal blast hits and visualized on Uniform Manifold Approximation and Projection (UMAP) and principal component analysis (PCA) coordinates ([Fig 1](#)). The projected data reveal a circular pattern shared across all *Babesia* species, that is characteristic of asexual replication in both *T. gondii* and *Plasmodium* spp. [23,41,50]. This circular pattern of expression is a consequence of periodicity in gene expression during the replication cycle and the existence of clusters of co-expressed genes. In the UMAP projection, *B. divergens* in human host cells and *B. divergens* in bovine host cells cluster together, indicating that the different hosts result in relatively few changes overall compared to the differences between *Babesia* species. The same analysis was repeated for each dataset independently using all genes, and also showed a circular pattern ([Fig A in S1 Text](#)).

Pseudo-time analysis maps the progression of gene expression and transition points in the asexual replicative cycle

To map the progression of gene expression during the replication cycle, we performed a pseudo-time analysis by fitting an elliptic principal curve to the first 2 PCA coordinates of *B. divergens* [54]. Cells were orthogonally projected onto the curve and ordered to mimic pseudo-time ([Fig 2A](#)). The start point was arbitrarily picked and set to 0 and the end point was set to 12 h [37,38], representing the approximate time of a division cycle in *Babesia* spp. [38,39]. We partitioned the projected cells along the pseudo-time course into 20-min time intervals. Cells in each partition were considered “synchronized replicates.” Expression of genes along the pseudo-time course were then used to construct gene expression curves.

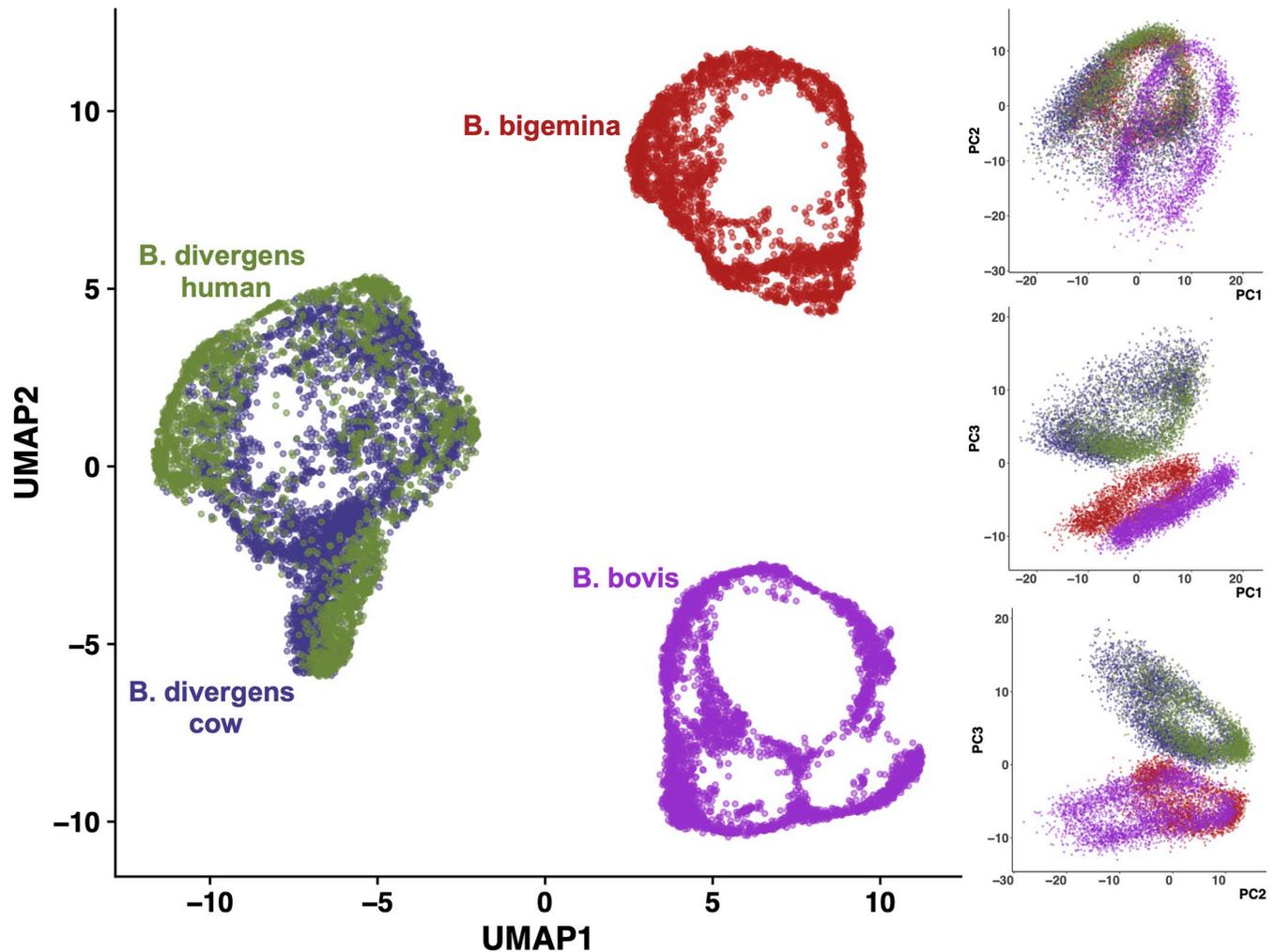


Fig 1. Projection of single-cell RNA sequencing data onto UMAP coordinates and the first 3 PC coordinates. *B. divergens* in human and bovine host cells cluster together. PCA, principal component; UMAP, Uniform Manifold Approximation and Projection. Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g001>

Genes whose expression did not show temporal correlation with pseudo-time were identified by fitting a generalized additive model (GAM) and calculating the goodness of fit (adjusted p -value < 0.01). These include constitutively expressed genes as well as genes that are sporadically and randomly expressed. Depending on the dataset, a total of 2,504–2,741 genes passed the criteria and were retained for further analysis (S1 Table). It should be noted that the scale of pseudo-time differs from that of the actual cell division cycle time. Information on the actual span of cell cycle phases is needed to match the 2 through piecewise linear scaling. However, this scaling has no impact on the subsequent analysis of relative timing of expression, identification of differentially expressed genes (DEGs), or clustering of gene curves.

To adjust the start time and map the direction of replicative cycle progression, we utilized gene expression data from synchronized bulk RNA-seq from *B. divergens* [39]. The time course data consisted of 7 bulk RNA-seq measurements of synchronized *B. divergens* parasites over a 12-h period. A total number of 3,993 genes were detected in the synchronized data, of which 2,236 showed significant expression changes over the time period (adjusted p -value < 0.01). We calculated the cross-correlation of gene expression curves in single cells and

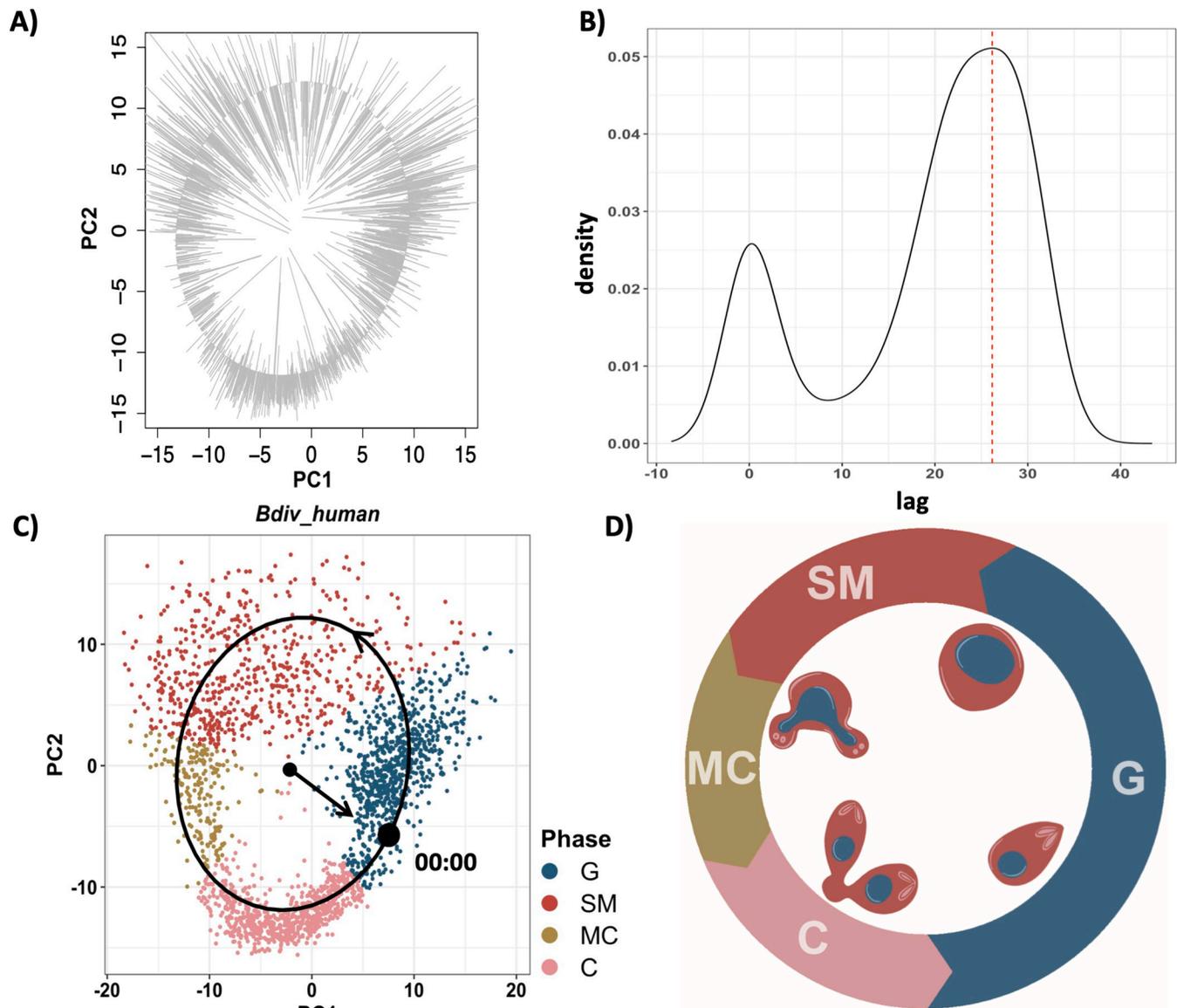


Fig 2. Pseudo-time reconstruction. (A) Cells were orthogonally projected on the elliptic principal curve fitted to the first 2 PCA coordinates and ordered sequentially using a random start point. The pseudo-time gene curve was constructed by mapping the cell orders to 0–12 h. (B) Distribution of the lag time between pseudo-time gene curves and matched gene curves from synchronized bulk measurements. Lag time was calculated using cross-correlation. The bimodal distribution indicates 2 optimal lag times corresponding to monotonic curves (lag 0) and cyclic curves (red dashed line). (C) PCA plot of *B. divergens* in human RBCs. Start time (black dot) was adjusted according to the optimal lag time corresponding to the larger peak in (B). The black arrow on the principal elliptic curve indicates the flow of time. Transition points were mapped to the PCA coordinates in *B. divergens* in human RBCs, and the colors represent each inferred phase. (D) Schematic representation of the *Babesia* spp. cell division cycle in RBCs. PC, principal component; PCA, principal component analysis; RBC, red blood cell. Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g002>

the corresponding genes in synchronized bulk experiments and identified the lag time that maximized the correlation. Essentially, in the cross-correlation analysis, we shift the start point of the expression curve in the scRNA-seq data and calculate the correlation with the bulk data. This is done for each gene independently, and the lag time maximizing the cross-correlation is calculated. Fig 2B shows the lag time distribution across all genes in *B. divergens*. The 2 distinct peaks correspond to the optimal shift for monotonically expressed genes (no lag) and peaking

genes (lag at 28). The lag corresponding to the second peak was used to adjust the lag time across all genes and to set the start point of the replication cycle clock in single cells (Fig 2C). Fig B in S1 Text shows a few example curves of genes corresponding to the optimal lag time, as well as genes corresponding to the lag time 0.

To identify the transition points through the replication cycle of *Babesia* spp., we used previously assigned replication cycle genes from *T. gondii*. Although transition points can be inferred computationally, we reasoned that *T. gondii* would be most informative because of its similar binary division modes shared with *Babesia* spp. and because the *T. gondii* phases of the cell division cycle were identified using DNA content analysis [50]. Moreover, computationally inferred phases generally agree with *T. gondii* inferred phases (Fig C in S1 Text). There is a high correlation between peak expression time of DEGs and the replicative cycle in *T. gondii* (Fig D in S1 Text). The timing of peak expression of the *Babesia* spp. orthologous genes of the top 20 *T. gondii* replicative cycle marker genes was used to map transition time points. There is an overlap between the distribution of peak time expression of S and M as well as M and C markers (Fig D in S1 Text). Further information on mapping the transition points can be found in S1 Text.

Based on this, we opted to define 4 “inferred” phases labeled as G, SM, MC, and C and marked the transitions to maximize the separation between the phases. These transition time points were used in pseudo-time to infer developmental phase transition points in *Babesia* spp. Fig 2C shows the PCA projection of *B. divergens* in human host cells, with the colors indicating the “inferred” phases, accompanied by a schematic depiction of the replication cycle (Fig 2D). These results show that by leveraging the geometry of gene expression data in asynchronously dividing parasites, single-cell gene expression can be converted to “pseudo-synchronized” time-course data at a fine resolution. This approach can be generally used to overcome some of the limitations of time-consuming and costly time-course experiments.

Gene expression is highly correlated between synchronized bulk and pseudo-synchronized single-cell sequencing

Using the newly inferred phases, we sought to assess the correlation of peak time as well as overall similarity of gene curves in the single-cell and synchronized bulk data. First, we fitted the curves with smoothing splines and interpolated points at regular intervals, aligned the scRNA-seq and bulk curves [55], and measured the distance between the curves (Materials and Methods). Time warping was applied to account for scaling differences between pseudo-time and real cell division time. Fig 3A shows some representative examples of high (left) and low (right) alignment for genes that peak (top) or deplete (bottom) in inferred SM phase. Next, we calculated peak expression time for each gene in both datasets. The peak assignment was limited to markers of S/M/C phases, where genes tend to have a more defined peak that lasts for a short period of time. The boundary cases of transitioning from C back to G were excluded, as marker genes assigned to G tend to have a flat expression period, followed by depletion at SM phase that slightly upticks in late C and transitions back to G. Fig 3B (top left) shows the correlation between peak expression times for S/M/C genes. The overall normalized alignment distance as measured by the dynamic time warping (DTW) algorithm was categorized into high, mid, and low, based on the distribution of the alignment distance (Fig 3B, top right and bottom). Overall, the data show that there is a high level of agreement between reconstructed pseudo-time in single-cell and synchronized time-course data, both in peak time expression as well as overall alignment. The bulk RNA-seq was used as a calibration point to set the start of the cell cycle in scRNA-seq and to validate the reconstructed expression curve, and played no role in subsequent analysis.

