

Single-cell analysis reveals divergent developmental trajectories and regulatory networks in *Plasmodium falciparum* lab strains and field isolates

DENARIO¹

¹Anthropic, Gemini & OpenAI servers. Planet Earth.

ABSTRACT

Adaptation of *Plasmodium falciparum* to distinct environments, such as laboratory culture versus the dynamic human host, leads to significant changes in parasite development and fate determination, but the underlying dynamic transcriptional programs and regulatory networks governing these processes remain poorly understood. To elucidate these differences, we employed single-cell RNA sequencing to dissect and compare the dynamic transcriptional programs and inferred regulatory networks controlling stage transitions in laboratory-adapted strains and field isolates. We analyzed a comprehensive dataset of over 45,000 single cells from both sources, reconstructing developmental trajectories, characterizing gene expression dynamics along pseudotime, identifying co-expressed gene modules, and inferring candidate regulators exhibiting transient expression patterns. Our analysis revealed marked differences in cellular composition and extensive differential gene expression between lab and field parasites, even within shared life cycle stages. Trajectory inference highlighted divergent developmental paths, most notably an extended sexual development program in field isolates that progresses to late-stage gametocytes not typically observed in standard laboratory culture. Furthermore, analysis of gene modules and transiently expressed regulators indicated substantial rewiring of the regulatory networks controlling stage transitions, suggesting differences in the timing and complexity of regulatory events, particularly a potentially more complex regulatory landscape governing gametocytogenesis in field parasites. These findings provide a high-resolution, dynamic perspective on how adaptation shapes *Plasmodium falciparum* development and identify key transcriptional regulators potentially mediating host-specific developmental control.

Keywords: Clustering, Dimensionality reduction, Regression, Smoothing, Principal component analysis

1. INTRODUCTION

Malaria, a disease caused by the eukaryotic parasite *Plasmodium falciparum*, remains a significant global health burden. The parasite's complex life cycle involves distinct developmental stages within both mosquito vectors and human hosts. Within human red blood cells, *P. falciparum* undergoes rapid asexual multiplication, which is responsible for the clinical manifestations of malaria. A crucial aspect of the parasite's life cycle is its ability to differentiate from the asexual proliferative stage into sexual precursor stages called gametocytes. These gametocytes are essential for transmission to the mosquito vector, making the decision to commit to sexual development and the subsequent maturation process critical determinants of parasite fitness and disease spread.

Precise control over the timing and regulation of these developmental stage transitions is paramount for parasite survival and transmission success. However, studying these dynamic processes presents significant challenges. *P. falciparum* populations are inherently asynchronous, comprising cells at various developmental stages simultaneously, making bulk analyses inherently limited. Furthermore, the parasite's biology is profoundly influenced by its environment. Laboratory-adapted strains, maintained under stable *in vitro* culture conditions, often diverge significantly in their developmental characteristics and gene expression patterns compared to field isolates circulating in diverse human populations, which are exposed to varying host immune responses, drug pressures, and nutrient availability. While it is recognized that adaptation to these distinct environments leads to changes in parasite development and fate determination, the underlying dynamic

transcriptional programs and the regulatory networks that govern these differences at the single-cell level are not well understood. Resolving the heterogeneity within parasite populations and capturing the continuous progression through developmental stages requires high-resolution approaches capable of analyzing individual cells.

Single-cell RNA sequencing (scRNA-seq) has emerged as a transformative technology for dissecting complex biological systems, particularly those involving dynamic processes like differentiation and development in heterogeneous cell populations. (Flores et al. 2021; Patel et al. 2024) By providing transcriptome-wide data from individual cells, scRNA-seq enables the reconstruction of developmental trajectories and the detailed characterization of gene expression dynamics along these inferred paths, often referred to as pseudotime. (Flores et al. 2021; Patel et al. 2024)

In this study, we leveraged the power of scRNA-seq to provide a high-resolution, dynamic perspective on *P. falciparum* development and to compare the transcriptional landscapes and inferred regulatory networks between laboratory-adapted strains and field isolates (Li 2018; Zitnik et al. 2024). We analyzed a comprehensive dataset comprising over 45,000 single cells originating from both sources. To dissect the differences shaped by laboratory adaptation versus natural *in vivo* conditions, we employed advanced computational methods. These methods allowed us to reconstruct the developmental trajectories of the parasites, characterize the dynamic changes in gene expression along pseudotime, identify modules of co-expressed genes representing coordinated biological processes (Zitnik et al. 2024), and infer candidate regulatory factors exhibiting transient expression patterns associated with stage transitions (Li 2018). We ensured the robustness of our findings through rigorous data quality control, statistical analysis, and comparative examination across multiple levels of biological organization, from overall cellular composition and differential gene expression within stages to the fine-grained dynamics of gene expression and inferred regulatory interactions along the developmental paths.

Our analysis revealed striking differences in cellular composition and extensive differential gene expression between lab and field parasites, even within nominally shared life cycle stages. Trajectory inference highlighted significant divergences in developmental paths, most notably demonstrating a more pronounced and extended sexual development program in field isolates that progresses to late-stage gametocytes, which are typically absent or rare in standard laboratory cultures (Cao et al. 2016). Furthermore, by analyzing gene co-expression

modules and identifying candidate transiently expressed regulators, we found evidence for substantial rewiring of the regulatory networks controlling stage transitions (Ditz et al. 2023). These findings suggest differences in the timing, complexity, and key players involved in regulatory events, particularly indicating a potentially more complex regulatory landscape governing gametocytogenesis in field parasites (Cao et al. 2016).

Collectively, our results provide a detailed, dynamic perspective on how adaptation shapes *P. falciparum* development and pinpoint key transcriptional regulators that may mediate host-specific developmental control, offering valuable insights into parasite biology and identifying potential targets for future interventions (Cao et al. 2016; Ditz et al. 2023).

2. METHODS

Here we describe the experimental data and the computational methods employed to analyze single-cell RNA sequencing data from *Plasmodium falciparum* laboratory-adapted strains and field isolates. The goal was to reconstruct and compare developmental trajectories, characterize gene expression dynamics, identify co-expressed gene modules, and infer candidate transcriptional regulators controlling stage transitions, thereby providing a dynamic, high-resolution view of differences shaped by adaptation.

2.1. Data Source and Initial Processing

The dataset analyzed in this study consists of single-cell RNA sequencing data from *P. falciparum*. The raw input data were provided as a gene expression matrix and associated cell metadata. The gene expression matrix, provided in ‘gene_expression.csv’, contained normalized expression values for genes across individual cells. The cell metadata, provided in ‘labels_csv’, included crucial annotations for each cell, such as a unique cell identifier (‘CELL_ID’), assigned life cycle stage (e.g., Ring, Trophozoite, Schizont, Gametocyte) (Motta et al. 2024), the parasite strain, and the source (laboratory-adapted or field isolate), along with specific identifiers for field isolates (MSC1, MSC3, MSC13, MSC14) and days in culture for laboratory strains.

Data loading was performed using standard bioinformatics libraries (e.g., in Python or R). The gene expression matrix was ingested and subsequently merged with the cell metadata based on the common ‘CELL_ID’.

For compatibility with common single-cell analysis workflows, the expression matrix was transposed such that rows represented individual cells and columns represented genes (Riffle et al. 2025).

2.2. Quality Control and Filtering

Initial quality control steps were performed to ensure the reliability of the single-cell data. Genes with minimal expression across the entire dataset were removed; specifically, genes detected in fewer than 3 cells or exhibiting zero total expression counts were filtered out (Dajani et al. 2025; Dandala et al. 2025).

While the provided data was pre-processed, standard cell-level quality metrics were assessed, including the total number of detected genes per cell and the total normalized expression per cell (Dajani et al. 2025; Dandala et al. 2025).

Cells that presented as extreme outliers in these distributions were considered for potential exclusion, though primary reliance was placed on the provided metadata for cell quality status (Dajani et al. 2025; Dandala et al. 2025). The sparsity of the expression matrix (percentage of zero values) was also calculated to understand the nature of the single-cell data.