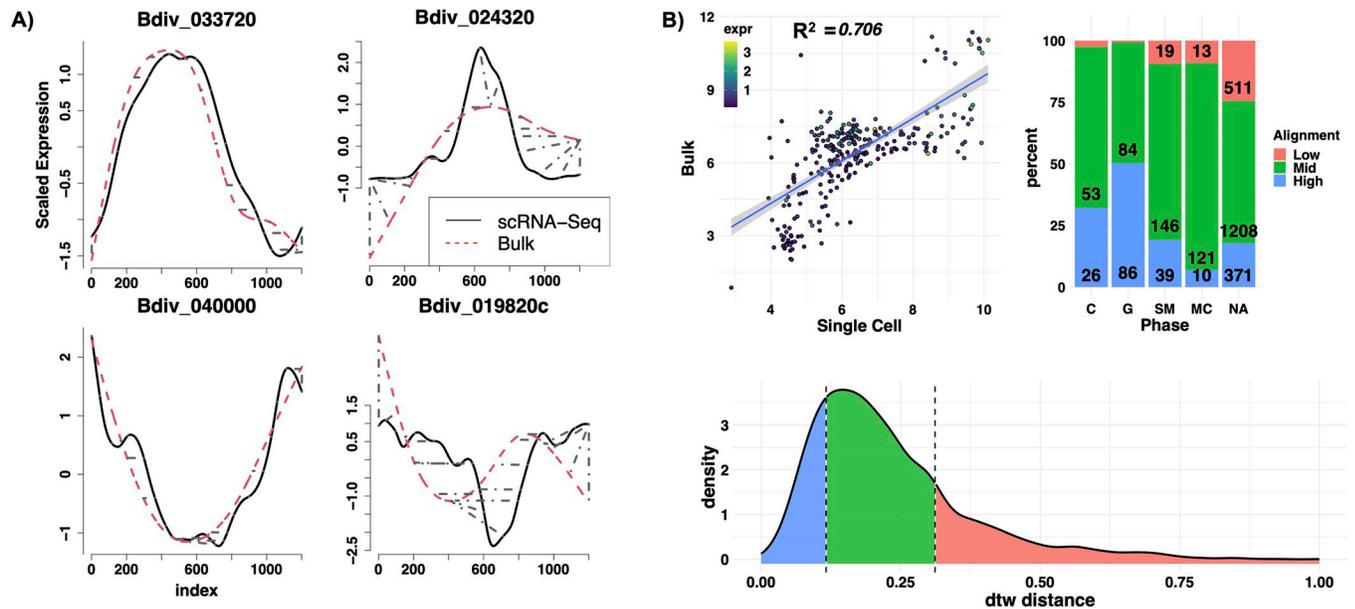


Fig 3. Quantification of alignment of gene curves between synchronized bulk and single-cell sequencing. (A) Dynamic time warping alignment of a sample gene: high alignments (left), low alignments (right). Gray dashed lines indicate the matched time index (warping). (B) Top left: Correlation between calculated peak times in S/M/C phases, excluding the boundary. Top right: Percent alignment distance categorized into high (distance < lower 20th percentile), low (distance > upper 80th percentile), and mid. Curves with no peak contain a higher percentage of poor alignment. Bottom: Distribution of dynamic time warping (dtw) alignment distance between single-cell and bulk sequencing. Shaded areas show boundaries of alignment categories. NA, not available. Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g003>

Differential gene expression analysis of inferred phases

We performed differential gene expression analysis to identify (1) cross-species conserved replication cycle DEGs, (2) species-specific phase DEGs (i.e., replication cycle phase DEGs highly upregulated in 1 species), and (3) replication cycle phase DEGs of each species independent of other species (fold change [FC] > 2 and adjusted p -value < 0.01). The bar plots in Fig 4A, 4B, 4C, and 4D respectively represent the number of species-specific DEGs, the percentage of DEGs in each phase in each species, conserved replication cycle DEGs, and species-specific DEGs. Conserved DEGs were calculated using the built-in functions in the Seurat R package (Materials and Methods), which is more lenient than calculating conserved DEGs (and species-specific DEGs) by taking the intersection (and set difference) of DEGs across species. We repeated this analysis using intersection and set differences as well (Fig E in S1 Text). There is a minor difference between total number of DEGs detected; however, the overall results regarding the enrichment analysis are similar. Full lists of DEGs are provided in S3 Table. We performed a Gene Ontology enrichment analysis (GOEA) to investigate biological functions linked to each set of DEGs. For this analysis, genes were mapped to their *T. gondii* orthologs, and enrichment analysis was performed using the available Gene Ontology (GO) terms on ToxoDB. GO term and gene set categories with fewer than 10 genes were excluded from this analysis. Multiple principal significant GO terms are highlighted for the conserved genes in Fig 4E. All significant GO term hits can be found in S4 Table. The top-ranked GO term for the C inferred phase across the *Babesia* species is actin cytoskeleton, including several genes involved in actin polymerization (profilin, Bdiv_003910c; actin depolymerizing factor [ADF], Bdiv_021160), as well as myosin A (Bdiv_001770c). These genes are consistent with the cytoskeletal remodeling that occurs during cytokinesis. The SM and MC most highly enriched term relates to the pellicle. In *Babesia* parasites the pellicle is the structural site that organizes

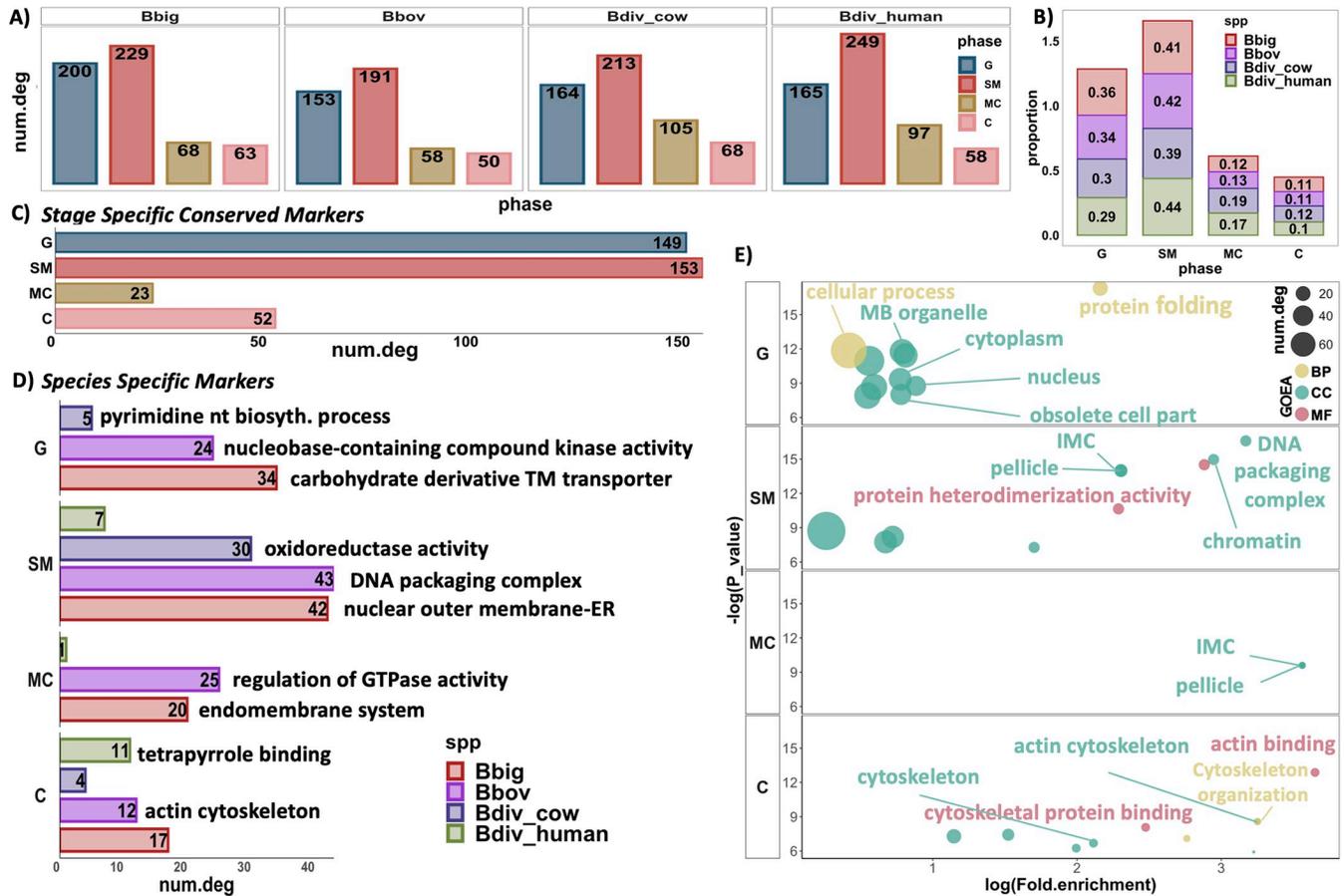


Fig 4. Differential expression and Gene Ontology enrichment analysis (GOEA). (A) Total number of inferred replication cycle differentially expressed genes (DEGs) in each species independent of other species. (B) Stack bar plot showing proportion of DEGs in each phase in all species. (C) Bar plot showing the total number of conserved DEGs of inferred replication cycle phases (rows) across all species. (D) Total number of DEGs of the indicated phase unique to the indicated species (colors). The most significant Gene Ontology (GO) term (Benjamini < 0.1) associated with the set of DEGs (each bar) is presented next to the bar. (E) Significant GO terms associated with conserved DEGs of the indicated phase (rows). Colors indicate the GO term category. Fold change > 2 and adjusted *p*-value < 0.01 were applied as cutoffs for all DEG analyses performed in this section. Note: All significant DEGs and GO hits can be found in S3 and S4 Tables. Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g004>

budding of daughter cells; thus, gene expression would need to occur prior to this process [56]. In *T. gondii*, expression of many of the orthologous genes peaks during the transition from S to M phase [23], which is consistent with the expression we observe here in *Babesia* spp. A major distinction between SM and MC inferred phase is the distinct upregulation of genes involved in biogenesis and segregation of the Golgi apparatus during SM phase—in mammalian cells this process is known to occur just prior to mitosis [57]. This is also in line with the previously observed timing of Golgi formation in *T. gondii*, where the Golgi undergoes elongation and segregation prior to the formation of the scaffolding of daughter cells (mitosis) [58]. Additionally, genes involved in vesicular transport and vesical formation (clathrin heavy chain, Bdiv_019640c; adaptin N terminal region family protein, Bdiv_009600c) were identified in SM phase. Finally, the genes identified as enriched in G phase are generally involved in organelle biogenesis including the apicoplast. In *P. falciparum*, apicoplast growth begins in the trophozoite phase prior to DNA replication [59]. Further, genes involved in protein folding, including several chaperones and heat shock proteins, are expressed in this stage.

Genes involved in DNA replication start expression in late G phase (Fig F in [S1 Text](#)), suggesting that *Babesia* species follow a “just-in-time” expression pattern, in this case expressing genes needed during S phase DNA replication during G phase. Further details of phase-based DEGs can be found in [S1 Text](#). These results show that the identified phases and the DEGs concur with known biology, although there is some degree of overlap between the phases, indicating the need for defining alternative transition points that more accurately reflect the progression of the cell cycle.

Species-specific differential gene expression analysis

In addition to conserved DEGs of replication cycle progression, we also investigated species-specific DEGs of each inferred state using GO term enrichment, focusing on the most highly enriched processes (Fig 4D). The topmost significant GO term for each set of DEGs is shown next to the bar plot in Fig 4D. For each species, a different common thread emerges throughout their respective replicative cycle. Many of the enriched processes involve the endomembrane system and membrane transport in *B. bigemina*. These include nucleotide (ATP) transport (BBBOND_0211990), known to be important in mitochondrial transport in related parasites [60,61], and several genes with predicted function in the endomembrane system (BBBOND_0309310, BBBOND_0401740, BBBOND_0307820, BBBOND_0102060, BBBOND_0210875, BBBOND_0210880). However, in *B. bovis*, an emphasis on various signaling pathways emerges, including kinase activity (nucleoside diphosphate kinase family protein, BBOV_III005290; adenylate kinase, BBOV_IV002930) and hydrolase activity (putative GTPase activating protein for Arf, BBOV_IV012060; GTPase activator protein, BBOV_IV007530). For the latter, both enriched genes work to activate GTPases, which are known to be important regulators in other systems [62,63]. Species-specific enrichment in *B. divergens* mainly occurs in metabolic pathways, including fatty acid metabolism/pyrimidine biosynthesis (bovine: cytidine diphosphate-diacylglycerol synthase, Bdiv_002810c; human: choline/ethanolamine kinase, Bdiv_020970) and oxidoreductase activity (bovine: Bdiv_019910, Bdiv_030660, Bdiv_040430c). Taken together, these results suggest varying levels of dependence on vesicular transport, signaling, and metabolic pathways among *Babesia* species. Further description of differences by phase can be found in [S1 Text](#). We also repeated the same analysis using a cutoff threshold of 1.5 for FC to test the robustness of the analysis (Fig G in [S1 Text](#)). As expected, the total number of identified DEGs increased, but the enrichment analysis was generally the same.

Progression of gene expression during the replication cycle

We sought to cluster and visualize the expression of DEGs during the replication cycle. DEGs of each phase were identified in each species independently (Fig 4A). We fitted smoothing B-splines to all gene curves and calculated the trends along the pseudo-time course. Fig 5 shows the heatmap of scaled expression of mean curves of the inferred replication cycle DEGs in each species. Genes are ordered by their peak expression time in each phase. Vertical lines show the mapped transition time points, while horizontal lines delineate the DEGs of each phase. The transition from C back to G consists of genes with “flat” peaks that cross the boundary. To better visualize this, we divided the G phase into late (G1 L) and early (G1 E), with G1 E directly proceeding the C phase. The heatmaps show that clusters of genes peak at similar times, with peak time expression gradually shifting through the replication cycle. Interestingly, there are visible points of transition in the peak expression time that seem to match well with cell division cycle phases, indicating that shifts in peak expression govern the progression of the cell division cycle. This “just-in-time” expression pattern profile is also seen in other parasite

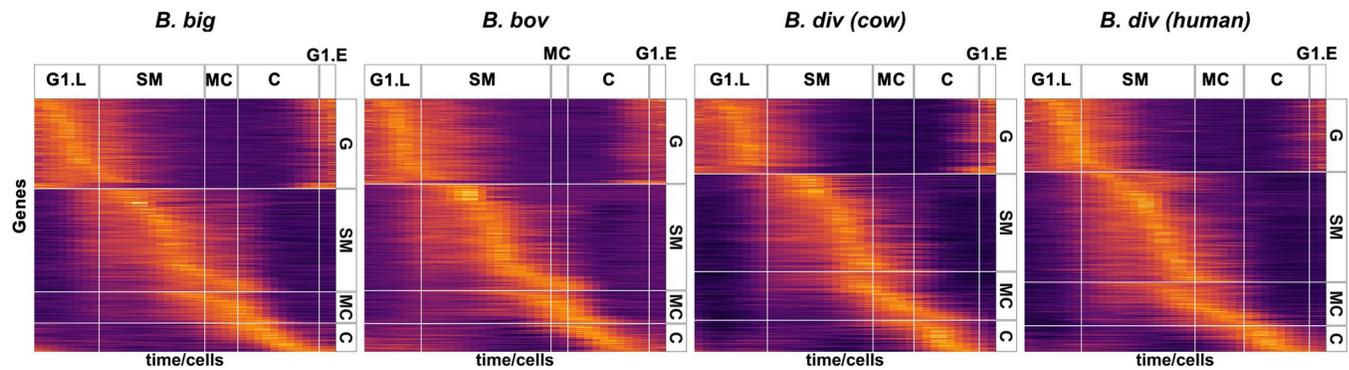


Fig 5. Transcription profile of replication cycle regulated genes: Heatmap of mean expression curves of differentially expressed genes of inferred replication cycle phases. Rows represent genes and columns represent pseudo-time-ordered cells. Horizontal facets group genes in the indicated phase. Vertical facets mark the transition time points. Analysis is done in each species independently (Fig 4A). Rows are ordered according to peak expression time. Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g005>

species [23,24,64–66], supporting the notion that replication cycle genes are organized into clusters of co-expressed/co-regulated genes that may be functionally related. A heatmap of 377 conserved replication cycle DEGs across all species (Fig 4C) ordered by their progression in *B. divergens* adapted to either human or bovine RBCs shows a more similar progression pattern compared to *B. bovis* or *B. bigemina* (Fig H in S1 Text), as also seen with the UMAP projection (Fig 1).