2.3. Normalization and Feature Selection

The provided "normalized expression value" was treated as the primary expression metric. Assuming these values represented library-size corrected counts, a log-transformation (specifically, $\log_e(x+1)$ or 'log1p') was applied to the expression matrix. This transformation is standard practice in scRNA-seq analysis to stabilize variance across the range of expression values and render the data distribution more suitable for linear dimensionality reduction techniques. To reduce computational complexity and focus on biologically relevant variation, highly variable genes (HVGs) were identified across all cells. These genes, exhibiting significant variance in expression relative to their mean expression, are typically more informative for distinguishing cell states. The identified HVGs were then used for subsequent dimensionality reduction and downstream analyses.

2.4. Dimensionality Reduction and Visualization

Principal Component Analysis (PCA) was applied to the scaled expression matrix of the selected HVGs (Ali et al. 2017). PCA identifies the principal components (PCs) that capture the maximum variance in the dataset (Ali et al. 2017). The number of PCs to retain for subsequent steps was determined by examining an elbow plot of the explained variance per PC, selecting the point where the gain in explained variance plateaus. To visualize the cellular landscape in two dimensions, Uniform Manifold Approximation and Projection (UMAP) was applied to the selected PCs. UMAP is a non-linear dimensionality reduction technique that aims to preserve local and global relationships between cells. UMAP plots were generated and colored by various metadata

attributes, including the assigned life cycle stage, source (lab vs. field), specific parasite strain or field isolate identifier, and days in culture. This visualization facilitated the assessment of cellular heterogeneity, the separation of distinct cell populations, potential batch effects, and the overall structure of the dataset.

2.5. Cell Stage Annotation

The primary annotation for cell identity relied on the "life cycle stage" column provided in the 'labels_csv' metadata. These labels represent the developmental stage assigned to each cell based on prior information or sorting. To verify the consistency and accuracy of these annotations within the scRNA-seq data, the expression patterns of well-established stage-specific marker genes for **P. falciparum** were examined and visualized on the UMAP embeddings. This step confirmed that cells clustered broadly according to their assigned biological stage.

2.6. Differential Gene Expression Analysis

To identify transcriptional differences between laboratory-adapted strains and field isolates, differential gene expression (DGE) analysis was performed. Comparisons were conducted for each major life cycle stage (e.g., Ring, Trophozoite, Schizont, and potentially early/late Gametocytes if sufficient numbers were present and distinguishable) that was represented in both lab and field datasets. For each stage, gene expression levels in cells from lab strains were statistically compared to those from field isolates. Statistical tests robust to the characteristics of single-cell data, such as the Wilcoxon rank-sum test or methods based on negative binomial distributions (e.g., using tools designed for scRNA-seq DGE), were employed. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method to control the false discovery rate (FDR), with a threshold (e.g., FDR < 0.05) used to define statistical significance. Additionally, stage-specific marker genes were identified separately within the lab and field datasets by comparing cells of a given stage against all other cells within the same source group (Birkholtz et al. 2006).

2.7. Developmental Trajectory Inference and Pseudotime Assignment

To reconstruct the continuous progression of *P. falciparum* through its developmental cycle and capture the dynamics of stage transitions, developmental trajectory inference was performed. Cells identified as belonging to the asexual intraerythrocytic cycle (Ring, Trophozoite, Schizont) were subsetted for asexual trajectory reconstruction. Similarly, if a sufficient number and diversity

of gametocyte stages were present, a separate trajectory focusing on sexual commitment and development was inferred. Trajectories were reconstructed independently for the combined laboratory strains and the combined field isolates using a suitable trajectory inference algorithm (e.g., Monocle 3, Slingshot, or graph-based methods like PAGA implemented in Scanpy). The inferred trajectories represent a principal path through the high-dimensional gene expression space, ordered by an abstract measure of progress called pseudotime. The starting point (root) of the trajectories was set based on biological knowledge, typically anchoring the asexual cycle at the Ring stage. Each cell was assigned a pseudotime value reflecting its position along the inferred path.

2.8. Comparative Trajectory Analysis and Pseudotemporal Dynamics

The inferred developmental trajectories were compared between laboratory-adapted strains and field isolates at multiple levels (Zhou et al. 2024; Hutton & Meyer 2025; Zhang et al. 2025). The overall topology of the trajectories, including linearity, branching points (such as the putative branch towards sexual development), and the relative density of cells along different paths, was compared (Zhou et al. 2024; Zhang et al. 2025). Graph-based approaches (e.g., comparing PAGA graphs) were used where appropriate to quantitatively assess differences in trajectory structure (Zhang et al. 2025).

A key aspect of the comparative analysis involved examining the dynamics of gene expression along the pseudotime axis (Chen et al. 2025,?). For genes identified as differentially expressed or as stage markers, their expression levels were modeled as a function of pseudotime for both lab and field trajectories (Chen et al. 2025,?). Gene expression profiles along pseudotime were smoothed (e.g., using Generalized Additive Models (GAMs) or LOESS regression) to capture underlying trends. These smoothed dynamics were then compared between conditions to identify differences in the timing of gene activation or repression, the magnitude of expression changes during transitions, and the overall shape of the expression profile (Chen et al. 2025,?). This analysis allowed for a detailed comparison of how transcriptional programs unfold dynamically in different adaptive contexts (Chen et al. 2025,?).

2.9. Identification and Characterization of Gene Modules

Modules of co-expressed genes were identified along the inferred pseudotime trajectories for both laboratory and field isolates. This was achieved by clustering

genes based on the similarity of their smoothed pseudotemporal expression patterns. Methods such as hierarchical clustering or k-means were applied to the fitted gene expression profiles along pseudotime. Alternatively, graph-based clustering methods on gene correlation networks derived from pseudotime-ordered cells were used. Each identified module represents a set of genes with coordinated expression dynamics, potentially reflecting shared regulatory control or involvement in a common biological process. For each module, functional enrichment analysis was performed using resources like Gene Ontology (GO) and KEGG pathways, leveraging annotations available in databases such as PlasmoDB. This step aimed to assign biological meaning to the co-expressed gene sets. The composition of gene modules (which genes belong to which module) and their pseudotemporal expression profiles were then compared between lab and field isolates to identify conserved or condition-specific modules and differences in their timing or prominence.

2.10. Identification of Candidate Transient Regulators

A specific analysis was conducted to identify candidate transcriptional regulators hypothesized to exhibit transient expression patterns associated with developmental transitions, as suggested by their potential role in initiating or coordinating downstream transcriptional changes. This approach focused on genes with relatively low overall expression but showing significant, transient increases in expression immediately preceding major transcriptional shifts or stage transitions.

First, genes were filtered based on their mean expression across cells within a relevant trajectory (e.g., asexual cycle), retaining those with mean expression below a defined threshold (e.g., the 25th percentile of all gene means) to focus on genes that are not constitutively highly expressed.

For these low-expression genes, their smoothed expression profiles along pseudotime were analyzed for significant peaks. A peak was defined as a local maximum in the smoothed profile where the expression level exceeded a threshold relative to the gene's baseline expression within that trajectory (e.g., a fold change greater than 2 relative to the gene's median expression).

Candidate regulators were prioritized if their transient expression peaks occurred immediately prior to known life cycle stage transitions (as defined by cell labels and marker gene expression) or the activation of major gene modules identified in the previous step (Grishechkin et al. 2025).

Functional annotations from PlasmoDB and other relevant databases were used to filter candidates for known

or predicted regulatory functions (e.g., transcription factors, kinases, phosphatases, chromatin modifiers).

The sets of identified candidate regulators and the timing of their transient expression were compared between lab and field isolates to highlight potential differences in regulatory control.

2.11. Inference of Putative Regulatory Interactions

To gain insight into the potential regulatory networks, a simplified approach was used to infer putative downstream targets of the identified candidate transient regulators. For each high-confidence candidate regulator exhibiting a transient peak in expression, the expression dynamics of other genes were examined in the pseudotime window immediately following the regulator's peak. Genes whose expression significantly changed (either upregulated or downregulated) in the period immediately after the regulator's peak were considered putative targets. This inference was based on the principle that a regulator's transient activation often precedes changes in the expression of its target genes (Kommu et al. 2024; Hegde et al. 2025).

Statistical methods, such as correlation analysis considering a potential time lag or differential expression testing between cells before and after the regulator's peak within a defined pseudotime window, were employed to identify these putative target genes (Kommu et al. 2024; Hegde et al. 2025).