Comparative differential expression analysis of *B. divergens* in human versus bovine RBCs

To investigate the impact of the host cell on the transcriptome, we performed a differential expression analysis followed by GO term enrichment and identified DEGs that are uniquely associated with growth in different host cells. First, we compared the transcriptome of *B. divergens* in human versus bovine RBCs in each of the inferred phases. This analysis quantifies species-specific host cell impact on the transcriptome. Fig 6A (top) shows the expression of the top DEG in human (Bdiv_006490c) and bovine (Bdiv_040280) RBCs. Bdiv_006490c is annotated as an aspartyl protease (*asp3*) that shares homology with plasmepsin X in *P. falciparum*, which is essential for egress and invasion of the parasites [67,68]. Bdiv_040280 is an unannotated protein that contains a mago binding domain. Mago domain containing proteins have been shown to be involved in post-transcriptional regulation in other parasites [69]. Mago proteins have also been shown to be involved in splicing and trafficking of mRNA [70]. The bar plot in Fig 6A (bottom) shows the total number of upregulated genes (FC > 1.5 and adjusted *p*-value < 0.01) specific to *B. divergens* in human RBCs and *B. divergens* in bovine RBCs in each phase. In total, we identified 28 genes across the inferred replication cycle phases that were differentially expressed between *B. divergens* grown in different host RBCs (S3 Table). The majority of changes observed occurred in metabolic pathways including lipid metabolism and pyrimidine biosynthesis (S1 Text). These underscore differences in nutrient transport and metabolism based on resident host cell.

Second, to quantify the impact of the host cell in a species-independent manner, we merged the transcriptomes for all *Babesia* species isolated from bovine RBCs and compared the merged transcriptome with that of *B. divergens* in human RBCs for each of the replication cycle phases. For this analysis, to ensure that parasite-specific differences did not confound the

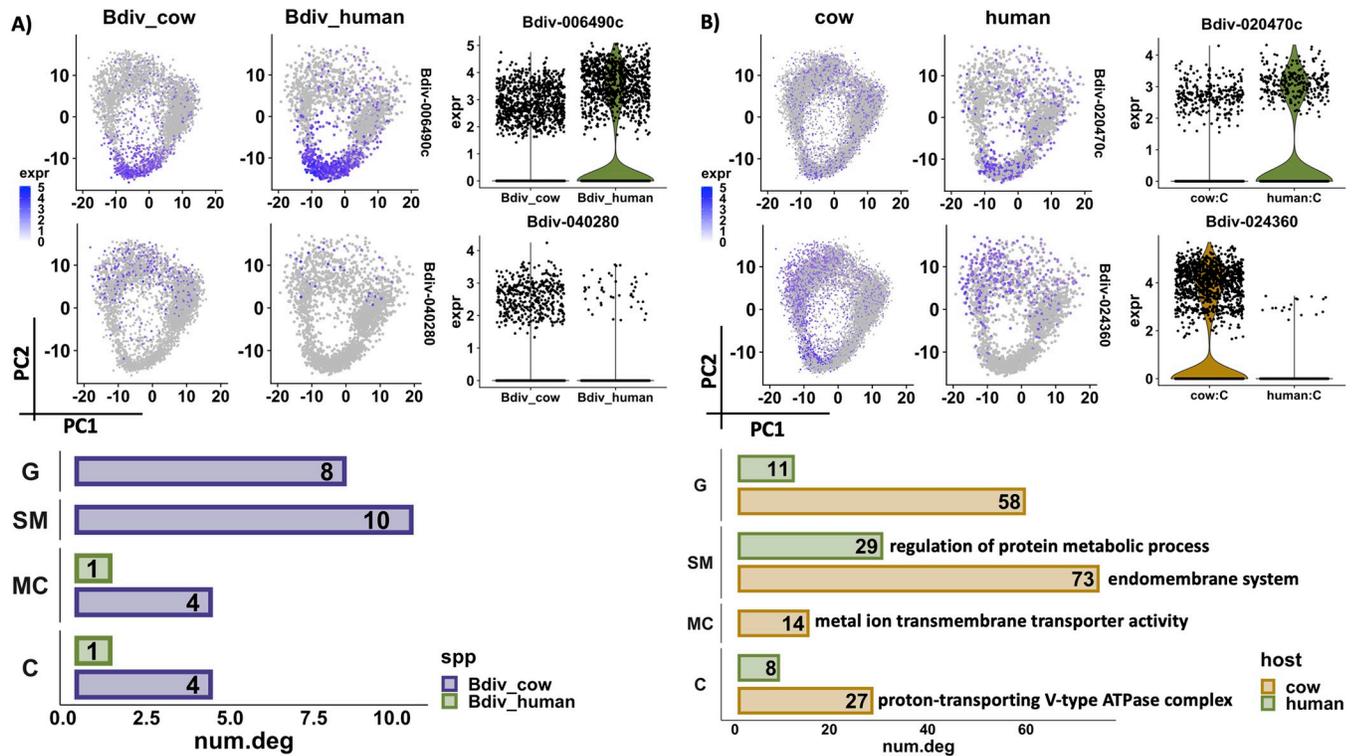


Fig 6. Host-specific differential expression and GO term enrichment analysis. (A) Top: Expression of top differentially expressed genes in *B. divergens* in bovine and human RBCs. Violin plots illustrate the distribution of expression of the top DEGs. Bottom: Bar plots showing the total number of DEGs in the indicated phase (rows) in each sample (colors). (B) Top: Expression of top DEGs in combined *B. bigemina*, *B. bovis*, and *B. divergens* in bovine RBCs versus *B. divergens* in human RBCs. Violin plots illustrate the distribution of expression of top DEGs. Bottom: Bar plots showing the total number of DEGs in the indicated phase (rows) in each sample (colors). To eliminate the confounding effect caused by species differences, any DEG between *B. divergens* and *B. bigemina* or between *B. divergens* and *B. bovis* in bovine RBCs was excluded from this analysis. The topmost significant GO term associated with each set of DEGs (Benjamini < 0.1) is presented next to the bar plot. DEG, differentially expressed gene; GO, Gene Ontology; PC, principal component; RBC, red blood cell. Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g006>

results, we excluded any gene that was differentially expressed between *B. divergens* in bovine RBCs and *B. bigemina* and *B. bovis* in bovine RBCs. Fig 6B (top) shows the expression of the top DEGs in human (Bdiv_020470c) and in bovine (Bdiv_024360) RBCs, projected on the PCA plot or shown using a violin plot to facilitate visualization of the expression distribution. The human DEG Bdiv_020470c is a cytochrome b5-like Heme/Steroid binding domain containing protein, while the bovine DEG Bdiv_024360 is a conserved hypothetical protein that contains a SNARE-associated Golgi protein domain as well as several transmembrane domains, which may be reflective of differences in membrane composition and nutrient availability between the host cells. The bar plot in Fig 6B (bottom) shows the total number of up-regulated genes in parasites cultured in each host per phase as well as the top significant GO terms. GO terms associated with genes enriched in human-adapted *B. divergens* are mainly involved in protein kinase signaling (Bdiv_035640) and protein metabolism (Bdiv_022680, Bdiv_002540) during the C and SM phases. These findings suggest differences in protein dynamics/degradation and in signaling pathways in the parasites that reside in human RBCs.

In contrast to the human-adapted parasite, in the bovine-adapted parasites (*B. bigemina*, *B. bovis*, *B. divergens*) the majority of enriched GO terms are related to the Golgi apparatus, endomembrane system, and vesicular formation/transport (i.e., Bdiv_000390c, Bdiv_005170,

Bdiv_009600c, Bdiv_018340, Bdiv_022710, Bdiv_028900c, Bdiv_032480), suggesting that vesicular transport and protein trafficking may be upregulated in parasites cultured in bovine RBCs. Further, there is an enrichment of genes involved in transport (i.e., Bdiv_010590, Bdiv_033310c). Taken together these results suggest that there are some differences in cellular processes of parasites propagated in bovine versus human RBCs that are driven by the host cell environment. Particularly, there appear to be consistent differences in protein metabolism, signaling, nutrient transport, and vesicular transport. A full list of DEGs and GO terms are available in [S3](#) and [S4](#) Tables and is described in [S1 Text](#).

Reconstructing the co-expression network

The presence of clusters of co-expressed genes ([Fig 5](#)) indicates that shared mechanisms may orchestrate the expression of gene clusters and the progression of the replication cycle. To shed light on the interaction between co-expressed genes, we assembled co-expression gene–gene interaction networks using a probabilistic graphical model [71] ([Materials and Methods](#)). A gene interaction network is a graph where nodes are genes and edges represent connection between the genes. The connected genes in the network represent genes that are potentially functionally related. Gene interaction networks are extensively used for a variety of applications, such as identifying clusters of functionally related genes or inferring function through association [72]. Highly connected genes in the network (i.e., “hub genes”) correspond to genes that may have an essential function. Disruption of highly connected hubs in interaction networks causes a major shift in the topology of the network [73]. Since the cell cycle is the major source of variation in our data, the hub genes may have a significant role in the progression of the cell cycle. We identified and ranked the top hub genes in each species and analyzed the overlap of their interactomes (i.e., genes connected to the hub-genes) across all species ([Fig 7A](#)). There are 10 genes at the intersection of hub genes in all 4 datasets. Of these 10 genes, 3 are annotated as surface antigens—41K blood-stage antigen precursor 41–3 (Bdiv_026840c), 12D3 antigen (Bdiv_020800), and 200 kDa antigen p200 (Bdiv_003210)—and there is also the highly conserved rhoptry-associated protein 1 (*rap-1*, Bdiv_025600). Rap-1 is known to be important in host cell invasion, an essential and conserved process in *Babesia* [34,74–76]. [Fig 7B](#) shows the pseudo-time expression of the 10 genes, clustered in 2 groups. [Fig 7C](#) shows the *B. divergens* (human) sub-network of the 10 genes at the intersection. The network appears to be modular, with most genes in the C or SM phase. Hub nodes in the interactome include *asf-1* (Bdiv_015780c), mitogen activated protein kinase (*mapk*, Bdiv_023270), and calcium-dependent protein kinase 4 (*cdpk4*, Bdiv_024410). ASF-1 has been shown to play an essential role in histone organization and progression through S phase in other systems [77,78]. The MAPK identified shares sequence homology with ERK7 kinase in *T. gondii* (TGGT1_233010), which has a role in the stability of the apical complex [79–81]. The hub gene *cdpk4* has been previously shown to be essential in egress [39]. Interestingly, several hypothetical proteins emerge at the intersection of hub genes, including Bdiv_011410c, Bdiv_024700, and Bdiv_028580c. By determining the interactome of these genes, some inferences about their cellular function can be made.

The hypothetical protein Bdiv_028580c shares the majority of its connections with genes in the SM phase (20 of 36 connections). Interestingly, it is connected to an aspartyl protease (*asp6*, Bdiv_022420c) that shares sequence homology with plasmepsin VII of *P. falciparum*. In *P. falciparum*, this aspartyl protease plays an essential role in the invasion of the mosquito mid-gut in the ookinete stage [82]. In addition to this aspartyl protease, this hypothetical protein is also connected to Bdiv_010620, which is orthologous to the secreted ookinete protein *psop13* in *B. microti*, and the orthologous gene in *Plasmodium* (PF3D7_0518800) is known to play an

associated protein 50 (GAP50, PF3D7_0918000) in *P. falciparum*, which plays a role in the organization of the inner membrane complex [88]. Indeed, similar cytoskeletal components are required for invasion in *T. gondii* [89]. Furthermore, Bdiv_024700 is connected to an aspartyl protease (*asp2*, Bdiv_023140c) similar to plasmepsin IX and X in *P. falciparum*, and has a known role in invasion [39]. Together, these data suggest a role in the invasion process for Bdiv_024700.

Finally, looking at the interactome of Bdiv_011410c, most gene connections occur in the inferred C phase. This hypothetical protein is connected to many genes involved in signal transduction, most notably 2 calcium-dependent protein kinases (*cdpk4*, Bdiv_024410; protein kinase domain containing protein, Bdiv_033990c) and the aspartyl protease *asp3* (Bdiv_006490c), suggesting a possible role in signaling processes that control egress [39]. Taken together, the network analysis identifies genes with essential function in the cell cycle as hubs of the network and enables inference of the function of unknown genes through analysis of the function of neighboring genes.

Expression profiles of functionally related gene families

To examine the expression changes in functionally related genes, we performed a clustering analysis on a list of curated gene families: (1) putative transcription factors (TFs) identified by the presence of a DNA binding domain in the sequence, (2) putative AP2 TFs identified by orthology with *P. falciparum*, (3) putative epigenetic factors similarly identified by orthology with *P. falciparum*, and (4) variable erythrocyte surface antigen (VESA) genes (S5 Table). The reconstructed expression curves of each gene family in each species were clustered into 4 groups, selected to match the 4 inferred phases of the cell division cycle. Fig 8A shows the clustering of epigenetic markers with the cyclic expression profile in all species. Four clear clusters emerge, with many genes peaking at similar times across all species, although the expression patterns of some genes are species-specific. Many of the core histone proteins are co-expressed in cluster 1, showing timing of expression mainly during the inferred SM phase. This pattern of coordinated expression of core histone proteins has been observed in related parasites [90]. There is also a coordinated expression in the MC phase in cluster 3, including epigenetic factors involved in chromatin organization (Bdiv_023810, Bdiv_024470c, Bdiv_036910c) and histone modification (Bdiv_023060, Bdiv_034310c). Cluster 2 shows peak gene expression of 3 epigenetic genes that occur in C phase, a histone demethylase (Bdiv_012930), histone acetyltransferase (Bdiv_034310c), and a zinc-finger domain containing protein (Bdiv_016880c). Cluster 4 shows broad expression over the replication cycle, and the genes that compose this cluster do not fall into similar gene families. Fig 8B shows a heatmap representing the presence of each cyclical epigenetic marker in each cluster and species. Absent genes in a species either are not cyclically expressed or belong to a different cluster. We next identified the inferred interactome of these markers using the co-expression networks. Fig 8C shows the interaction network of epigenetic markers in cluster 1 for *B. divergens*. Interestingly, many of these genes appear as hubs in the network including *asf-1*, which is the top hub gene in *B. divergens* in bovine blood and among the top in *B. divergens* human blood and appears in the inferred interactome of 10 conserved hubs (Fig 7C). ASF-1 shares connections with histone H2A (Bdiv_011310c), histone 2B (Bdiv_011450c), and chromatin assembly factor 1 (*caf-1*) subunit C (Bdiv_013610), each of these occurring in cluster 1 (Fig 8A). Identifying this interaction provides support for the validity of the networks generated from these data in *Babesia* spp.