By comparing the sets of inferred targets for conserved and condition-specific regulators between lab and field conditions, potential rewiring of regulatory networks controlling specific developmental transitions was assessed (Kommu et al. 2024; Hegde et al. 2025).

2.12. Statistical Analysis and Reproducibility

All statistical analyses, including differential expression testing, trajectory inference, module detection, and regulator identification, were performed with attention to statistical rigor. Multiple testing correction (e.g., Benjamini-Hochberg FDR) was applied where appropriate to control the rate of false positives (Sarkar & Tang 2021; He et al. 2024; Timans et al. 2025). A significance level, typically $FDR < 0.05$ or $FDR < 0.01$, was used to determine statistical significance. All computational analyses were performed in a reproducible manner, with code and parameters used for each step documented.

3. RESULTS

3.1. Dataset overview and quality control

We analyzed a single-cell RNA sequencing dataset comprising 45,691 individual *Plasmodium falciparum* cells. Following initial processing and quality control

steps, which included filtering genes detected in fewer than three cells, the dataset retained expression profiles for 5,274 genes. These cells originated from two distinct sources: laboratory-adapted strains and field isolates. The majority of cells (37,624, or 82.3%) were derived from laboratory cultures, while 8,067 cells (17.7%) were obtained from field isolates from four asymptomatic patients in Mali. After normalization and log-transformation as described in the Methods, the resulting gene expression matrix exhibited a sparsity of 80.25%, consistent with the nature of single-cell transcriptomic data. To inform subsequent dimensionality reduction, we examined the variance explained by principal components (PCs), as shown in Figure 1.

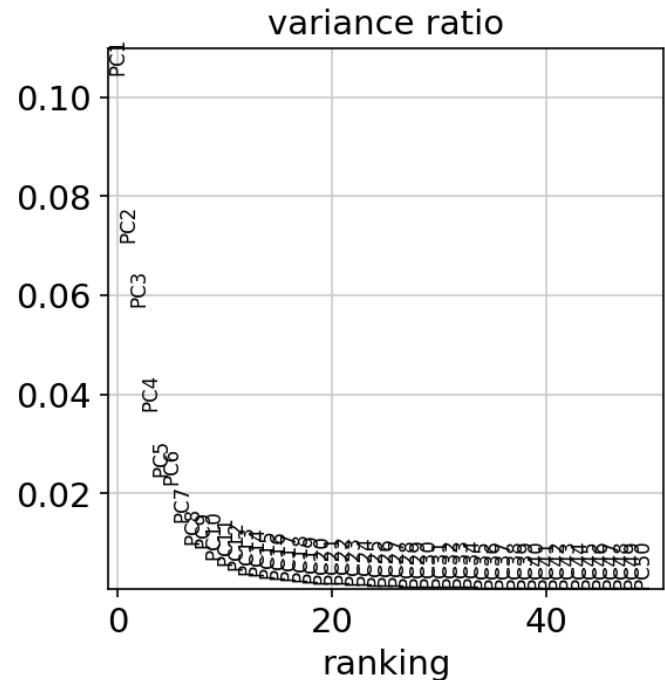


Figure 1. Variance ratio explained by principal components (PCs) ranked by importance. This plot illustrates the proportion of total variance captured by each PC, informing the selection of components used for subsequent dimensionality reduction.

A comparison of basic dataset characteristics revealed notable differences between the lab and field populations (Table 1). Cells from laboratory strains generally displayed a higher median number of detected genes per cell (954 vs. 802) and a higher median total normalized expression per cell (2093.79 vs. 1813.51) compared to field isolates. These differences might reflect variations in parasite health, transcriptional activity, or RNA integrity influenced by the distinct environments of laboratory culture versus the human host.

A critical divergence between the two datasets was observed in the distribution of cells across the parasite's life cycle stages (Table 2). The laboratory dataset provided a comprehensive representation of the asexual intraerythrocytic development cycle (IDC), encompassing early ring, late ring, early trophozoite, late trophozoite, early schizont, and late schizont stages, alongside developing and mature male and female gametocytes. In stark contrast, the field isolates predominantly consisted of sexual stages (gametocytes), with asexual stages limited to late rings and early trophozoites. Early ring, early schizont, and late schizont stages were entirely absent from the field samples. Furthermore, very late-stage gametocytes, specifically annotated as 'LE' (likely representing stage V gametocytes), were exclusively identified within the field isolate population. This striking difference in stage composition highlights that standard laboratory culture conditions, while supporting the asexual cycle and early sexual development, do not fully recapitulate the conditions necessary for progression to the most mature gametocyte stages observed in the human host. This observation underscores the importance of analyzing field isolates to capture the full spectrum of parasite development relevant to transmission.

3.2. The transcriptional landscape of lab and field parasites

To visualize the global transcriptional landscape and assess the relationships between cells, we performed dimensionality reduction using UMAP on the expression data of 2,000 highly variable genes. As shown in Figure 2, when the UMAP embedding was colored by the assigned life cycle stage, cells organized into a structure that faithfully represents the known progression of the *P. falciparum* life cycle. The asexual IDC formed a large, semi-circular arrangement, beginning with rings, transitioning through trophozoites, and culminating in schizonts. Sexual stages branched off from the asexual cycle, forming distinct clusters for developing, female, and male gametocytes. This spatial arrangement validates the quality of the single-cell data and the biological relevance of the provided stage annotations.

Coloring the same UMAP embedding by the source of the cells (lab vs. field), as presented in Figure 3, revealed both shared and distinct regions occupied by the two populations. As anticipated from the cell count distribution described previously, areas corresponding to early asexual stages (early rings, schizonts) were exclusively populated by laboratory cells. Conversely, the clusters representing late-stage (LE) male and female gametocytes were occupied solely by field isolates.

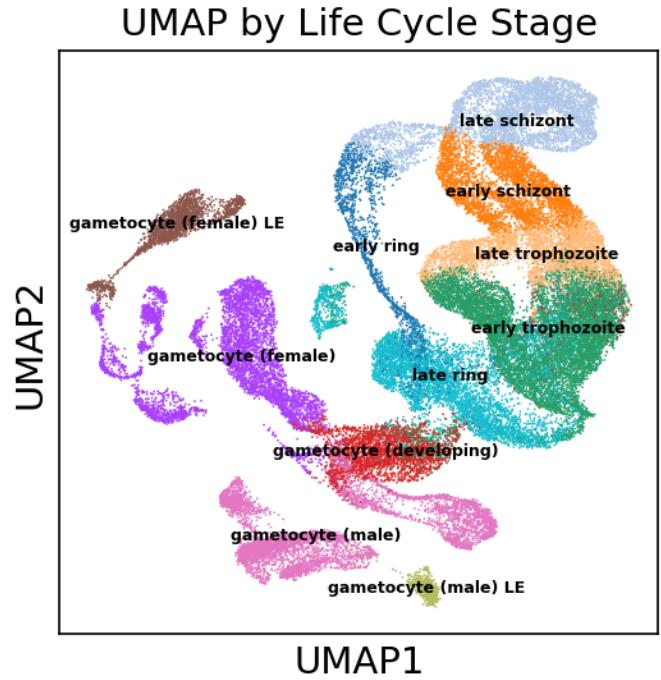


Figure 2. UMAP projection of single-cell *P. falciparum* transcriptomes, colored by annotated life cycle stage. The visualization reveals distinct clusters and a continuous structure reflecting the asexual intraerythrocytic development cycle (IDC) and separate sexual stages, confirming the quality of the data and stage annotations.

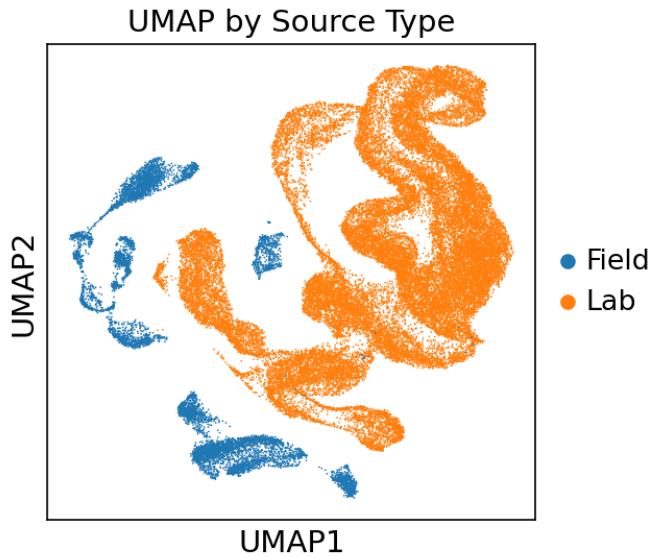


Figure 3. UMAP embedding of single cells colored by source (Field or Lab). The plot shows both shared cell populations and distinct clusters unique to either lab strains or field isolates, highlighting differences in dataset composition.

Importantly, even within stages present in both datasets (late rings, early trophozoites, developing/mature gametocytes), the distributions of lab and field cells, while overlapping, often occupied slightly different spaces within the clusters. Figure 4, which shows the UMAP colored by individual sources, further illustrates this. This visual separation suggested that beyond the differences in stage representation, there were likely underlying transcriptional distinctions between lab and field parasites even within nominally equivalent developmental stages.

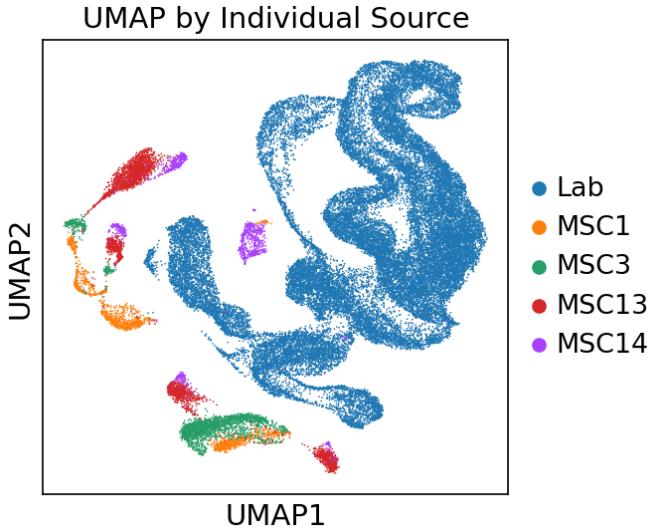


Figure 4. UMAP embedding of *P. falciparum* single-cell transcriptomes colored by source. Cells from laboratory strains (Lab) and individual field isolates (MSC1, MSC3, MSC13, MSC14) are shown. The plot reveals that while cells from both sources co-cluster within shared developmental stages, distinct regions are exclusively populated by either lab or field cells, highlighting differences in dataset composition.

3.3. Stage-specific differential gene expression between lab and field isolates

To quantify the transcriptional differences hinted at by the UMAP visualization, we performed differential gene expression (DGE) analysis comparing laboratory-adapted strains and field isolates within four major life cycle stages where both sources were represented: late ring, early trophozoite, female gametocyte, and male gametocyte.

In the **late ring** stage, comparing 428 field cells to 5,438 lab cells revealed a substantial number of differentially expressed genes (DEGs). The volcano plot in Figure 5 shows a strong asymmetry, with a large proportion of genes significantly upregulated in field iso-

lates compared to lab strains. For example, genes like *PF3D7_1372200* and *PF3D7_0831800* showed significant upregulation in field isolates with log-fold changes (LFCs) of 4.87 and 2.33, respectively (*FDR* < 0.05). This widespread transcriptional divergence at the late ring stage indicates that field parasites are already on a distinct transcriptional path early in their IDC.

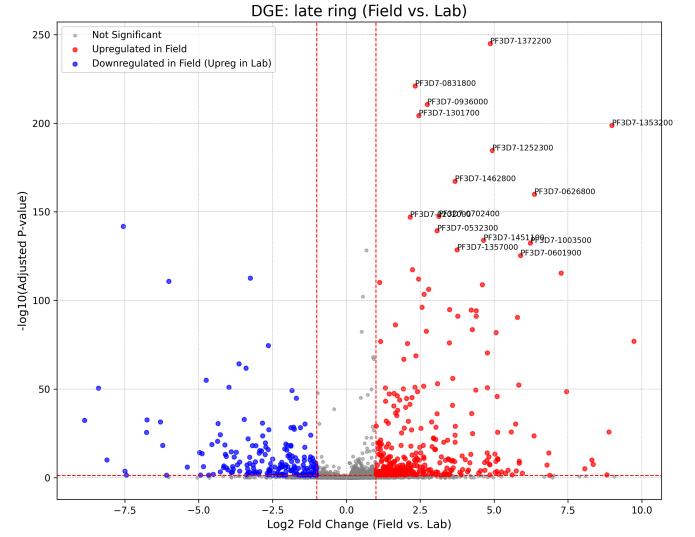


Figure 5. Volcano plot showing differential gene expression (DGE) between field isolates and lab strains at the late ring stage. Log₂ fold change (Field vs. Lab) is plotted against $-\log_{10}$ (adjusted P-value). Red points indicate genes significantly upregulated in field isolates, blue points indicate genes significantly downregulated (upregulated in lab), and gray points are not significant. The plot demonstrates a strong transcriptional divergence at this stage, with a notable skew towards upregulation in field isolates.

The transcriptional differences persisted and were even more pronounced in the **early trophozoite** stage (122 field vs. 9,635 lab cells). As illustrated in the volcano plot in Figure 6, a large number of genes were again significantly differentially expressed, with a strong bias towards upregulation in field isolates. Genes such as *PF3D7_1372200* (LFC = 10.44) and *PF3D7_1001500* (LFC = 5.39) were among the most highly upregulated in field parasites. These substantial differences suggest that the primary growth phase of field isolates is characterized by a unique transcriptional program, potentially influenced by factors encountered in the human host environment that are absent in standard laboratory culture.

The most striking transcriptional divergence was observed within the **sexual stages (gametocytes)**. In female gametocytes (1,656 field vs. 3,903 lab cells), thousands of genes were found to be differentially expressed,

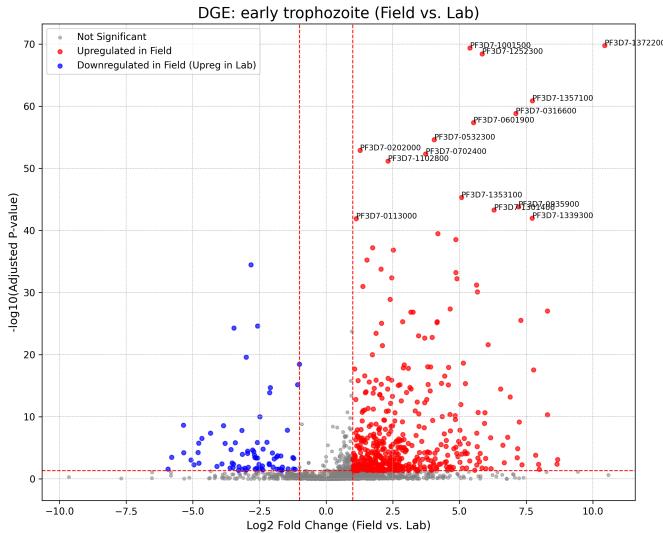


Figure 6. Volcano plot showing differential gene expression between field isolates and laboratory strains in the early trophozoite stage. Each point represents a gene, colored by significance and direction of change (red: significantly upregulated in field, blue: significantly downregulated in field). The plot reveals a strong skew towards gene upregulation in field parasites, highlighting transcriptional differences at this stage.

as depicted in Figure 7. While many genes were upregulated in both directions, the sheer number of significant DEGs highlighted a profound difference in the transcriptional state of female gametocytes from field isolates compared to laboratory strains. Top upregulated genes in field female gametocytes included *PF3D7_1423600* (LFC=5.72) and, intriguingly, the canonical male gametocyte surface protein gene *PF3D7_1031000* (*Pfs25*, LFC=1.22). The upregulation of a male marker in cells annotated as female gametocytes from field isolates could potentially indicate less stringent sexual lineage separation in the field, a different transcriptional state of field female gametocytes that includes expression of some male-associated genes, or subtle differences in cell sorting/annotation between the two sources that reflect *in vivo* heterogeneity.