The same analysis was performed for TFs with cyclical expression profiles, including AP2 domain containing proteins, identified through reciprocal blast with identified TFs in *P. falciparum* [91] (Fig J in S1 Text). We identified 4 clusters of co-expressed genes, which map to the

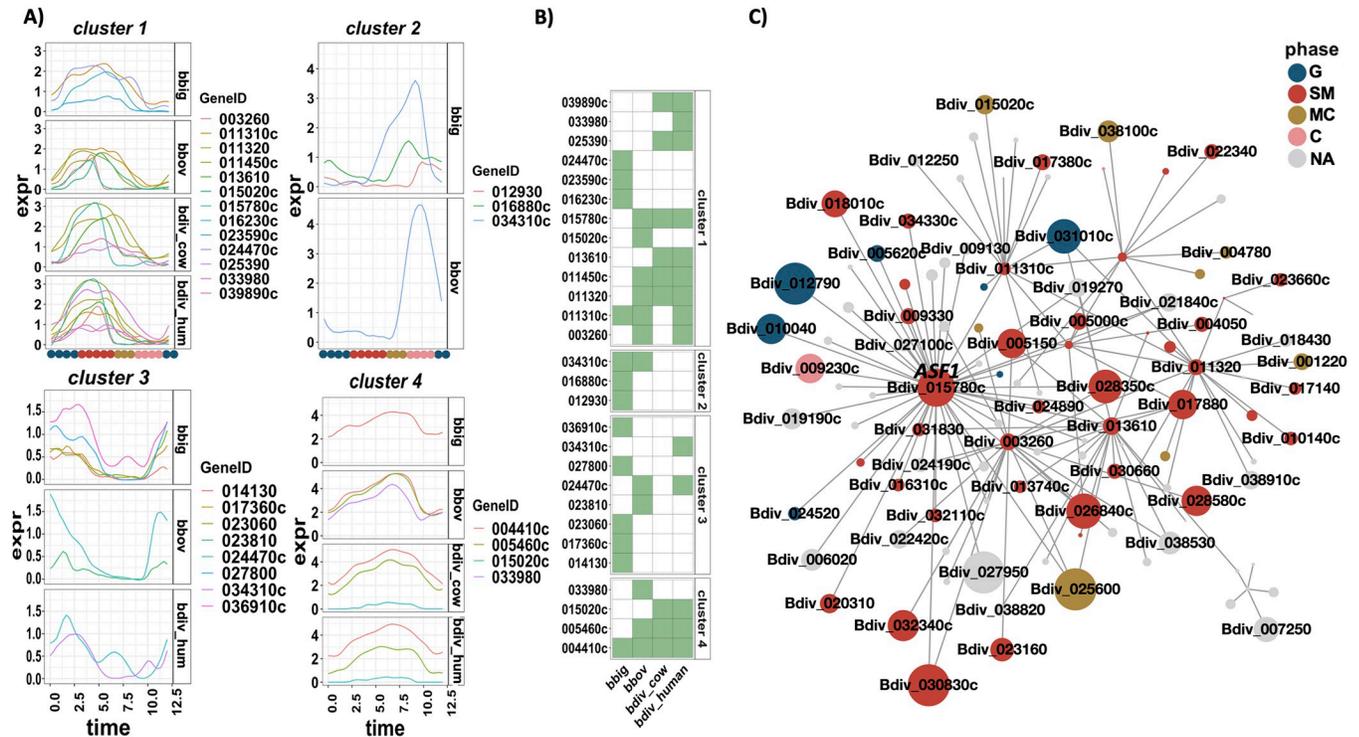


Fig 8. Expression profile of epigenetic markers with a cyclic pattern of expression. (A) The expression curves of epigenetic marker genes clustered into 4 groups according to their expression similarity, split by species. (B) Presence (green) or absence (white) of the gene (rows) in the indicated sample (column); absence indicates that the gene is either not cyclically expressed or belongs to a different cluster (033980, 024470c, 015020c, and 034310c). (C) Interaction co-expression network of epigenetic marker genes in *B. divergens* (human) in cluster 1 (9 total). Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g008>

4 inferred replication cycle phases, with each gene identified by its ID in *B. divergens* (Fig J in S1 Text). Several AP2 domain containing proteins appear to show a cyclical expression profile in 1 species. For example, Bdiv_037050 (AP2-G3) peaks at G phase in *B. divergens*, and Bdiv_010110c peaks during SM in *B. bigemina*. In contrast, the AP2 domain containing proteins Bdiv_000800c and Bdiv_024900c are expressed at the same time in all species. There appears to be a highly conserved function in the inferred SM phase for Bdiv_024900c—an AP2 domain containing protein with sequence homology to PF3D7_1239200. The expression of this gene is earlier than in the expression profile of *P. falciparum*, where it peaks in the later stages of schizogony [92]. The most similar ortholog of this gene in *T. gondii* is *AP2VIIb-3* (TGGT1_255220), which is implicated in the replication cycle progression into S phase, which is more in line with the observed expression across *Babesia* species [93,94]. A list of gene cluster IDs is available in S6 Table.

Motif analysis of the interactome of TFs and AP2

To examine whether the interactome of the TF and AP2 family of regulators is transcriptionally co-regulated, we performed a motif search analysis using MEME suite [95] on the promoter genes in the interactome. For each AP2 (and TF), we identified the genes interacting with the AP2 as determined by the assembled co-expression network. Next, we took the union of all genes across the 3 species to assemble a shared interactome for each AP2. The promoter sequences of the genes in the interactome of each AP2 were extracted for each species independently and inputted into the MEME algorithm. The analysis identified a significant motif

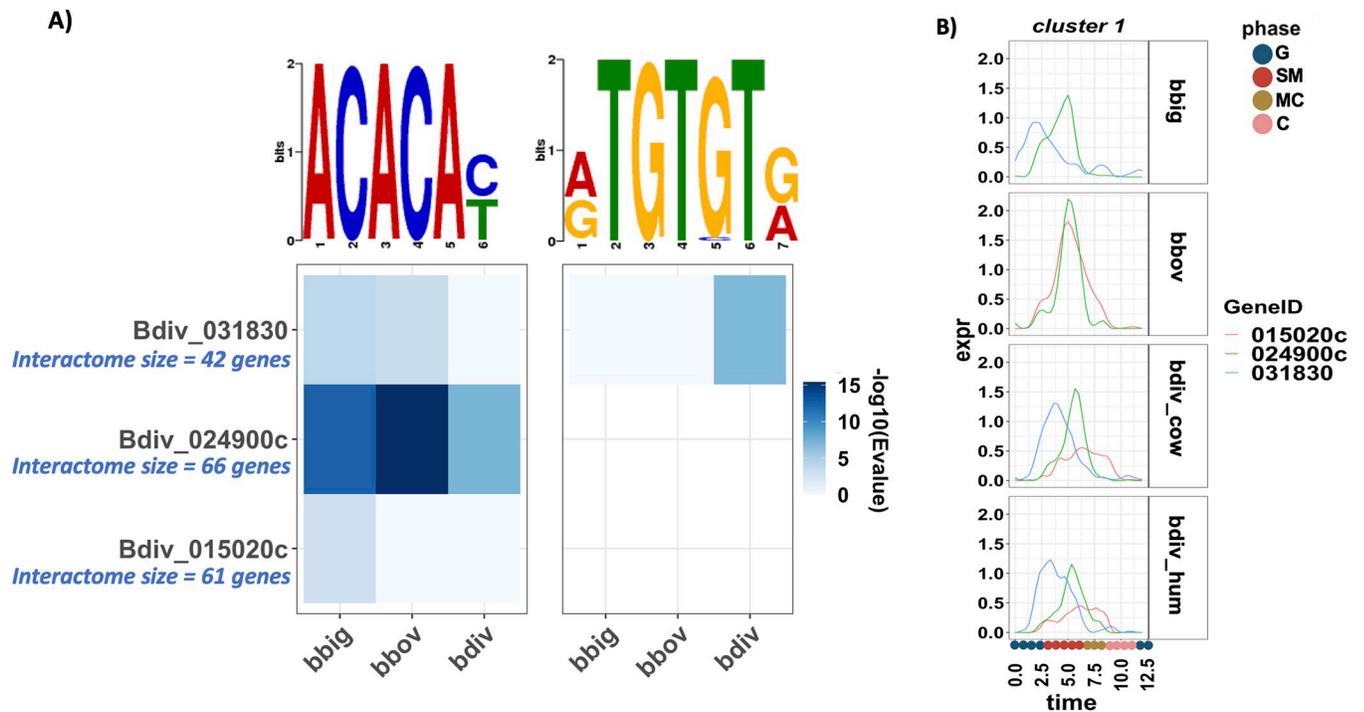


Fig 9. Motif search analysis identified a significant motif in the promoter of the interactome of 3 AP2s. (A) The heatmap shows the significance ($-\log_{10}(\text{E-value})$) of identified motifs, with rows corresponding to the interactome of the indicated AP2 and columns corresponding to the species. (B) Expression curve of the 3 AP2s. Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g009>

“ACACA” in the promoter of the interactome of 3 of the AP2s: Bdiv_015020c, Bdiv_024900c, and Bdiv_031830 (Fig 9A). These AP2s are orthologous to *P. falciparum* AP2s PF3D7_0604100 (SIP2), PF3D7_1239200 (an unstudied AP2), and PF3D7_0802100 (AP2-I), respectively. Interestingly, the motif “ACACA” has previously been reported to play a role in DNA replication in *Plasmodium* spp. [96,97]. Moreover, analysis of ATAC-seq regions in *P. falciparum* has identified and associated the same motif with AP2-I (PF3D7_0802100) [98]. Fig 9B shows the peak time expression of the 3 AP2s in all species. The AP2s mostly cluster together and seem to peak at S/M phase.

Interactive web app

To facilitate usage, we developed a user-friendly interactive web app using web dashboard. The app provides functionality to explore and visualize gene expression during the IDC across thousands of asynchronously dividing single cells projected on PCA or UMAP coordinates. Users can examine the timing of expression using pseudo-time analysis and perform comparative transcriptomic analysis across the *Babesia* species. Moreover, users can generate co-expression networks and interactively visualize and explore the inferred interactome of genes. Fig 10 illustrates the main functions implemented in the app. The app can be accessed at <https://umbibio.math.umb.edu/babesiase/>. The source code for the app is available on GitHub at https://github.com/umbibio/babesia_dash_app.

Discussion

In this work, we present the first single-cell sequencing data in asynchronously replicating *Babesia* parasites and characterize the progression of the replication cycle using newly

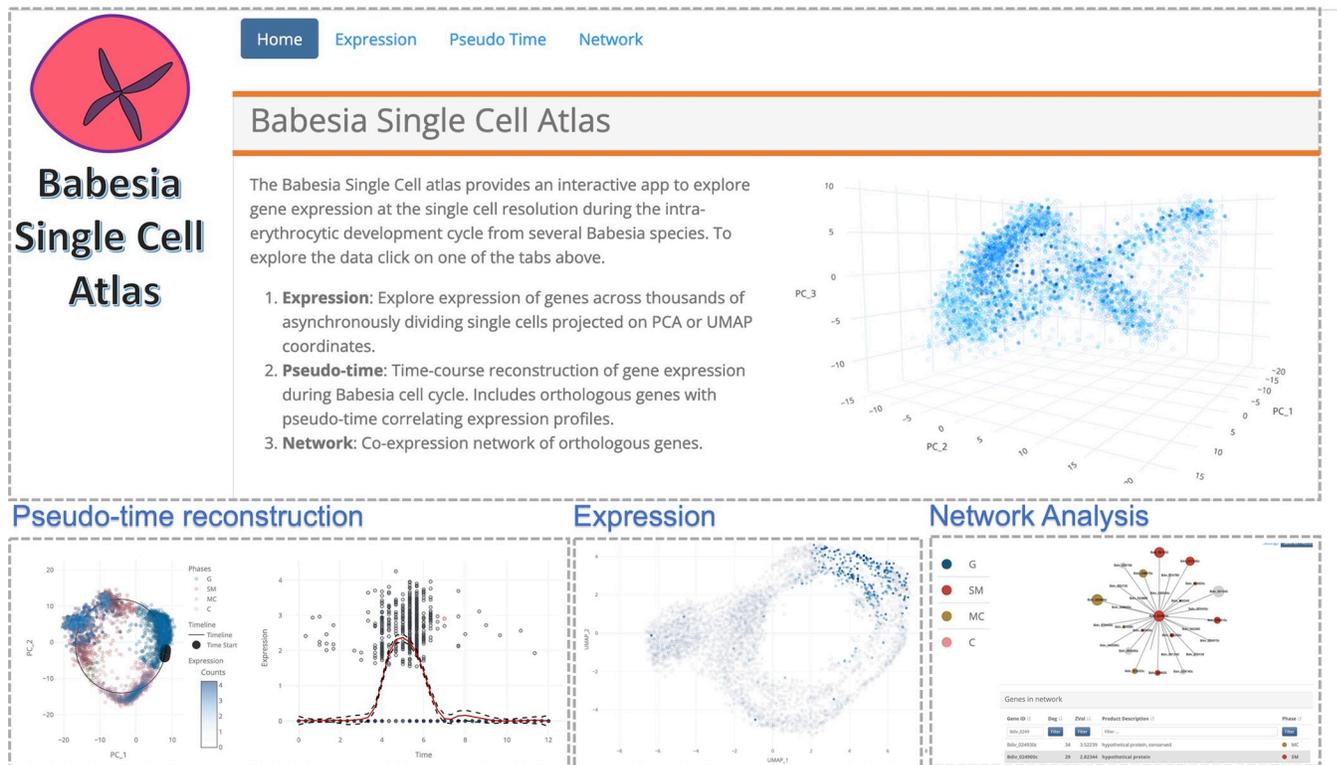


Fig 10. Babesia single-cell atlas screenshots. The app provides functionality for pseudo-time analysis, visualizing expression, network analysis, and comparative transcriptomics. URL: <https://umbibio.math.umb.edu/babesiasc/>. Data and code for generating the figure are available at <https://github.com/umbibio/scBabesiaAtlases>.

<https://doi.org/10.1371/journal.pbio.3001816.g010>

developed computational approaches. The replication cycle in *Babesia* spp. likely relies heavily on various signaling pathways (reviewed in [28]). In the asexual replicative cycle of *B. bovis* and *B. bigemina* and many other species, after invasion, the vast majority of parasites will grow, mature, and divide once to form 2 daughter cells prior to egress; for other species, parasites will divide into 4 daughter cells [99]. Data on the division cycle for *B. bovis* in asynchronous culture suggests that cell division is complete in approximately 5 h, yet this does not describe the time from invasion to egress [100]. However, in a more recent study on synchronized *B. bovis*, this process was approximately 12 h [38]. In contrast, *B. divergens* has a much more complex replicative cycle [37]. Indeed, conflicting literature exists suggesting that the replication cycle varies from 4 to 12 h; however, in most cases it appears the minimal time for division is between 4 and 5 h [37,101,102]. Of these studies, only one was conducted on synchronous parasites [37]. However, data also show that the time for the majority of parasites in *B. divergens* to transition from single parasites to paired piriforms is between 10 and 12 h [37]. Subsequent division cycles are possible in *B. divergens*, and the timing of these cycles can vary between 9 and 14 h [37]. In all studies of the dynamics of *B. divergens* division, there is a significant range of time for a single division, highlighting the difficulty in generating an exact measurement [37,101,102]. This variability leads to difficulty in synchronizing *Babesia* parasites, which currently relies on mechanical release of parasites from RBCs using filtration [37,38], and we have recently shown that parasites may be egress competent at various times in their intraerythrocytic development [39]. Unfortunately, no such detailed data on the division time in *B. bigemina* exist; however, the replication rate in culture is similar to that in *B. bovis*, and

the 2 parasites follow a similar pattern of dividing from single to double parasites prior to egressing. Based on these data, we opted to set the window of time for the replication cycle at 12 h for the 3 parasites [39]. Of note, in our analysis we did not distinguish separate clusters of expression for secondary division cycles in *B. divergens*, suggesting that the core replication process is conserved and that multiple rounds of replication within the same RBC do not require separate gene regulators.