Similarly, male gametocytes (3,364 field vs. 1,964 lab cells) exhibited extensive transcriptional differences, as seen in Figure 8, with genes like *PF3D7_0205000* (LFC=2.46) and *PF3D7_1201600* (LFC=2.67) showing significant upregulation in field isolates. The widespread and large-magnitude differential expression in both male and female gametocytes underscores that the sexual development program proceeds fundamentally differently in the human host environment compared to laboratory culture.

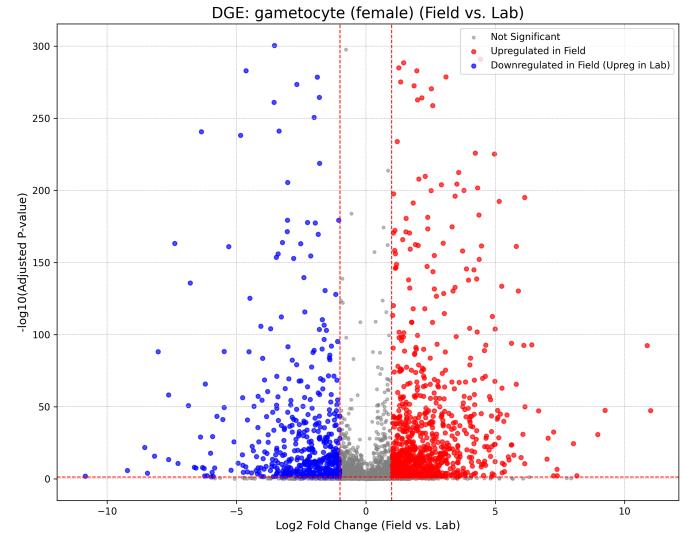


Figure 7. Volcano plot illustrating differential gene expression (DGE) between field and lab isolates of *Plasmodium falciparum* female gametocytes. The x-axis represents the Log₂ fold change (Field vs. Lab), and the y-axis shows the -Log₁₀ adjusted P-value. Points represent individual genes, colored by significance (adjusted P-value < 0.05) and direction of change (red: upregulated in field, blue: downregulated in field/upregulated in lab, gray: not significant). The plot reveals a large number of significantly differentially expressed genes, indicating profound transcriptional differences in female gametocytes between *in vivo* and *in vitro* conditions.

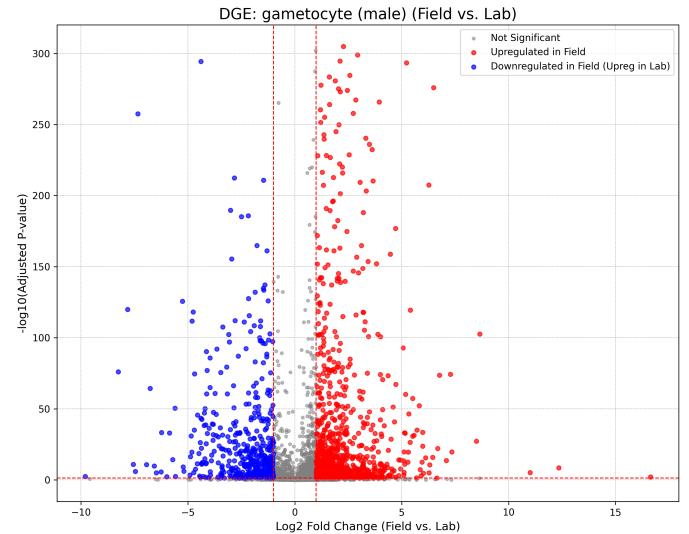


Figure 8. Differential gene expression analysis of male gametocytes from field isolates compared to lab strains. Volcano plot shows log₂ fold change (Field vs. Lab) versus -log₁₀ adjusted p-value. Numerous genes are significantly differentially expressed, revealing substantial transcriptional divergence in male gametocytes between these two sources.

3.4. Developmental trajectories reveal altered progression dynamics

To gain a dynamic perspective on these transcriptional differences and understand how stage transitions unfold, we performed developmental trajectory inference using PAGA and diffusion pseudotime.

For the **asexual development** in laboratory strains, we successfully reconstructed a complete, cyclical trajectory representing the IDC. As shown in the UMAP embedding (Figure 9), anchored at the 'early ring' stage as the root, the pseudotime progression accurately captured the known sequence of development from rings through trophozoites to schizonts.

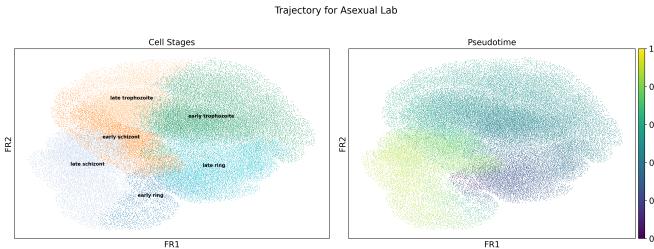


Figure 9. Two-dimensional embedding of single cells from laboratory-adapted *Plasmodium falciparum* asexual stages. Cells are colored by annotated life cycle stage (left) and inferred pseudotime (right). The cyclical arrangement of stages and the smooth progression of pseudotime demonstrate the accurate reconstruction of the intraerythrocytic developmental cycle trajectory in laboratory strains.

The PAGA graph (Figure 10 and Figure 11) confirmed strong connectivity and smooth transitions between consecutive asexual stages, reflecting the continuous nature of this cycle in culture.

For the field isolates, due to the limited stage representation (as detailed in Table 2), the asexual trajectory reconstruction was necessarily incomplete, encompassing only the late ring and early trophozoite stages present in the dataset (Figure 12). While this segment showed the expected pseudotemporal progression within these stages (Figure 13), the absence of early rings and schizonts precluded a full comparison of the entire asexual cycle's dynamics and duration between lab and field.

The **sexual development** trajectories revealed even more pronounced differences. In laboratory strains, the trajectory, rooted at 'gametocyte (developing)', showed a clear bifurcation into distinct mature male and female gametocyte lineages (Figure 14 and Figure 15), consistent with the known sexual differentiation pathway in culture.

In contrast, the trajectory inferred for field isolates, visualized using PAGA (Figure 16 and Figure 17) and a

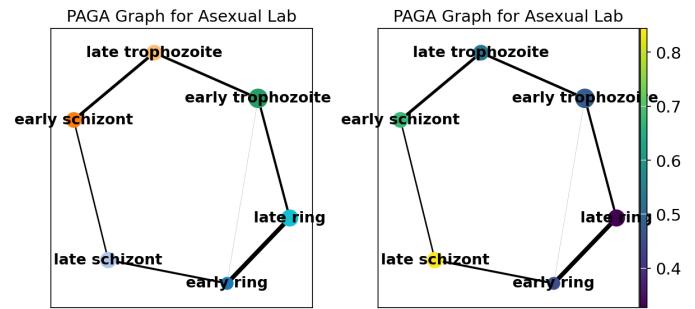


Figure 10. PAGA graph showing the reconstructed asexual life cycle trajectory for laboratory-adapted **Plasmodium falciparum** strains. Nodes represent parasite stages, and edge thickness indicates inferred connectivity strength between stages. The left panel displays stage labels, and the right panel colors the nodes by diffusion pseudotime. The cyclical structure and strong connectivity between consecutive stages confirm the expected progression of the intraerythrocytic development cycle.