Taking advantage of the unique geometry of replicating parasites, we developed a pseudo-time analysis and used the synchronous bulk RNA-seq data in *B. divergens* to calibrate the progression of time in single-cell data. This technique allows us to generate pseudo-synchronized time-course data at a fine resolution and reconstruct the expression waves of genes during the replication cycle. Analysis of the data shows a high degree of agreement between the bulk and single-cell data, demonstrating the ability of single-cell measurements to match (and overcome some of the limitations of) synchronized time-course measurements.

The limitations, challenges, and advantages of scRNA-seq have been extensively reviewed, including comparisons of the most robust tools, understanding dropouts, and discussion of the ability to understand genes with low expression [103–110]. While scRNA-seq offers extremely fine granularity of cell states, the resolution with which differential gene expression can be detected varies significantly [109]. Indeed, in these *Babesia* spp. datasets, we can clearly observe divergence between bulk RNA-seq and scRNA-seq in the pattern of expression over time in genes with low expression. For example, protein kinase G (*PKG*, *Bdiv_020500*) was recently identified as an essential gene in egress using synchronized bulk RNA-seq and reverse genetics [39]; however, this gene has low expression, and the single-cell data have difficulty detecting this gene (Fig K in *S1 Text*). This highlights a key challenge to all methods of differential gene expression: There is often not enough power to confidently characterize genes with low expression [109]. In mammalian cells, it is recommended that one sequence to a depth of 20,000 reads per cell. Based on the minimal *Babesia* genome size and decreased number of predicted coding sequences (between 3,700 and 4,000) in relation to mammalian cells (>12,000), we estimated that an average of approximately 6,000 reads per cell should sufficiently capture the expression profiles (*S1 Table*). We acknowledge that this depth is lower than that of scRNA-seq experiments performed in *Plasmodium* and *Toxoplasma*; however, *Babesia* has both a smaller genome and fewer coding sequences predicted than these related organisms (3,700 compared to approximately 5,200 and 8,100 for *Plasmodium* and *Toxoplasma*, respectively). We also note that the read depths do vary between samples (correlating with genome size) and may cause some limitations in the downstream data analysis of genes with low expression. These limitations are important to consider when attempting to characterize cell populations, especially those that may represent rare cell types.

However, scRNA-seq also provides major advantages over bulk sequencing techniques. In most synchronized bulk RNA-seq time-course studies, measurements are limited to a few discrete time points (typically <12 points), whereas in sequencing heterogeneous single-cell populations, the entire trajectory of expression can be captured, representing a continuum of time. This is advantageous in capturing subtle variations in curvature, which allow more precise mapping of peak expression as well as clustering genes by expression similarities. Because of the ability to capture time using asynchronous populations, the difficulties of tightly synchronizing parasites are completely avoided in single-cell studies. Indeed, heterogeneity is an advantage in scRNA-seq that allows capturing and grouping cells in similar states at single-cell resolution, and scRNA-seq can capture and characterize small cell populations not distinguishable at the bulk level. In this study, we highlight the benefit of combining the 2 techniques to leverage the detection power of bulk RNA-seq and the fine resolution of single-cell sequencing.

It should be noted that annotations of *Babesia* genomes are not extensive and do not include the untranslated regions (UTRs). Moreover, the gaps between the genes are shorter compared to *T. gondii* and *Plasmodium* spp. (S2 Table). However, there is generally a clear boundary between mapped reads to the genome, and intergenic regions (including unannotated UTRs) are excluded in calculating transcript abundance. While it would be optimal to include UTRs, the reads mapped to the exons provide a good approximation for transcript abundance, and there is no bias toward any specific gene when calculating relative abundance and DEGs.

Using a comparative approach, we utilized data from better studied *T. gondii*—where the replicative cycle phases have been characterized using single-cell sequencing and DNA content [50]—to map the replication cycle phases in *Babesia* spp. and identify the DEGs of “inferred” replication cycle phases. The significant overlap between peak expression times of *T. gondii*-based markers indicates that progression of cell division in *Babesia* spp. may not fit well into the canonical model of cell cycle progression (Fig D in S1 Text). Mapping transition points in the *Babesia* replication cycle can also be achieved using automatic clustering approaches. We initially used this approach and identified 5–6 phases, which mostly agree with *T. gondii* inferred phases, with little impact on GOEA results (Fig C in S1 Text). The progression of gene expression (Fig 5) shows a defined transition point between G and S/M phases, and a gradual shift through S/M/C with overlapping transition boundaries. More data, (e.g., time-course DNA content) is needed to map the phases more accurately. However, GO indicates that the inferred phases of canonical *T. gondii*-based phases agree with the known biology of the replication cycle.

The mapping of replication cycle phases allowed us to identify phase-regulated genes and perform comparative studies between the *Babesia* species. The analysis revealed conserved and species-specific genes that delineate the inferred states of the replication cycle. Intriguingly, although several of the genes have species-specific expression profiles, the total number of DEGs in each phase is comparable, and the progression of gene expression during the replication cycle shows a similar pattern in all species. We also investigated the impact of the host species on the transcriptome. Of the parasites tested, *B. divergens* is able to grow in both human and bovine RBCs, while *B. bigemina* and *B. bovis* can only be propagated in bovine RBCs. The genes that were identified with differential gene expression in *B. divergens* between bovine and human RBCs indicated changes in the expression of genes involved in transcriptional regulation, as well as those potentially involved in invasion and egress. We also identified several differentially regulated aspartyl proteases known to play roles in invasion and egress in various replication cycle phases [39]. The differential expression of these genes could be due to differences in host cell receptors, parasite ligands, or host cell membrane composition. A common thread through the DEGs between human- and bovine-adapted parasites is the presence of changes in protein metabolism, signaling, and vesicular transport. Despite both being mammalian hosts, bovine and human RBCs differ in their composition, size, and deformability [111–114]. One striking difference between bovine and human RBCs is the lipid composition of the cell membrane. The composition of bovine RBCs is significantly divergent from most other mammals in that they have low to absent levels of phosphatidylcholine, while having high sphingomyelin levels [112]. This stark difference in lipid composition of the host cell membrane may be a driver of the differential expression; indeed we did observe upregulation of 2 genes involved in lipid metabolism in parasites resident in human RBCs (Bdiv_036170c, Bdiv_025380c). Lipid composition can also affect signal transduction [115], which could underlie the differences we observed in the expression of genes involved in protein kinase activity. Future experiments comparing these processes between bovine- and human-adapted parasites are warranted based on the differences observed here.

Interestingly, the major unifying difference in bovine parasites is the emphasis on vesicular transport and the endomembrane system. One hypothesis from this observation is that parasites grown in bovine RBCs require increased trafficking to and from the membrane, both to export proteins and scavenge nutrients. The ability of *B. bovis* to cause disease pathology by adhering to the vascular endothelium via altering the RBC membrane surface supports the idea that increased protein export may occur in bovine parasites [116–120]. Recently, expression of a multigene family of multi-transmembrane integral membrane proteins (*mtm*) was identified in the proteome of the infected RBC membrane, in addition to several other multigene families (*ves1*, *smorf*, *tpr*-related), showing this protein is expressed and exported by the parasite [121]. These processes in *B. bovis* could account for an increased use of the endomembrane system. However, this cannot fully explain the observed enrichment across the bovine-adapted parasites, as *B. bigemina* and *B. divergens* do not sequester. It may still be the case that parasites in bovine RBCs more dramatically alter the host cell. Indeed, when splenectomized calves were infected with stabilates of bovine-derived strains of *B. divergens*, an increase in the mean corpuscular volume of the host cell was observed in relation to infection by human-derived patient isolates (stabilates), suggesting changes in the membrane architecture. This suggests there is a parasite-specific effect on bovine host cells depending on which host (human versus cow) the parasite was originally isolated from [122]. Unfortunately, no such studies investigating protein export or RBC architecture exist as of yet in *B. bigemina* [120]. However, the evidence from *B. bovis* and *B. divergens*, combined with the pattern observed in this study, presents an intriguing possibility that protein export is affected by resident host cell.

Our work provides to our knowledge the first comparative single-cell transcriptomic study across 3 *Babesia* species. To date, comparative genomic studies have focused on gene annotation, understanding variant gene expression, and elucidation of virulence determinants [17,18,20,29,31–33,35,36,123]. In the context of this study, it is worth noting that the organization of the genomes of *B. bovis* and *B. divergens* share similarity [18]. Additionally, *B. bigemina* contains extensive duplications of certain gene families, leading to an increased genome size [32]. The genome sequences would suggest that similarities exist in general cellular development, and variations arise due to host evasion pathways and differential host tropism. Several comparative studies have sought to understand the expression and structure of VESAs, both within and across species [32,36,123]. Consistent with these studies, we also observed a sporadic expression pattern of VESA genes, with species-specific differences (Fig L in S1 Text).

We also utilized our dataset to examine the expression profile of other gene families, including, AP2s, TFs, and epigenetic markers [92,124]. The analysis revealed distinct clusters of genes with similar peak expression time across species, indicating that shared regulatory mechanisms may be orchestrating the progression of the replication cycle. Of note, we identified that several core histones are co-expressed, which is also observed in related parasites [91]. Interestingly, the timing of expression of these core histone proteins differs from that in related parasites: For example, histone H3 expression peaks in early schizonts towards the end of DNA replication in *P. falciparum* [25], while in the 3 *Babesia* species the expression peaks around 3.75 h, which is at the beginning of S phase. This difference is likely indicative of the different modes of division of the parasites (schizogony versus binary fission). We were also able to provide strong evidence for the utility of the interactome networks by showing the connection of *asf-1* to histone expression (Fig 8C). Interestingly, *asf-1* appeared throughout our analyses, suggesting an important role of the gene in *Babesia* asexual cycle development. Future studies disrupting *asf-1* in *Babesia* would provide insight into the nature of the regulation of chromatin formation, as well as reveal any novel functions for the gene in the parasite. These interactomes provide the foundation for future perturbation experiments to understand the directionality of regulatory interactions. Additionally, we were able to identify patterns of

expression of AP2 domain containing proteins that were dependent on replication cycle phase (Fig J in [S1 Text](#)) and were able to identify a conserved motif for a set of these TFs (Fig 9B).

Finally, to facilitate wider use, we present a web dashboard for interactive exploration of the data. The app provides functionality for analysis of the expression of single genes, comparative analysis of expression profiles and pseudo-time curves across all species, and co-expression network analysis of genes. The app is hosted at <https://umbibio.math.umb.edu/babesiasc/>. This resource will allow for novel studies to expand upon and add to the analyses of these rich transcriptomic datasets using the interactive interface, without the need for expertise in computational methods. The work described here lays the foundation for numerous functional studies to elucidate many facets of *Babesia* biology.

Materials and methods

Parasite culture

The *B. bovis* strain MO7 and the *B. bigemina* strain J29, provided by David Allred of the University of Florida, were maintained in purified bovine RBCs (hemostat) hematocrit in RPMI-1640 medium supplemented with 25 mM HEPES, 11.50 mg/l hypoxanthine, 2.42 mM sodium bicarbonate, and 4.31 mg/ml AlbuMAX II (Invitrogen). Before addition of AlbuMAX II and sodium bicarbonate, we adjusted the pH of the medium to 6.75. *B. divergens* strain Rouen 1987, kindly provided by Kirk Deitsch and Laura Kirkman (Weill Cornell Medical College), was maintained under the same conditions in purified white male O+ human RBCs (Research Blood Components). All cultures were maintained at 37°C in a hypoxic environment (1% O₂, 5% CO₂). Clonal lines of parasites were used for all selections and were derived from the provided strains via limiting dilution—these will be referenced as BdC9 (*B. divergens*), BigE1 (*B. bigemina*), and BOV2C (*B. bovis*).

Single-cell protocol

B. divergens, *B. bovis*, and *B. bigemina* were grown to >15% parasitemia. Health of parasites was assessed by thin blood smear. Parasites were collected and pelleted, and washed with warm 1× PBS, followed by a final wash with warm 0.4% BSA in 1× PBS. To ensure loading of the correct number of parasites, parasitemia was counted on stained thin blood smears (2,000 total cells), and RBCs were counted using a hemocytometer. These values were used to calculate the number of infected cells to be loaded into the Chromium Chip B. We aimed for recovery of 10,000 infected cells, thus loaded 16,500 infected cells in bulk culture. Cell suspensions were loaded into individual wells on the Chromium Chip B. After gel bead-in-emulsion (GEM) generation, single-cell libraries were processed according to the 10× Genomics Chromium 3' v2 User Guide protocol, using 13 amplification cycles for cDNA amplification, and 14 cycles in library construction. Libraries were subsequently sequenced on the Illumina Next-Seq platform following the 10× Genomics specifications, aiming for a minimum of 6,000 reads per cell.

RNA-seq alignment

The reference genomes and annotation files of *Babesia* spp. (release 54) were downloaded from PiroplasmaDB (<https://piroplasmadb.org/>). Custom references were generated using the 10× Genomics Cell Ranger (version 6.0.0) pipeline (cellranger mkref), and raw fastq files were aligned to the genome cellranger count with default parameters.

scRNA-seq data processing

All data analysis was performed in R (version 4.1.1). The R Seurat package version 3 [53] was utilized to process the count data. Seurat objects were created for each count independently. Cells and genes with low counts were filtered out from the analysis using the Seurat function `CreateSeuratObject` with parameters `min.cells = 10`, and `min.features = 100`. Expression data were normalized using the Seurat functions `FindVariableFeatures` (with parameter `nfeatures = 3,000`) and `ScaleData`. Dimensionality reduction was performed using PCA and UMAP as implemented in the Seurat functions `RunPCA`, and `FindNeighbors`, with parameters `dims = 1:10` and `reduction = pca`. Clustering analysis was performed using the *k*-nearest neighbors algorithm using Seurat function `FindClusters` with parameter `res = 0.2`. Datasets were down-sampled to include 800 cells per cluster. Orthologous genes in all 3 species were used to construct Seurat objects with same genes using *B. divergens* gene IDs. Datasets were then integrated using Seurat's `merge` and `IntegrateData` functions. The *B. divergens* sample in human host RBCs was used as the “reference” dataset for integration (Seurat `FindIntegrationAnchors` function).

Pseudo-time analysis

Pseudo-time analysis was performed in 3 steps. First, an ellipsoid was fitted to the first 2 PCA coordinates in each dataset using the `Ellipsoidfit` function from the `MyEllipsoidfit` R package (<https://github.com/MarkusLoew/MyEllipsoidfit>). Next, the elliptic fit was used as prior to fit a principal curve to each dataset [54]. The function `principal_curve` from the R package `princurve` was used to fit the principal curves. The parameter `smoother = “periodic_lowess”` was set to enforce closed curves. Data were then orthogonally projected onto the principal curves and ordered to generate pseudo-time. The pseudo-time curve was mapped to interval [0,12 h] to mimic the *Babesia* spp. replication cycle. The pseudo-time interval was partitioned into 20-min bins, and cells that fell within the same bin were treated as synchronized replicates. Next, we calculated the correlation of gene expression with pseudo-time using a GAM and filtered out genes that did not correlate with the pseudo-time (FDR-adjusted *p*-value < 0.05). The R function `gam` from the package `gam` was used for this analysis. This resulted in a time-course expression matrix with dimensions $n \times \prod_{k=1}^N n_k$ with *n* representing the total number of genes, *N* representing the total number of time bins (36 intervals, each 20 min), and *n_k* representing the total number of cells mapping to the time bin *k*.