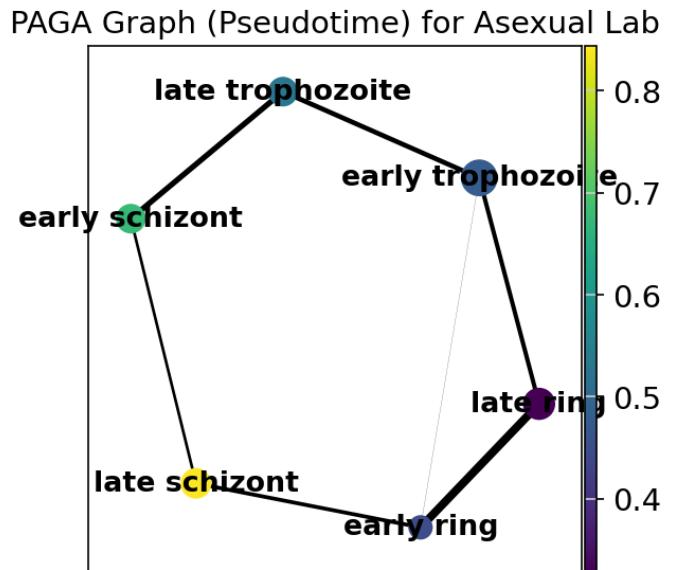


Figure 11. PAGA graph illustrating the reconstructed asexual developmental trajectory for lab-adapted *P. falciparum*. Nodes represent life cycle stages colored by pseudotime, and edge thickness indicates connectivity strength between stages. The cyclical structure and strong edge weights confirm the expected progression and connectivity of the intraerythrocytic development cycle in laboratory culture.

force-directed layout (Figure 18), and rooted at 'gametocyte (female)' (reflecting the initial abundance), presented a more complex structure. This trajectory not only showed the separation into male and female populations but also extended significantly to include the 'LE' (late-stage) gametocyte populations that were unique to the field dataset. The PAGA analysis confirmed strong

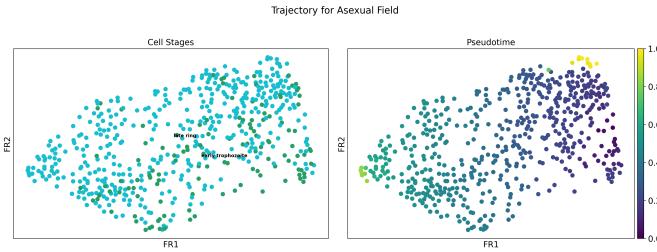


Figure 12. Developmental trajectory of asexual *P. falciparum* field isolates. Cells are embedded in a 2D layout (FR1/FR2) and colored by annotated stage (left) or calculated pseudotime (right). The plot shows the progression from late ring to early trophozoite stages, reflecting the limited asexual stages captured from field samples.

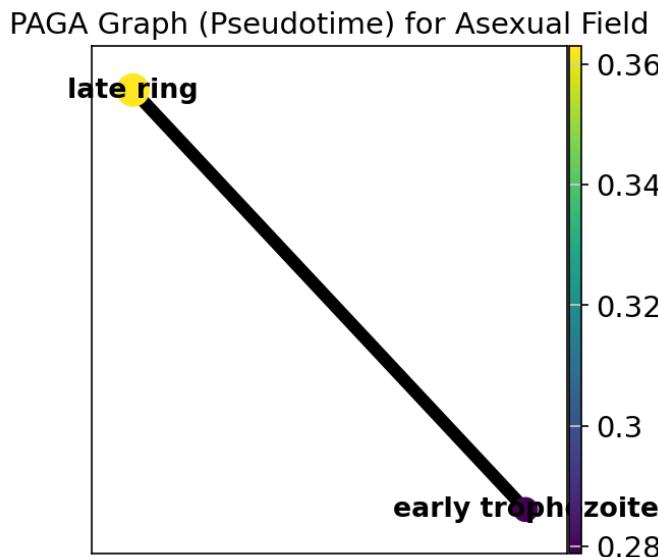


Figure 13. PAGA graph showing the asexual trajectory in field isolates. Strong connectivity between late ring and early trophozoite stages reflects the limited asexual stages captured in the field data.

connections from the main male and female gametocyte clusters to their respective 'LE' counterparts (Figure 17), indicating that these late-stage cells represent a further point in the maturation trajectory occurring in the human host. This finding provides compelling evidence that the sexual development pathway in field isolates is more prolonged or involves distinct late maturation steps that are typically not completed or maintained in standard laboratory culture conditions.

3.5. Identification of candidate master regulators and inferred regulatory networks

To investigate the potential regulatory mechanisms underlying the observed differences in developmental trajectories and transcriptional programs, we applied a

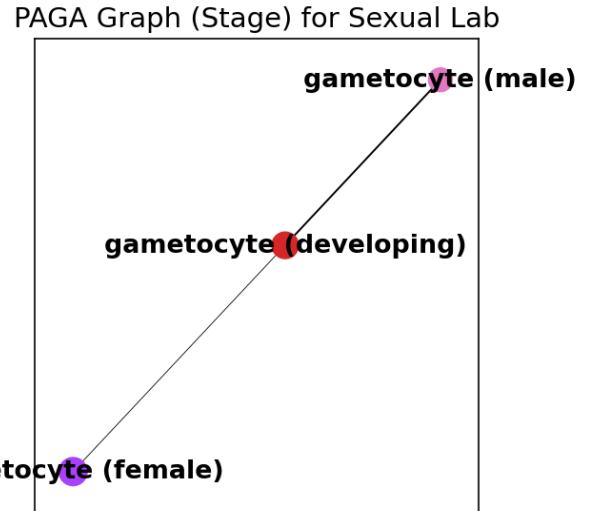


Figure 14. PAGA graph visualizing the sexual development trajectory in laboratory-adapted *P. falciparum* strains. Nodes represent life cycle stages, and edges indicate inferred transcriptional connectivity. The graph shows a clear bifurcation from developing gametocytes into distinct male and female lineages.

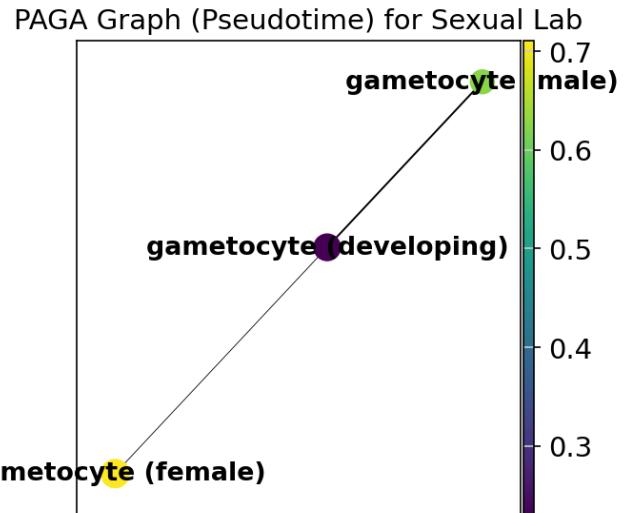


Figure 15. PAGA graph of the sexual development trajectory in lab-adapted *Plasmodium falciparum*. Nodes represent cell stages, colored by pseudotime, with edges indicating connectivity. The graph shows the bifurcation from developing gametocytes into mature female and male lineages.

computational pipeline designed to identify candidate transcriptional regulators exhibiting transient expression peaks that precede major transcriptional shifts or the activation of co-expressed gene modules along pseudotime. We performed this analysis separately for the reconstructed asexual and sexual trajectories in both lab and field datasets.

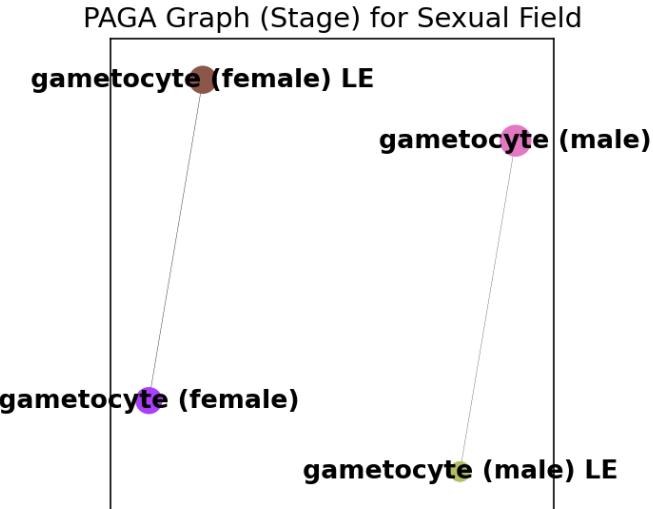


Figure 16. PAGA graph showing inferred connectivity between sexual life cycle stages of *P. falciparum* field isolates. Nodes represent gametocyte stages, including late-stage (LE) forms. Edges indicate strong connections, revealing progression from female and male gametocytes to their respective LE stages in field isolates.

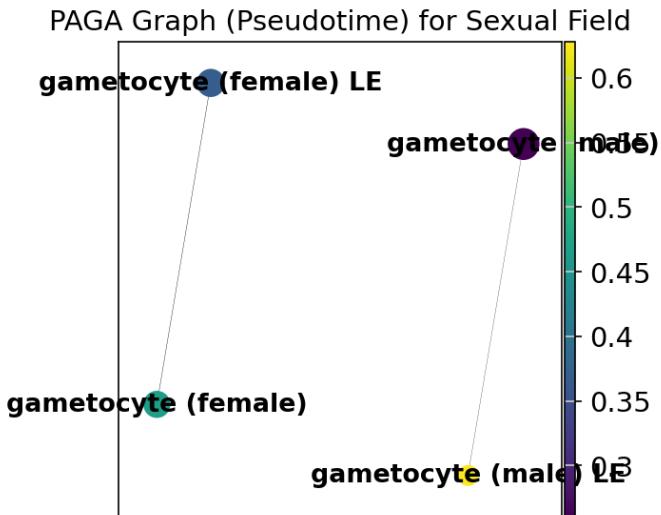


Figure 17. PAGA graph of sexual stage development in field isolates. Nodes represent distinct gametocyte stages, including late-stage ('LE') male and female populations unique to field samples. Edges indicate inferred developmental connections, revealing progression to later stages and an extended developmental pathway in the host environment.

In the **Asexual Lab** trajectory, we identified 17 distinct gene modules exhibiting coordinated expression dynamics along pseudotime. Our analysis of transient regulators predicted a remarkable 3,428 putative regulatory links. A large number of candidate regulators were

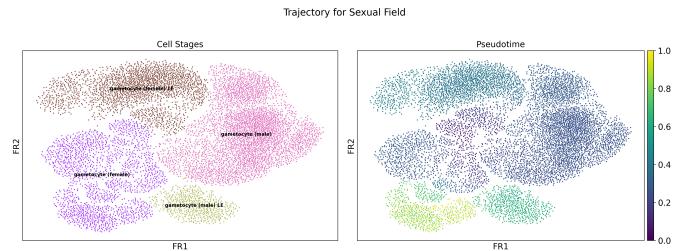


Figure 18. Force-directed layout (FR) of single-cell transcriptomes from sexual stage *Plasmodium falciparum* field isolates colored by annotated life cycle stage (left) and inferred developmental pseudotime (right). The trajectory reveals progression from earlier gametocytes to late-stage (LE) populations, illustrating extended maturation observed only in field isolates.

predicted to have peak expression at the very beginning of the trajectory (pseudotime bin 0), immediately preceding the activation of multiple gene modules in subsequent bins (e.g., bin 1). This pattern suggests that the initiation of the asexual IDC in laboratory culture is driven by a major, coordinated wave of regulatory activity involving numerous factors acting early in the cycle.

For the **Asexual Field** trajectory (limited to late ring and early trophozoite stages), we identified 13 gene modules. However, the analysis predicted significantly fewer putative regulatory links (945) compared to the lab asexual trajectory. The timing of peak expression for top candidate regulators in the field asexual trajectory appeared different, with some predicted to peak later in the observed trajectory segment (e.g., bin 5). An example showing the expression profile of a candidate regulator and its target module activation along pseudotime is presented in Figure 19. This suggests that the regulatory triggers and dynamics within the observed late ring and early trophozoite stages may differ in field parasites, or that the key initiating regulatory events of the asexual cycle occur in the unobserved early ring stages.

Analysis of the **Sexual Lab** trajectory identified 13 gene modules and 404 putative regulatory links. Similar to the asexual lab trajectory, many candidate regulators were found to exhibit peak expression at the earliest pseudotime bins of the sexual trajectory (Figure 20 shows an example), suggesting a concentrated regulatory cascade initiating sexual commitment and early gametocyte development in culture.

The **Sexual Field** trajectory presented a dramatically different picture of the regulatory landscape, aligning with the expanded trajectory and transcriptional divergence observed in gametocytes. We identified 18 gene modules, and the analysis predicted 1,917 putative regulatory links. This represents a nearly five-fold increase in

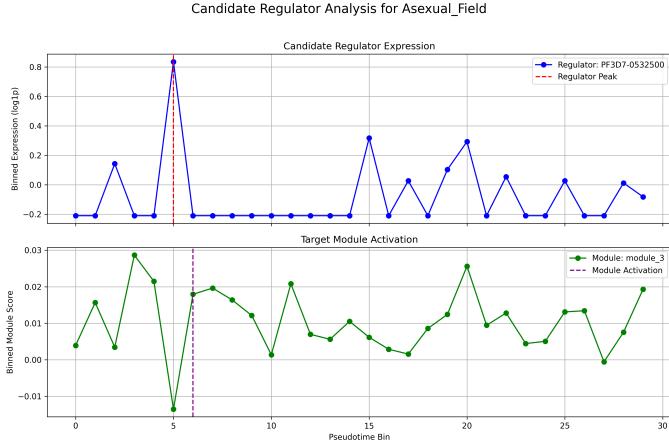


Figure 19. Candidate regulator expression and target module activation along the asexual field trajectory. Top panel shows the binned expression profile of candidate regulator *PF3D7_0532500* across pseudotime bins, with its peak expression indicated by a dashed red line. The bottom panel shows the corresponding binned activation score for target module 3, with the inferred module activation time marked by a dashed purple line. This example illustrates a temporal relationship where the candidate regulator's peak expression precedes the activation of its putative target module, supporting the inferred regulatory link.

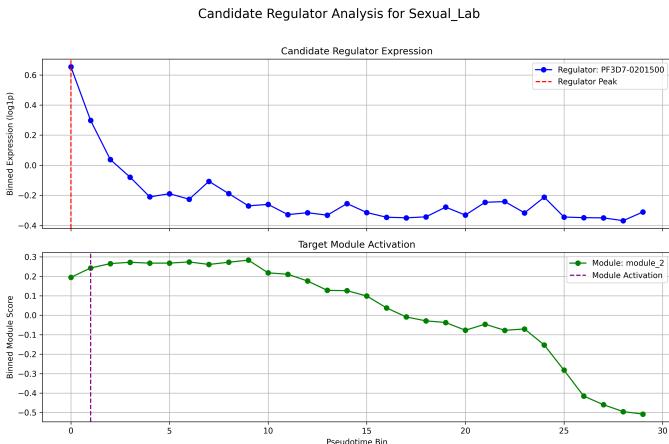


Figure 20. Example of an inferred regulatory relationship in lab sexual stage parasites. The upper panel shows the binned expression of candidate regulator *PF3D7_0201500* along the sexual development pseudotime trajectory. The lower panel shows the corresponding binned activation score of target gene module 2. The peak expression of the candidate regulator precedes the activation of the gene module, supporting a potential regulatory link during sexual development in laboratory culture.

predicted regulatory interactions compared to the sexual lab trajectory, suggesting a substantially more complex or finely tuned regulatory program governing gametocytogenesis *in vivo*. Furthermore, candidate regulators

were predicted to act later in the trajectory compared to the lab, with some peaking in later pseudotime bins (e.g., bin 17) before the activation of downstream modules (e.g., bin 18). Figure 21 provides an example of such a relationship. This timing likely corresponds to the regulatory events controlling the progression to the late-stage 'LE' gametocytes that are unique to the field environment.

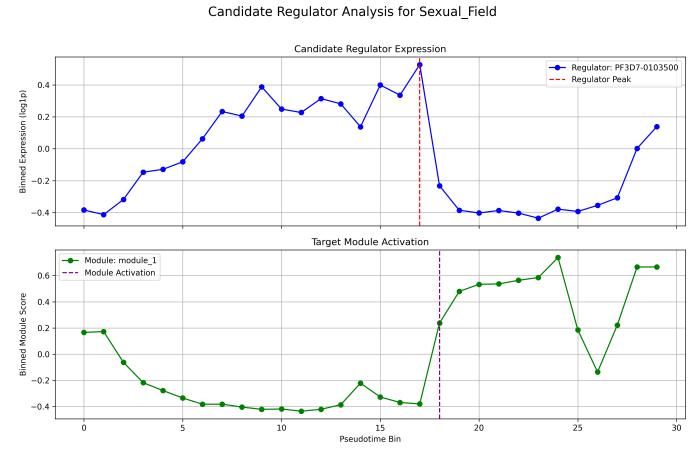


Figure 21. Expression profile of candidate regulator *PF3D7_0103500* (blue) and activation score of target gene module_1 (green) across pseudotime bins for the Sexual Field trajectory. The dashed red line indicates the regulator peak, and the dashed purple line indicates target module activation, illustrating how regulator expression precedes module activation.