Fitting gene curves

We fitted the expression curves with smoothing splines using R's “`smooth.spline`” function with the smoothing parameter set by cross-validation. The fitted splines were then sampled at regular time points to interpolate the curves.

Alignment with bulk RNA-seq data

Time-course bulk RNA-seq data from synchronized *B. divergens* parasites was processed as previously described [37]. The bulk time-course data consisted of 7 measurements of synchronized *B. divergens* parasites over the 12-h period spanning the replication cycle. There was a total of 2 biological replicates and 2 technical replicates. Technical replicates were merged for this analysis. Smoothing splines were fitted to the common genes between the bulk time-course data and single-cell pseudo-time-course data. The smoothing spline fits to the bulk data were sampled to 36 points, matching the total number of points in the single-cell gene curve data. Cross-correlation between corresponding genes in bulk and *B. divergens* single-cell

datasets was calculated by

$$g_b \star g_{sc}(t) = \sum_{\tau=0}^N g_b(t) g_{sc}((\tau + t)_{\text{mod } N})$$

where $g_b(t)$ and $g_{sc}(t)$ are gene curves in the bulk and single-cell data, respectively, and N is the total number of sample points in the gene curves. The lag time maximizing the cross-correlation was then calculated for each gene, and the distribution of lag times across all genes was examined to identify a single optimal lag time for all genes. Single-cell gene curves in all single-cell datasets were shifted by the optimal lag time to adjust the start time.

Inferred replication cycle phases

T. gondii scRNA-seq data were obtained from the single-cell atlas of *T. gondii* [50], where replication cycle phases were determined using DNA content and computational analysis. Markers of each phase were determined by performing differential expression analysis using the Seurat R function `FindAllMarkers` with parameters `only.pos = TRUE`, `min.pct = 0`. Significance was determined using `FC > 2` and adjusted p -value < 0.01 . The top 20 markers of each phase were then used and mapped to their *Babesia* spp. orthologs. The timing of peak expression of each marker was calculated by examining the local maxima of the fitted pseudo-time gene expression curves in each species. Transition time points between phases were then determined by examining the quantiles of the peak time distributions, and adjusted by visual inspection of the overlap of distributions.

Differential expression analysis

DEGs of the inferred replication cycle phases were identified as follows. Conserved DEGs of each phase across species were identified using the Seurat function `FindConservedMarkers` with default parameters. DEGs of each phase were also calculated in each species independently using the `FindAllMarkers` function from the Seurat R package with parameters `only.pos = TRUE`. DEGs of each phase unique to a specific species were determined using the same function and by setting an appropriate contrast in cell identities (e.g., *B. bigemina* G phase versus *B. bovis*, *B. divergens* (bovine), and *B. divergens* (human) G phase). For these analyses, `FC > 2` and adjusted p -value < 0.01 were used to determine significance. For host-cell-specific differences, differential expression analysis was performed between *B. divergens* in human host and *B. divergens* in bovine host as well as between *B. divergens* in human host and merged *B. bigemina*, *B. bovis*, and *B. divergens* in bovine host in a phased-matched specific manner. For these analyses, `FC > 1.5` and adjusted p -value < 0.01 were used to determine significance. Genes differentially expressed between *B. divergens* and *B. bigemina* as well as between *B. divergens* and *B. bovis* in bovine RBCs were excluded from this analysis to minimize the confounding effect caused by differences in parasites.

Enrichment analyses

DEGs were mapped to their orthologs in *T. gondii*. GOEA was performed using available GO terms on ToxoDB (<https://toxodb.org/>), and significant GO terms (Benjamini < 0.1) were determined. The log fold enrichment and log p -value were used to visualize the significant GO terms.

Time-course clustering

Mean gene expression curves of inferred replication cycle genes were used to construct an $n \times N$ time-course matrix, with rows representing genes and columns representing pseudo-time

bins. Data were scaled to z-scores, and a DTW metric was used to measure the similarity between curves and to perform a hierarchical clustering. For this analysis we used the `tsclust` function from the R package `dtwclust` (<https://github.com/asardaes/dtwclust>) with parameters `control = hierarchical_control(method = "complete")`, `args = tsclust_args(dist = list(window.size = 4L))`. The total number of clusters was set empirically with trial and error. Genes were ordered according to peak expression time, and cells according to their inferred replication cycle phase (transition points along the pseudo-time course). A heatmap was used to visualize the expression profile of the genes.

Reconstruction of the gene–gene interaction network

We used a Gaussian graphical model (GGM) to assemble a gene–gene interaction network using the scRNA-seq expression data. The GGM can be used to calculate the partial correlation between gene pairs conditioned on the rest of the genes, and thus it captures pairwise relationships between the nodes in the interaction graph. Partial correlation is then used to assemble a gene–gene interaction network, where genes represent nodes and edges represent a direct interaction between them after accounting for tertiary effects. The objective of the GGM is given by

$$\max_{\Theta} \log(\det \Theta) - \text{tr}(S\Theta) - \lambda \|\Theta\|_1,$$

where S and Θ are the empirical covariance and precision matrices and λ is the sparsity penalty.

To fit this model, we estimated the empirical covariance matrix S using the pseudo-time-course gene expression data as follows. First, the mean trend $\mu_i(t_j)$ of each gene i at time point t_j was estimated using the expression of replicate cells that mapped to time partition i . This mean trend was removed from the expression of genes to de-trend the data:

$$y_i^{j,\ell,d}(t_j) = y_i^{j,\ell}(t_j) - \mu_i(t_j).$$

The superscript represents (replicate) cell j_ℓ at time bin j . The trended data were used to estimate the empirical covariance matrix S . As gene expression is periodic during the replication cycle, and assuming a non-time-varying covariance matrix, the GGM model can be directly applied to the de-trend data to capture direct covariations in gene expression. The Objective function of the GGM was then fitted for a grid of λ values ranging from 0.01 to 1.0 with step size 0.01. The R package `glassoFast` was used for fitting the model (<https://github.com/JClavel/glassoFast>). The fitted precision matrices were converted to partial correlation matrices P , which in turn were converted to network adjacency matrices. The scale-free network property for each network was calculated, and the penalty value that maximized the scale-free property was used to identify the “optimal” network.

Motif search analysis

Annotated *P. falciparum* TFs and AP2s were mapped to their *Babesia* spp. orthologs when available (S4 Table). For each TF, the interactomes were identified in the assembled co-expression networks in each species, and the union of all interacting genes across all species was taken as the TF’s overall interactome. The promoter of genes in the overall interactomes of each TF was defined as the sequence N nucleotides upstream of the start codon, where N was set to 167 in *B. bigemina*, 219 in *B. bovis*, and 190 in *B. divergens*. These promoter lengths were selected as one-half of the mean gap between consecutive genes in each species, calculated from genome annotation files (S2 Table). The promoters were extracted using the `getfasta`

command from the BEDTools package [125] and the latest version of the genome of each species, downloaded from PiroplasmaDB (<https://piroplasmadb.org/>). Motif search analysis was carried out on promoter sequences using the MEME suite [95]. The meme command with the parameters -dna -minw 4 -maxw 10 anr -nmotifs 10 -revcomp was used to perform the motif search.

Data visualization

A web dashboard was built to provide easy access to single-cell expression data and analysis results. Data were preprocessed and loaded as tables to a SQL database. The web interface was implemented using the “dash” python framework, which allows building of dynamic, interactive, data-driven apps. The current instance of the app is running in a single server using docker containers. However, the app design and the stateless approach of the framework allows for easy scalability to support increasing traffic as needed. The app for the dashboard is organized as a python module and separated into a sub-module for each of the pages available. URL requests are processed using an index python script, which loads the appropriate UI layout and backend logic. The python app is served by a Gunicorn WSGI HTTP Server, which allows it to spawn multiple workers for the app. MariaDB is used as the SQL server for the app, with connection pools of size 32 for each python worker, allowing multiple persistent connections to the database. A series of bash and python scripts were written to automate the process of database initialization from the TSV datafiles. As the data are expected to remain static through the running time of the app, SQL tables were created with tight data size allocations to help with performance. SQL indices were created such that queries remain fast. SQL unique indices are used where possible, as a way of ensuring the input data’s integrity. The resulting relational database contains 9 tables for each of the species, holding data and metadata for genes, orthologous genes, single-cell expression experiments, spline fits for pseudo-time analysis, and nodes and edges for each of the interaction networks available. Finally, a docker-compose configuration script was written, which contains all relevant configuration parameters in one place to easily deploy the app+sql server services. The app can be accessed at <https://umbi-bio.math.umb.edu/babesiasc/>. The source code for the app is available on GitHub at https://github.com/umbibio/babesia_dash_app.

Supporting information

S1 Table. Quality control. Tab 1 contains the alignment metrics. Tab 2 summarizes the total number of genes and cells that pass the Seurat filters. Tab 3 summarizes the total number of significant genes in the fitted generalized additive model (GAM; adjusted p -value < 0.01). (XLSX)

S2 Table. Genomic statistics. Tabs 1 and 2 contain information and pie charts on the total length and percentages of exons, introns, and intergenic regions in *Babesia* species (*B. bigemina*, *B. bovis*, and *B. divergens*). *T. gondii* and *P. falciparum* are also included for comparison. Tab 3 contains strand-specific as well as the overall percentage of overlapping genes. Tab 4 contains summary statistics (mean and median) of gaps between consecutive genes. (XLSX)

S3 Table. Differential expression. Tabs 1–5 contain the list of differentially expressed genes (DEGs) of the inferred cell cycle phases (FC > 2, adjusted p -value < 0.01). Tabs 5 and 6 contain the list of conserved and species-specific DEGs. Tabs 7 and 8 contain the list of *B. divergens* and species-independent DEGs in the human versus bovine comparison. (XLSX)

S4 Table. GO terms. Tabs 1–3 contain the list of enriched GO terms for conserved and species-specific DEGs and human versus bovine DEGs.

(XLSX)

S5 Table. Functionally related genes. Tabs 1–3 contain the list of orthologous *Babesia* transcription factors (TFs), AP2s, and epigenetic markers. Tabs 4–6 contain the list of VESA genes in *B. bovis*, *B. bigemina*, and *B. divergens*.

(XLSX)

S6 Table. Gene clusters. Tabs 1 and 2 contain the list of genes in each cluster for epigenetic markers (Fig 8) and TFs and AP2s (Fig J in S1 Text).

(XLSX)

S1 Text. Additional figures. S1 Text also contains extended analysis on species-specific markers and the human versus bovine marker analysis.

(PDF)

Acknowledgments

Thanks to David Allred of the University of Florida for assistance in identifying VESA genes for our analysis and many helpful discussions. Also, thanks to David Degras for helpful discussion regarding network analysis.

Author Contributions

Conceptualization: Marc-Jan Gubbels, Manoj T. Duraisingh, Kouros Zarrinhalam.

Data curation: Caroline D. Keroack.

Formal analysis: Yasaman Rezvani, Caroline D. Keroack, Brendan Elsworth, Kouros Zarrinhalam.

Investigation: Marc-Jan Gubbels, Manoj T. Duraisingh, Kouros Zarrinhalam.

Methodology: Yasaman Rezvani, Caroline D. Keroack, Brendan Elsworth, Marc-Jan Gubbels, Manoj T. Duraisingh, Kouros Zarrinhalam.

Software: Yasaman Rezvani, Argenis Arriojas, Kouros Zarrinhalam.

Validation: Caroline D. Keroack.

Writing – original draft: Yasaman Rezvani, Caroline D. Keroack, Brendan Elsworth, Argenis Arriojas, Marc-Jan Gubbels, Manoj T. Duraisingh, Kouros Zarrinhalam.

References

1. Hunfeld K, Hildebrandt A, Gray J. Babesiosis: recent insights into an ancient disease. *Int J Parasitol.* 2008; 38:1219–37. <https://doi.org/10.1016/j.ijpara.2008.03.001> PMID: 18440005
2. Bock R, Jackson L, De Vos A, Jorgensen W. Babesiosis of cattle. *Parasitology.* 2004; 129:S247–69. <https://doi.org/10.1017/s0031182004005190> PMID: 15938514
3. Kivaria FM. Estimated direct economic costs associated with tick-borne diseases on cattle in Tanzania. *Trop Anim Health Prod.* 2006; 38:291–9. <https://doi.org/10.1007/s11250-006-4181-2> PMID: 17137131
4. Homer MJ, Aguilar-Delfin I, Telford SR, Krause PJ, Persing DH. Babesiosis. *Clin Microbiol Rev.* 2000; 13:451–69. <https://doi.org/10.1128/CMR.13.3.451> PMID: 10885987
5. Vannier E, Krause PJ. Human babesiosis. *N Engl J Med.* 2012; 366:2397–407. <https://doi.org/10.1056/NEJMra1202018> PMID: 22716978

6. Vannier E, Gewurz BE, Krause PJ. Human babesiosis. *Infect Dis Clin North Am*. 2008; 22:469–88. <https://doi.org/10.1016/j.idc.2008.03.010> PMID: 18755385
7. Wormser GP, Prasad A, Neuhaus E, Joshi S, Nowakowski J, Nelson J, et al. Emergence of resistance to azithromycin-atovaquone in immunocompromised patients with *Babesia microti* infection. *Clin Infect Dis*. 2010; 50:381–6. <https://doi.org/10.1086/649859> PMID: 20047477
8. Krause PJ, Gewurz BE, Hill D, Marty FM, Vannier E, Foppa IM, et al. Persistent and relapsing babesiosis in immunocompromised patients. *Clin Infect Dis*. 2008; 46:370–6. <https://doi.org/10.1086/525852> PMID: 18181735
9. Goethert HK, Telford SR. What is *Babesia microti*? *Parasitology*. 2003; 127:301–9. <https://doi.org/10.1017/s0031182003003822> PMID: 14636016
10. Bish EK, Moritz ED, El-Amine H, Bish DR, Stramer SL. Cost-effectiveness of *Babesia microti* antibody and nucleic acid blood donation screening using results from prospective investigational studies. *Transfusion*. 2015; 55:2256–71. <https://doi.org/10.1111/trf.13136> PMID: 25995054
11. Bloch EM, Herwaldt BL, Leiby DA, Shaieb A, Herron RM, Chervenak M, et al. The third described case of transfusion-transmitted *Babesia duncani*. *Transfusion*. 2012; 52:1517–22. <https://doi.org/10.1111/j.1537-2995.2011.03467.x> PMID: 22168221
12. Saito-Ito A, Tsuji M, Wei Q, He S, Matsui T, Kohsaki M, et al. Transfusion-acquired, autochthonous human babesiosis in Japan: isolation of *Babesia microti*-like parasites with hu-RBC-SCID mice. *J Clin Microbiol*. 2000; 38:4511–6. <https://doi.org/10.1128/JCM.38.12.4511-4516.2000> PMID: 11101588
13. Curcio SR, Tria LP, Guwca AL. Seroprevalence of *Babesia microti* in individuals with Lyme disease. *Vector Borne Zoonotic Dis*. 2016; 16:737–43. <https://doi.org/10.1089/vbz.2016.2020> PMID: 27911694
14. Laha R, Das M, Sen A. Morphology, epidemiology, and phylogeny of *Babesia*: an overview. *Trop Parasitol*. 2015; 5:94–100. <https://doi.org/10.4103/2229-5070.162490> PMID: 26629451
15. He L, Bastos RG, Sun Y, Hua G, Guan G, Zhao J, et al. Babesiosis as a potential threat for bovine production in China. *Parasit Vectors*. 2021; 14:460. <https://doi.org/10.1186/s13071-021-04948-3> PMID: 34493328
16. Suarez CE, Noh S. Emerging perspectives in the research of bovine babesiosis and anaplasmosis. *Vet Parasitol*. 2011; 180:109–25. <https://doi.org/10.1016/j.vetpar.2011.05.032> PMID: 21684084
17. Eichenberger RM, Ramakrishnan C, Russo G, Deplazes P, Hehl AB. Genome-wide analysis of gene expression and protein secretion of *Babesia canis* during virulent infection identifies potential pathogenicity factors. *Sci Rep*. 2017; 7:3357. <https://doi.org/10.1038/s41598-017-03445-x> PMID: 28611446
18. González LM, Estrada K, Grande R, Jiménez-Jacinto V, Vega-Alvarado L, Sevilla E, et al. Comparative and functional genomics of the protozoan parasite *Babesia divergens* highlighting the invasion and egress processes. *PLoS Negl Trop Dis*. 2019; 13:e0007680. <https://doi.org/10.1371/journal.pntd.0007680> PMID: 31425518
19. Peloakgosi-Shikwambani K. Analysis of *Babesia rossi* transcriptome in dogs diagnosed with canine babesiosis. Pretoria: University of South Africa Institutional Repository; 2018 [cited 2021 Nov 15]. Available from: <https://uir.unisa.ac.za/handle/10500/24424>.
20. Silva JC, Cornillot E, McCracken C, Usmani-Brown S, Dwivedi A, Ifeonu OO, et al. Genome-wide diversity and gene expression profiling of *Babesia microti* isolates identify polymorphic genes that mediate host-pathogen interactions. *Sci Rep*. 2016; 6:35284. <https://doi.org/10.1038/srep35284> PMID: 27752055
21. Ueti MW, Johnson WC, Kappmeyer LS, Herndon DR, Mousel MR, Reif KE, et al. Transcriptome dataset of *Babesia bovis* life stages within vertebrate and invertebrate hosts. *Data Brief*. 2020; 33:106533. <https://doi.org/10.1016/j.dib.2020.106533> PMID: 33294524
22. Yamagishi J, Natori A, Tolba MEM, Mongan AE, Sugimoto C, Katayama T, et al. Interactive transcriptome analysis of malaria patients and infecting *Plasmodium falciparum*. *Genome Res*. 2014; 24:1433–44. <https://doi.org/10.1101/gr.158980.113> PMID: 25091627
23. Behnke MS, Wootton JC, Lehmann MM, Radke JB, Lucas O, Nawas J, et al. Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of *Toxoplasma gondii*. *PLoS ONE*. 2010; 5:e12354. <https://doi.org/10.1371/journal.pone.0012354> PMID: 20865045
24. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLOS Biol*. 2003; 1:e5. <https://doi.org/10.1371/journal.pbio.0000005> PMID: 12929205
25. Chappell L, Ross P, Orchard L, Russell TJ, Otto TD, Berriman M, et al. Refining the transcriptome of the human malaria parasite *Plasmodium falciparum* using amplification-free RNA-seq. *BMC Genomics*. 2020; 21:395. <https://doi.org/10.1186/s12864-020-06787-5> PMID: 32513207

26. Llinás M, Bozdech Z, Wong ED, Adai AT, DeRisi JL. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res.* 2006; 34:1166–73. <https://doi.org/10.1093/nar/gkj517> PMID: 16493140
27. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Böhme U, et al. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol Microbiol.* 2010; 76:12–24. <https://doi.org/10.1111/j.1365-2958.2009.07026.x> PMID: 20141604
28. Elsworth B, Duraisingh MT. A framework for signaling throughout the life cycle of *Babesia* species. *Mol Microbiol.* 2021; 115:882–90. <https://doi.org/10.1111/mmi.14650> PMID: 33274587
29. Brayton KA, Lau AOT, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, et al. Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotezoa. *PLoS Pathog.* 2007; 3:1401–13. <https://doi.org/10.1371/journal.ppat.0030148> PMID: 17953480
30. Cornillot E, Hadj-Kaddour K, Dassouli A, Noel B, Ranwez V, Vacherie B, et al. Sequencing of the smallest Apicomplexan genome from the human pathogen *Babesia microti*. *Nucleic Acids Res.* 2012; 40:9102–14. <https://doi.org/10.1093/nar/gks700> PMID: 22833609
31. Cuesta I, González LM, Estrada K, Grande R, Zaballo A, Lobo CA, et al. High-quality draft genome sequence of *Babesia divergens*, the etiological agent of cattle and human babesiosis. *Genome Announc.* 2014; 2:e01194–14. <https://doi.org/10.1128/genomeA.01194-14> PMID: 25395649
32. Jackson AP, Otto TD, Darby A, Ramaprasad A, Xia D, Echaide IE, et al. The evolutionary dynamics of variant antigen genes in *Babesia* reveal a history of genomic innovation underlying host–parasite interaction. *Nucleic Acids Res.* 2014; 42:7113–31. <https://doi.org/10.1093/nar/gku322> PMID: 24799432
33. Lemieux JE, Tran AD, Freimark L, Schaffner SF, Goethert H, Andersen KG, et al. A global map of genetic diversity in *Babesia microti* reveals strong population structure and identifies variants associated with clinical relapse. *Nat Microbiol.* 2016; 1:16079. <https://doi.org/10.1038/nmicrobiol.2016.79> PMID: 27572973
34. Niu Q, Liu Z, Yang J, Yu P, Pan Y, Zhai B, et al. Expression analysis and biological characterization of *Babesia* sp. BQ1 (Lintan) (*Babesia motasi*-like) rhoptry-associated protein 1 and its potential use in serodiagnosis via ELISA. *Parasit Vectors.* 2016; 9:313. <https://doi.org/10.1186/s13071-016-1573-7> PMID: 27245213
35. Sachman-Ruiz B, Lozano L, Lira JJ, Martínez G, Rojas C, Álvarez JA, et al. A comparative genomic study of attenuated and virulent strains of *Babesia bigemina*. *Pathogens.* 2021; 10:318. <https://doi.org/10.3390/pathogens10030318> PMID: 33800466
36. Yamagishi J, Asada M, Hakimi H, Tanaka TQ, Sugimoto C, Kawazu S. Whole-genome assembly of *Babesia ovata* and comparative genomics between closely related pathogens. *BMC Genomics.* 2017; 18:832. <https://doi.org/10.1186/s12864-017-4230-4> PMID: 29078748
37. Cursino-Santos JR, Singh M, Pham P, Rodriguez M, Lobo CA. *Babesia divergens* builds a complex population structure composed of specific ratios of infected cells to ensure a prompt response to changing environmental conditions. *Cell Microbiol.* 2016; 18:859–74. <https://doi.org/10.1111/cmi.12555> PMID: 26663747
38. Hakimi H, Asada M, Ishizaki T, Kawazu S. Isolation of viable *Babesia bovis* merozoites to study parasite invasion. *Sci Rep.* 2021; 11:16959. <https://doi.org/10.1038/s41598-021-96365-w> PMID: 34417510
39. Elsworth B, Keroack CD, Rezvani Y, Tennessen JA, Sack SA, Paul A, et al. *Babesia divergens* egress from host cells is orchestrated by essential and druggable kinases and proteases. *bioRxiv.* 2022 Feb 18. <https://doi.org/10.1101/2022.02.17.480550>
40. Cubi R, Vembar SS, Biton A, Franetich J-F, Bordessoulles M, Sossau D, et al. Laser capture microdissection enables transcriptomic analysis of dividing and quiescent liver stages of *Plasmodium relapsing* species. *Cell Microbiol.* 2017; 19:e12735. <https://doi.org/10.1111/cmi.12735> PMID: 28256794
41. Howick VM, Russell AJC, Andrews T, Heaton H, Reid AJ, Natarajan K, et al. The Malaria Cell Atlas: single parasite transcriptomes across the complete *Plasmodium* life cycle. *Science.* 2019;365: eaaw2619. <https://doi.org/10.1126/science.aaw2619> PMID: 31439762
42. Poran A, Nötzel C, Aly O, Mencia-Trinchant N, Harris CT, Guzman ML, et al. Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature.* 2017; 551:95–9. <https://doi.org/10.1038/nature24280> PMID: 29094698
43. Rawat M, Srivastava A, Gupta I, Karmodiya K. Single cell RNA-sequencing reveals cellular heterogeneity, stage transition and antigenic variation during stress adaptation in synchronized *Plasmodium falciparum*. *bioRxiv.* 2019 Aug 31. <https://doi.org/10.1101/752543>
44. Reid AJ, Talman AM, Bennett HM, Gomes AR, Sanders MJ, Illingworth CJR, et al. Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *Elife.* 2018; 7:e33105. <https://doi.org/10.7554/eLife.33105> PMID: 29580379

45. Ruberto AA, Bourke C, Merienne N, Obadia T, Amino R, Mueller I. Single-cell RNA sequencing reveals developmental heterogeneity among *Plasmodium berghei* sporozoites. *Sci Rep*. 2021; 11:4127. <https://doi.org/10.1038/s41598-021-82914-w> PMID: 33619283
46. Sà JM, Cannon MV, Caleon RL, Wellem's TE, Serre D. Single-cell transcription analysis of *Plasmodium vivax* blood-stage parasites identifies stage- and species-specific profiles of expression. *PLOS Biol*. 2020; 18:e3000711. <https://doi.org/10.1371/journal.pbio.3000711> PMID: 32365102
47. Walzer KA, Fradin H, Emerson LY, Corcoran DL, Chi J-T. Latent transcriptional variations of individual *Plasmodium falciparum* uncovered by single-cell RNA-seq and fluorescence imaging. *PLOS Genet*. 2019; 15:e1008506. <https://doi.org/10.1371/journal.pgen.1008506> PMID: 31856180
48. Real E, Howick VM, Dahalan FA, Witmer K, Cudini J, Andradi-Brown C, et al. A single-cell atlas of *Plasmodium falciparum* transmission through the mosquito. *Nat Commun*. 2021; 12:3196. <https://doi.org/10.1038/s41467-021-23434-z> PMID: 34045457
49. Waldman BS, Schwarz D, Wadsworth MH, Saeij JP, Shalek AK, Lourido S. Identification of a master regulator of differentiation in *Toxoplasma*. *Cell*. 2020; 180:359–72.e16. <https://doi.org/10.1016/j.cell.2019.12.013> PMID: 31955846
50. Xue Y, Theisen TC, Rastogi S, Ferrel A, Quake SR, Boothroyd JC. A single-parasite transcriptional atlas of *Toxoplasma gondii* reveals novel control of antigen expression. *eLife*. 2020; 9:e54129. <https://doi.org/10.7554/eLife.54129> PMID: 32065584
51. Keroack CD, Elsworth B, Duraisingh MT. To kill a piroplasm: genetic technologies to advance drug discovery and target identification in *Babesia*. *Int J Parasitol*. 2019; 49:153–63. <https://doi.org/10.1016/j.ijpara.2018.09.005> PMID: 30391230
52. Schuster FL. Cultivation of *Babesia* and *Babesia*-like blood parasites: agents of an emerging zoonotic disease. *Clin Microbiol Rev*. 2002; 15:365–73. <https://doi.org/10.1128/CMR.15.3.365-373.2002> PMID: 12097245
53. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. *Cell*. 2019; 177:1888–902.e21. <https://doi.org/10.1016/j.cell.2019.05.031> PMID: 31178118
54. Hastie T, Stuetzle W. Principal curves. *J Am Stat Assoc*. 1989; 84:502–16. <https://doi.org/10.1080/01621459.1989.10478797>
55. Giorgino T. Computing and visualizing dynamic time warping alignments in R: the dtw package. *J Stat Softw*. 2009; 31:1–24. <https://doi.org/10.18637/jss.v031.i07>
56. Kreier JP, Gravely SM, Seed TM, Smucker R, Pfister RM. *Babesia* sp.: the relationship of stage of development to structure of intra- and extracellular parasites. *Tropenmed Parasitol*. 1975; 26:9–18. PMID: 1145729
57. Tang D, Wang Y. Cell cycle regulation of Golgi membrane dynamics. *Trends Cell Biol*. 2013; 23:296–304. <https://doi.org/10.1016/j.tcb.2013.01.008> PMID: 23453991
58. Nishi M, Hu K, Murray JM, Roos DS. Organellar dynamics during the cell cycle of *Toxoplasma gondii*. *J Cell Sci*. 2008; 121:1559–68. <https://doi.org/10.1242/jcs.021089> PMID: 18411248
59. van Dooren GG, Marti M, Tonkin CJ, Stimmler LM, Cowman AF, McFadden GI. Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*. *Mol Microbiol*. 2005; 57:405–19. <https://doi.org/10.1111/j.1365-2958.2005.04699.x> PMID: 15978074
60. van Dooren GG, Yeoh LM, Striepen B, McFadden GI. The import of proteins into the mitochondrion of *Toxoplasma gondii*. *J Biol Chem*. 2016; 291:19335–50. <https://doi.org/10.1074/jbc.M116.725069> PMID: 27458014
61. Nozawa A, Ito D, Ibrahim M, Santos HJ, Tsuboi T, Tozawa Y. Characterization of mitochondrial carrier proteins of malaria parasite *Plasmodium falciparum* based on in vitro translation and reconstitution. *Parasitol Int*. 2020; 79:102160. <https://doi.org/10.1016/j.parint.2020.102160> PMID: 32574727
62. Chircop M. Rho GTPases as regulators of mitosis and cytokinesis in mammalian cells. *Small GTPases*. 2014; 5:e29770. <https://doi.org/10.4161/sgtp.29770> PMID: 24988197
63. Miserey-Lenkei S, Colombo MI. Small RAB GTPases regulate multiple steps of mitosis. *Front Cell Dev Biol*. 2016; 4:2. <https://doi.org/10.3389/fcell.2016.00002> PMID: 26925400
64. Cernetich-Ott A, Daly TM, Vaidya AB, Bergman LW, Burns JM. Remarkable stability in patterns of blood-stage gene expression during episodes of non-lethal *Plasmodium yoelii* malaria. *Malar J*. 2012; 11:265. <https://doi.org/10.1186/1475-2875-11-265> PMID: 22866913
65. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*. 2003; 301:1503–8. <https://doi.org/10.1126/science.1087025> PMID: 12893887