Collectively, the regulatory network analysis indicates significant rewiring of the transcriptional control mechanisms governing parasite development in response to adaptation. Field isolates appear to employ altered developmental trajectories and are regulated by a distinct set of timed regulatory events, with a particularly pronounced increase in regulatory complexity observed during sexual development, potentially facilitating progression through the full maturation process required for transmission in the human host. The candidate transient regulators identified in this analysis provide a rich set of testable hypotheses for future studies aimed at dissecting the molecular mechanisms of parasite adaptation and stage-specific control.

4. CONCLUSIONS

Malaria caused by *Plasmodium falciparum* remains a major health challenge, driven by the parasite's complex life cycle and ability to adapt to diverse environments. Understanding the dynamic transcriptional programs and regulatory networks that govern developmental stage transitions is crucial, particularly how these dif-

fer between laboratory-adapted strains and field isolates reflecting adaptation to the human host. This study addressed this knowledge gap by employing single-cell RNA sequencing to provide a high-resolution, dynamic comparison of development in over 45,000 cells from lab strains and field isolates.

We utilized advanced computational methods including quality control, normalization, dimensionality reduction (UMAP), differential gene expression analysis, developmental trajectory inference (PAGA, pseudotime), gene module identification, and the inference of candidate transient transcriptional regulators and their putative targets. This comprehensive approach allowed us to dissect the transcriptional landscape and regulatory dynamics underlying parasite development in different adaptive contexts.

Our analysis revealed significant differences between laboratory-adapted strains and field isolates across multiple levels. Basic cellular metrics and stage composition showed stark contrasts, with field isolates possessing a much higher proportion of sexual stages and uniquely including late-stage gametocytes. The transcriptional landscape, visualized by UMAP, demonstrated clear separation between lab and field cells, even within shared life cycle stages. Differential gene expression analysis confirmed extensive transcriptional divergence within late ring, early trophozoite, and sexual stages, often with a bias towards upregulation in field parasites, indicating distinct transcriptional programs are active *in vivo*.

Developmental trajectory inference highlighted divergent paths, most notably an extended and distinct sex-

ual development program in field isolates culminating in late-stage gametocytes not observed in standard laboratory culture. This finding provides a dynamic view of how the human host environment supports the full maturation necessary for transmission. Furthermore, analysis of co-expressed gene modules and candidate transient regulators indicated substantial rewiring of the regulatory networks. The sexual development trajectory in field isolates was associated with a significantly larger number of predicted regulatory interactions and later-acting regulators compared to laboratory strains, suggesting a more complex and potentially prolonged regulatory landscape governing gametocytogenesis *in vivo*.

From these results, we learned that adaptation to the laboratory environment leads to significant divergence in *P. falciparum* development, characterized by altered transcriptional programs, truncated sexual development trajectories, and substantial rewiring of regulatory networks compared to field isolates circulating in human populations. Field isolates capture the full spectrum of sexual development and appear to employ a more complex regulatory strategy for gametocyte maturation. These findings underscore the importance of studying field isolates to fully understand parasite biology relevant to transmission and identify key transcriptional regulators that potentially mediate host-specific developmental control. This study provides a valuable resource and a foundation for future investigations into the molecular mechanisms of parasite adaptation and stage-specific regulation.

REFERENCES

- Ali, M. U., Ahmed, S., Ferzund, J., Mehmood, A., & Rehman, A. 2017, Using PCA and Factor Analysis for Dimensionality Reduction of Bio-informatics Data, doi: <https://doi.org/10.14569/IJACSA.2017.080551>
- Birkholtz, L. M., Bastien, O., Wells, G., et al. 2006, Integration and mining of malaria molecular, functional and pharmacological data: how far are we from a chemogenomic knowledge space?, doi: <https://doi.org/10.1186/1475-2875-5-110>
- Cao, P., Klonis, N., Zaloumis, S., et al. 2016, A dynamic stress model explains the delayed drug effect in artemisinin treatment of Plasmodium falciparum, doi: <https://doi.org/10.1128/AAC.00618-17>
- Chen, X., Lin, S., Chen, X., Li, W., & Li, Y. 2025, Timestamp calibration for time-series single cell RNA-seq expression data. <https://arxiv.org/abs/2412.03027>
- Dajani, S. A. A., Sanchez, A., & Williams, J. R. 2025, DeepSeq: High-Throughput Single-Cell RNA Sequencing Data Labeling via Web Search-Augmented Agentic Generative AI Foundation Models. <https://arxiv.org/abs/2506.13817>
- Dandala, B., Danziger, M. M., Barkan, E., et al. 2025, BMFM-RNA: An Open Framework for Building and Evaluating Transcriptomic Foundation Models. <https://arxiv.org/abs/2506.14861>
- Ditz, J. C., Wistuba-Hamprecht, J., Maier, T., et al. 2023, PlasmoFAB: A Benchmark to Foster Machine Learning for Plasmodium falciparum Protein Antigen Candidate Prediction. <https://arxiv.org/abs/2301.06454>
- Flores, M., Liu, Z., Zhang, T.-H., et al. 2021, Deep learning tackles single-cell analysis A survey of deep learning for scRNA-seq analysis. <https://arxiv.org/abs/2109.12404>

- Grishechkin, A., Mukherjee, A., & Karin, O. 2025, Hierarchical Control of State Transitions in Dense Associative Memories. <https://arxiv.org/abs/2412.11336>
- He, J., Gang, B., & Fu, L. 2024, False Discovery Control in Multiple Testing: A Brief Overview of Theories and Methodologies. <https://arxiv.org/abs/2411.10647>
- Hegde, A., Nguyen, T., & Cheng, J. 2025, Machine Learning Methods for Gene Regulatory Network Inference. <https://arxiv.org/abs/2504.12610>
- Hutton, A., & Meyer, J. G. 2025, Trajectory Inference for Single Cell Omics. <https://arxiv.org/abs/2502.09354>
- Kommu, S., Wang, Y., Wang, Y., & Wang, X. 2024, Gene Regulatory Network Inference from Pre-trained Single-Cell Transcriptomics Transformer with Joint Graph Learning. <https://arxiv.org/abs/2407.18181>
- Li, G.-D. 2018, Cell-Cycle-Associated Amplified Genomic-DNA Fragments (CAGFs) Might be Involved in Chloroquine Action and Resistance in Plasmodium falciparum. <https://arxiv.org/abs/1702.02076>
- Motta, F. C., McGoff, K., Cummins, B., & Haase, S. B. 2024, Generalized Measures of Population Synchrony. <https://arxiv.org/abs/2406.15987>
- Patel, M., Magre, N., Motwani, H., & Brown, N. B. 2024, Advances in Machine Learning, Statistical Methods, and AI for Single-Cell RNA Annotation Using Raw Count Matrices in scRNA-seq Data. <https://arxiv.org/abs/2406.05258>
- Riffle, D., Shirooni, N., He, C., et al. 2025, OLAF: An Open Life Science Analysis Framework for Conversational Bioinformatics Powered by Large Language Models. <https://arxiv.org/abs/2504.03976>
- Sarkar, S. K., & Tang, C. Y. 2021, Adjusting the Benjamini-Hochberg method for controlling the false discovery rate in knockoff assisted variable selection. <https://arxiv.org/abs/2102.09080>
- Timans, A., Straehle, C.-N., Sakmann, K., & Nalisnick, E. 2025, A powerful rank-based correction to multiple testing under positive dependency. <https://arxiv.org/abs/2311.10900>
- Zhang, Z., Sun, Y., Peng, Q., Li, T., & Zhou, P. 2025, Integrating Dynamical Systems Modeling with Spatiotemporal scRNA-seq Data Analysis, doi: <https://doi.org/10.3390/e27050453>
- Zhou, Z., Li, J., Xin, H., Pan, X., & Shen, H.-B. 2024, Simultaneously Infer Cell Pseudotime, Velocity Field and Gene Interaction from Multi-Branch scRNA-seq Data with scPN. <https://arxiv.org/abs/2410.18394>
- Zitnik, M., Li, M. M., Wells, A., et al. 2024, Current and future directions in network biology, doi: <https://doi.org/10.1093/bioadv/vbae099>