66. Lesage KM, Huot L, Mouveau T, Courjol F, Saliou J-M, Gissot M. Cooperative binding of ApiAP2 transcription factors is crucial for the expression of virulence genes in *Toxoplasma gondii*. *Nucleic Acids Res.* 2018; 46:6057–68. <https://doi.org/10.1093/nar/gky373> PMID: 29788176
67. Nasamu AS, Glushakova S, Russo I, Vaupel B, Oksman A, Kim AS, et al. Plasmepsins IX and X are essential and druggable mediators of malaria parasite egress and invasion. *Science.* 2017; 358:518–22. <https://doi.org/10.1126/science.aan1478> PMID: 29074774
68. Šnebergerová P, Bartošová-Sojtková P, Jalovecká M, Sojka D. Plasmepsin-like aspartyl proteases in *Babesia*. *Pathogens.* 2021; 10:1241. <https://doi.org/10.3390/pathogens10101241> PMID: 34684190
69. Batugedara G, Lu XM, Saraf A, Sardu ME, Cort A, Abel S, et al. The chromatin bound proteome of the human malaria parasite. *Microb Genom.* 2020; 6:e000327. <https://doi.org/10.1099/mgen.0.000327> PMID: 32017676
70. Le Hir H, Gatfield D, Braun IC, Forler D, Izaurralde E. The protein Mago provides a link between splicing and mRNA localization. *EMBO Rep.* 2001; 2:1119–24. <https://doi.org/10.1093/embo-reports/kve245> PMID: 11743026
71. Uhler C. Gaussian graphical models: an algebraic and geometric perspective. *arXiv.* 2017 Jul 13. <https://doi.org/10.48550/arXiv.1707.04345>
72. Bebek G. Identifying gene interaction networks. In: Elston RC, Satagopan JM, Sun S, editors. *Statistical human genetics: methods and protocols.* Totowa (NJ): Humana Press; 2012. pp. 483–494. https://doi.org/10.1007/978-1-61779-555-8_26
73. Albert R. Scale-free networks in cell biology. *J Cell Sci.* 2005; 118:4947–57. <https://doi.org/10.1242/jcs.02714> PMID: 16254242
74. Mosqueda J, McElwain TF, Stiller D, Palmer GH. *Babesia bovis* merozoite surface antigen 1 and rhoptry-associated protein 1 are expressed in sporozoites, and specific antibodies inhibit sporozoite attachment to erythrocytes. *Infect Immun.* 2002; 70:1599–603. <https://doi.org/10.1128/IAI.70.3.1599-1603.2002> PMID: 11854249
75. Rodriguez M, Alhassan A, Ord RL, Cursino-Santos JR, Singh M, Gray J, et al. Identification and characterization of the RouenBd1987 *Babesia* divergens rhoptry-associated protein 1. *PLoS ONE.* 2014; 9:e107727. <https://doi.org/10.1371/journal.pone.0107727> PMID: 25226276
76. Suarez CE, Palmer GH, Florin-Christensen M, Hines SA, Hötzel I, McElwain TF. Organization, transcription, and expression of rhoptry associated protein genes in the *Babesia bigemina* rap-1 locus. *Mol Biochem Parasitol.* 2003; 127:101–12. [https://doi.org/10.1016/s0166-6851\(02\)00311-0](https://doi.org/10.1016/s0166-6851(02)00311-0) PMID: 12672519
77. Wang P, Yang W, Zhao S, Nashun B. Regulation of chromatin structure and function: insights into the histone chaperone FACT. *Cell Cycle.* 2021; 20:465–79. <https://doi.org/10.1080/15384101.2021.1881726> PMID: 33590780
78. Zabarónick SR, Tyler JK. The histone chaperone anti-silencing function 1 is a global regulator of transcription independent of passage through S phase. *Mol Cell Biol.* 2005; 25:652–60. <https://doi.org/10.1128/MCB.25.2.652-660.2005> PMID: 15632066
79. Back PS O'Shaughnessy WJ, Moon AS, Dewangan PS, Hu X, Sha J, et al. Ancient MAPK ERK7 is regulated by an unusual inhibitory scaffold required for *Toxoplasma* apical complex biogenesis. *Proc Natl Acad Sci U S A.* 2020; 117:12164–73. <https://doi.org/10.1073/pnas.1921245117> PMID: 32409604
80. O'Shaughnessy WJ, Hu X, Beraki T, McDougal M, Reese ML. Loss of a conserved MAPK causes catastrophic failure in assembly of a specialized cilium-like structure in *Toxoplasma gondii*. *Mol Biol Cell.* 2020; 31:881–8. <https://doi.org/10.1091/mbc.E19-11-0607> PMID: 32073987
81. Pacheco NDS, Tosetti N, Krishnan A, Haase R, Maco B, Suarez C, et al. Revisiting the role of *Toxoplasma gondii* ERK7 in the maintenance and stability of the apical complex. *mBio.* 2021; 12:e0205721. <https://doi.org/10.1128/mBio.02057-21> PMID: 34607461
82. Li F, Bounkeua V, Pettersen K, Vinetz JM. *Plasmodium falciparum* ookinete expression of plasmepsin VII and plasmepsin X. *Malar J.* 2016; 15:111. <https://doi.org/10.1186/s12936-016-1161-5> PMID: 26911483
83. Nikolaeva D, Illingworth JJ, Miura K, Alanine DGW, Brian IJ, Li Y, et al. Functional characterization and comparison of *Plasmodium falciparum* proteins as targets of transmission-blocking antibodies. *Mol Cell Proteomics.* 2020; 19:155–66. <https://doi.org/10.1074/mcp.RA117.000036> PMID: 29089373
84. Laurentino EC, Taylor S, Mair GR, Lasonder E, Bartfai R, Stunnenberg HG, et al. Experimentally controlled downregulation of the histone chaperone FACT in *Plasmodium berghei* reveals that it is critical to male gamete fertility. *Cell Microbiol.* 2011; 13:1956–74. <https://doi.org/10.1111/j.1462-5822.2011.01683.x> PMID: 21899698

85. Hehl AB, Lekutis C, Grigg ME, Bradley PJ, Dubremetz J-F, Ortega-Barria E, et al. Toxoplasma gondii homologue of Plasmodium apical membrane antigen 1 is involved in invasion of host cells. *Infect Immun*. 2000; 68:7078–86. <https://doi.org/10.1128/IAI.68.12.7078-7086.2000> PMID: 11083833
86. Triglia T, Healer J, Caruana SR, Hodder AN, Anders RF, Crabb BS, et al. Apical membrane antigen 1 plays a central role in erythrocyte invasion by Plasmodium species. *Mol Microbiol*. 2000; 38:706–18. <https://doi.org/10.1046/j.1365-2958.2000.02175.x> PMID: 11115107
87. Hu X, O'Shaughnessy WJ, Beraki TG, Reese ML. Loss of the conserved alveolate kinase MAPK2 decouples Toxoplasma cell growth from cell division. *mBio*. 2020; 11:e02517–20. <https://doi.org/10.1128/mBio.02517-20> PMID: 33173004
88. Yeoman JA, Hanssen E, Maier AG, Klonis N, Maco B, Baum J, et al. Tracking Glideosome-associated protein 50 reveals the development and organization of the inner membrane complex of Plasmodium falciparum. *Eukaryot Cell*. 2011; 10:556–64. <https://doi.org/10.1128/EC.00244-10> PMID: 21239623
89. Hu K, Johnson J, Florens L, Fraunholz M, Suravajjala S, DiLullo C, et al. Cytoskeletal components of an invasion machine—the apical complex of Toxoplasma gondii. *PLoS Pathog*. 2006; 2:e13. <https://doi.org/10.1371/journal.ppat.0020013> PMID: 16518471
90. Miao J, Fan Q, Cui L, Li J, Li J, Cui L. The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation. *Gene*. 2006; 369:53–65. <https://doi.org/10.1016/j.gene.2005.10.022> PMID: 16410041
91. Toenhake CG, Bártfai R. What functional genomics has taught us about transcriptional regulation in malaria parasites. *Brief Funct Genomics*. 2019; 18:290–301. <https://doi.org/10.1093/bfpg/elz004> PMID: 31220867
92. van Biljon R, van Wyk R, Painter HJ, Orchard L, Reader J, Niemand J, et al. Hierarchical transcriptional control regulates Plasmodium falciparum sexual differentiation. *BMC Genomics*. 2019; 20:920. <https://doi.org/10.1186/s12864-019-6322-9> PMID: 31795940
93. Jeninga MD, Quinn JE, Petter M. ApiAP2 transcription factors in apicomplexan parasites. *Pathogens*. 2019; 8:47. <https://doi.org/10.3390/pathogens8020047> PMID: 30959972
94. Szatanek T, Anderson-White BR, Faugno-Fusci DM, White M, Saeji JPJ, Gubbels M-J. Cactin is essential for G1 progression in Toxoplasma gondii. *Mol Microbiol*. 2012; 84:566–77. <https://doi.org/10.1111/j.1365-2958.2012.08044.x> PMID: 22486860
95. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res*. 2015; 43:W39–49. <https://doi.org/10.1093/nar/gkv416> PMID: 25953851
96. Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinás M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog*. 2010; 6:e1001165. <https://doi.org/10.1371/journal.ppat.1001165> PMID: 21060817
97. Young JA, Johnson JR, Benner C, Yan SF, Chen K, Le Roch KG, et al. In silico discovery of transcription regulatory elements in Plasmodium falciparum. *BMC Genomics*. 2008; 9:70. <https://doi.org/10.1186/1471-2164-9-70> PMID: 18257930
98. Toenhake CG, Frasncka SA-K, Vijayabaskar MS, Westhead DR, van Heeringen SJ, Bártfai R. Chromatin accessibility-based characterization of the gene regulatory network underlying Plasmodium falciparum blood-stage development. *Cell Host Microbe*. 2018; 23:557–69.e9. <https://doi.org/10.1016/j.chom.2018.03.007> PMID: 29649445
99. Gubbels M-J, Keroack CD, Dangoudoubiyam S, Worliczek HL, Paul AS, Bauwens C, et al. Fussing about fission: defining variety among mainstream and exotic apicomplexan cell division modes. *Front Cell Infect Microbiol*. 2020; 10:269. <https://doi.org/10.3389/fcimb.2020.00269> PMID: 32582569
100. Franssen FFJ, Gaffar FR, Yatsuda AP, de Vries E. Characterisation of erythrocyte invasion by Babesia bovis merozoites efficiently released from their host cell after high-voltage pulsing. *Microbes Infect*. 2003; 5:365–72. [https://doi.org/10.1016/s1286-4579\(03\)00041-8](https://doi.org/10.1016/s1286-4579(03)00041-8) PMID: 12737991
101. Conesa JJ, Sevilla E, Terrón MC, González LM, Gray J, Pérez-Berná AJ, et al. Four-dimensional characterization of the Babesia divergens asexual life cycle, from the trophozoite to the multiparasite stage. *mSphere*. 2020; 5:e00928–20. <https://doi.org/10.1128/mSphere.00928-20> PMID: 33055261
102. Rossouw I, Maritz-Olivier C, Niemand J, Biljon R van, Smit A, Olivier NA, et al. Morphological and molecular descriptors of the developmental cycle of Babesia divergens parasites in human erythrocytes. *PLoS Negl Trop Dis*. 2015; 9:e0003711. <https://doi.org/10.1371/journal.pntd.0003711> PMID: 25955414
103. Baran-Gale J, Chandra T, Kirschner K. Experimental design for single-cell RNA sequencing. *Brief Funct Genomics*. 2018; 17:233–9. <https://doi.org/10.1093/bfpg/elx035> PMID: 29126257
104. Chen G, Ning B, Shi T. Single-cell RNA-Seq technologies and related computational data analysis. *Front Genet*. 2019; 10:317. <https://doi.org/10.3389/fgene.2019.00317> PMID: 31024627

105. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol.* 2020; 38:737–46. <https://doi.org/10.1038/s41587-020-0465-8> PMID: 32341560
106. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 2017; 9:75. <https://doi.org/10.1186/s13073-017-0467-4> PMID: 28821273
107. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 2016; 17:29. <https://doi.org/10.1186/s13059-016-0888-1> PMID: 26887813
108. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. *Genome Biol.* 2020; 21:31. <https://doi.org/10.1186/s13059-020-1926-6> PMID: 32033589
109. Mou T, Deng W, Gu F, Pawitan Y, Vu TN. Reproducibility of methods to detect differentially expressed genes from single-cell RNA sequencing. *Front Genet.* 2020; 10:1331. <https://doi.org/10.3389/fgene.2019.01331> PMID: 32010190
110. Qiu P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat Commun.* 2020; 11:1169. <https://doi.org/10.1038/s41467-020-14976-9> PMID: 32127540
111. Amin TM, Sirs JA. The blood rheology of man and various animal species. *Q J Exp Physiol.* 1985; 70:37–49. <https://doi.org/10.1113/expphysiol.1985.sp002895> PMID: 4011828
112. Florin-Christensen J, Suarez CE, Florin-Christensen M, Wainszelbaum M, Brown WC, McElwain TF, et al. A unique phospholipid organization in bovine erythrocyte membranes. *Proc Natl Acad Sci U S A.* 2001; 98:7736–41. <https://doi.org/10.1073/pnas.131580998> PMID: 11427712
113. Jikuya T, Tsutsui T, Shigeta O, Sankai Y, Mitsui T. Species differences in erythrocyte mechanical fragility: comparison of human, bovine, and ovine cells. *ASAIO J.* 1998; 44:M452–5. PMID: 9804471
114. Stone WH. The relation of human and cattle blood groups. *Transfusion.* 1962; 2:172–7. <https://doi.org/10.1111/j.1537-2995.1962.tb00217.x> PMID: 13917479
115. Sunshine H, Iruela-Arispe ML. Membrane lipids and cell signaling. *Curr Opin Lipidol.* 2017; 28:408–13. <https://doi.org/10.1097/MOL.0000000000000443> PMID: 28692598
116. Hutchings CL, Li A, Fernandez KM, Fletcher T, Jackson LA, Molloy JB, et al. New insights into the altered adhesive and mechanical properties of red blood cells parasitized by *Babesia bovis*. *Mol Microbiol.* 2007; 65:1092–105. <https://doi.org/10.1111/j.1365-2958.2007.05850.x> PMID: 17640278
117. O'Connor RM, Allred DR. Selection of *Babesia bovis*-infected erythrocytes for adhesion to endothelial cells coselects for altered variant erythrocyte surface antigen isoforms. *J Immunol.* 2000; 164:2037–45. <https://doi.org/10.4049/jimmunol.164.4.2037> PMID: 10657656
118. O'Connor RM, Long JA, Allred DR. Cytoadherence of *Babesia bovis*-infected erythrocytes to bovine brain capillary endothelial cells provides an in vitro model for sequestration. *Infect Immun.* 1999; 67:3921–8. <https://doi.org/10.1128/IAI.67.8.3921-3928.1999> PMID: 10417157
119. Sondgeroth KS, McElwain TF, Allen AJ, Chen AV, Lau AO. Loss of neurovirulence is associated with reduction of cerebral capillary sequestration during acute *Babesia bovis* infection. *Parasit Vectors.* 2013; 6:181. <https://doi.org/10.1186/1756-3305-6-181> PMID: 23777713
120. Suarez CE, Alzan HF, Silva MG, Rathinasamy V, Poole WA, Cooke BM. Unravelling the cellular and molecular pathogenesis of bovine babesiosis: is the sky the limit? *Int J Parasitol.* 2019; 49:183–97. <https://doi.org/10.1016/j.ijpara.2018.11.002> PMID: 30690089
121. Hakimi H, Templeton TJ, Sakaguchi M, Yamagishi J, Miyazaki S, Yahata K, et al. Novel *Babesia bovis* exported proteins that modify properties of infected red blood cells. *PLOS Pathog.* 2020; 16: e1008917. <https://doi.org/10.1371/journal.ppat.1008917> PMID: 33017449
122. Lewis D, Purnell RE, Shaw SR, Revington JP. The isolation and characterization of human and bovine strains of *Babesia divergens* from Drumnadrochit, Scotland. *Parasitology.* 1980; 81:145–55. <https://doi.org/10.1017/S0031182000055116> PMID: 7422359
123. Guan G, Korhonen PK, Young ND, Koehler AV, Wang T, Li Y, et al. Genomic resources for a unique, low-virulence *Babesia* taxon from China. *Parasit Vectors.* 2016; 9:564. <https://doi.org/10.1186/s13071-016-1846-1> PMID: 27784333
124. Duffy MF, Selvarajah SA, Josling GA, Petter M. Epigenetic regulation of the *Plasmodium falciparum* genome. *Brief Funct Genomics.* 2014; 13:203–16. <https://doi.org/10.1093/bfpg/elt047> PMID: 24326119
125. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010; 26:841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278