Draft version June 25, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Divergent Transcriptional Programs and Regulatory Networks Govern Plasmodium falciparum Development in Laboratory-Adapted Strains and Field Isolates

Denario[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

ABSTRACT

Laboratory adaptation can significantly alter *Plasmodium falciparum* biology, impacting the relevance of research findings. To understand these effects, we investigated differences in the dynamic transcriptional programs and regulatory networks governing stage transitions between lab-adapted strains and field isolates. Using single-cell RNA sequencing data from 45,691 cells, including both lab strains and field isolates from asymptomatic patients, we reconstructed and compared developmental trajectories, performed differential gene expression analysis, and identified co-expression modules and candidate regulators. Our analysis revealed substantial differences in transcriptional profiles, developmental trajectories, and regulatory networks between lab and field parasites, particularly during sexual development. We observed distinct expression patterns, alternative developmental routes in field isolates leading to late-stage gametocytes absent in lab strains, and a rewiring of regulatory networks. Specifically, we identified a unique set of candidate master regulators and inferred regulatory interactions in field isolates, suggesting adaptation to *in vivo* conditions alters developmental control and fate determination. These findings highlight the importance of studying field isolates to fully understand *P. falciparum* biology and the molecular mechanisms underlying parasite adaptation to the human host.

Keywords: Astronomy data visualization, Astronomy data modeling, Computational astronomy, Astronomy data acquisition, Astronomy data reduction

1. INTRODUCTION

Plasmodium falciparum, the most virulent malaria parasite, undergoes a complex life cycle with distinct developmental stages in both the mosquito vector and the human host. A comprehensive understanding of the molecular mechanisms governing these stage transitions is crucial for developing effective interventions that target parasite development and transmission, ultimately aiming to control and eradicate malaria. However, a significant challenge arises from the observation that *P. falciparum* undergoes biological changes when adapted to laboratory culture. This adaptation can alter gene expression, metabolism, and even developmental programs, potentially compromising the relevance of research findings obtained solely from lab-adapted strains when applied to the complexities of natural infections. Understanding how laboratory adaptation affects parasite biology and identifying the key differences between lab-adapted strains and field isolates is a major hurdle. Field isolates, which represent the parasite populations circulating in infected individuals, reflect adaptation to the selective pressures of the *in vivo* environment, including host immune responses, drug treatments, and nutrient availability. However, detailed molecular characterization of field isolates is often limited by their genetic diversity and the challenges of obtaining sufficient parasite material from clinical samples. Disentangling the effects of laboratory adaptation from inherent differences between parasite strains requires careful experimental design and robust analytical approaches.

In this study, we directly address this challenge by comparing the dynamic transcriptional programs and inferred regulatory networks that govern stage transitions in *P. falciparum* between laboratory-adapted strains and field isolates using single-cell RNA sequencing (scRNA-seq). We hypothesized that adaptation to lab culture leads to alterations in the timing, path, or efficiency of developmental processes, particularly those related to parasite survival and adaptation within the human host. To test this, we leverage a dataset of 45,691 single-cell transcriptomes from both lab-adapted strains and field isolates derived from asymptomatic patients. This allows us to reconstruct and compare developmen-

tal trajectories, characterize gene modules, and identify candidate regulators at an unprecedented resolution.

Our approach involves reconstructing developmental trajectories to identify differences in the timing, path, or efficiency of transitions (e.g., asexual cycle progression, sexual commitment). We then characterize gene modules and candidate regulators associated with these dynamic programs to understand how adaptation to lab culture or *in vivo* conditions alters developmental control and fate determination. To verify our findings, we perform differential gene expression analysis between lab and field isolates within each major life cycle stage, identify stage-specific marker genes, and perform functional enrichment analysis of co-expression modules. We then identify low expression genes with transient increases prior to major transcriptional shifts. Finally, we infer regulatory interactions based on the temporal dynamics of gene expression along pseudotime, allowing us to identify potential master regulators and their downstream targets in both lab and field parasites. By comparing these inferred regulatory networks, we can identify potential rewiring events that contribute to the phenotypic differences between lab-adapted and field parasites. Our analysis reveals substantial differences in transcriptional profiles, developmental trajectories, and regulatory networks between lab and field parasites, particularly during sexual development. These findings emphasize the importance of studying field isolates to fully understand *P. falciparum* biology and the molecular mechanisms underlying parasite adaptation to the human host.

2. METHODS

This study investigates the differences in dynamic transcriptional programs and inferred regulatory networks controlling stage transitions in *Plasmodium falciparum* between laboratory-adapted strains and field isolates. We leverage a dataset of 45,691 single-cell transcriptomes from both lab-adapted strains and field isolates derived from asymptomatic patients to reconstruct and compare developmental trajectories, characterize gene modules, and identify candidate regulators.

2.1. *Data Loading, Preprocessing, and Quality Control*

2.1.1. *Data Ingestion*

The gene expression matrix from ‘gene_expression.csv‘ and the cell metadata from ‘labels.csv‘ were loaded into a Python environment

2.1.2. *Data Integration*

The gene expression data was merged with the cell metadata using the common ‘CELL_ID‘ column. The gene expression matrix was transposed such that rows represented cells and columns represented genes.

2.1.3. *Initial Filtering and Quality Control*

Genes with very low expression across all cells were removed. Specifically, genes detected in fewer than 3 cells were filtered out. This threshold was determined based on the distribution of gene detection rates. Although the data was pre-processed, we examined standard quality control metrics. The number of genes detected per cell and the total normalized expression per cell were calculated. Cells that were outliers in these distributions (e.g., very few genes detected or extremely low/high total expression) might be considered for removal, although we primarily relied on the provided metadata for cell quality. A log-transformation (log(X+1)) was applied to the normalized expression values to stabilize variance and make data more symmetric for downstream analyses.

2.2. *Exploratory Data Analysis (EDA)*

EDA was performed to understand the dataset's structure and characteristics.

2.2.1. *Dataset Overview Statistics*

The total number of cells and genes after initial filtering were calculated. The number of unique *P. falciparum* 3D7 gene IDs was determined. The sparsity of the expression matrix (percentage of zero values) was also calculated.

2.2.2. *Cell Metadata Summary*

The number of cells for each category in ‘labels.csv‘ was tabulated, including life cycle stage (e.g., Ring, Trophozoite, Schizont, Gametocyte), parasite strain, source (lab vs. field isolates: MSC1, MSC3, MSC13, MSC14), and days in culture (for lab strains). Cell counts per life cycle stage were summarized, stratified by source (lab total vs. field total, and individual field isolates).

2.2.3. *Expression Data Characteristics*

Per-cell metrics, including the number of genes expressed and total normalized expression, were calculated. Per-gene metrics, including the number of cells expressing the gene and the mean/median expression across cells, were also calculated. These statistics were compared between lab and field isolates.

2.3. *Data Normalization, Dimensionality Reduction, and Visualization*

2.3.1. *Normalization and Scaling*

A log-transformation (log1p) was applied to the normalized expression data. Highly variable genes (HVGs) were identified across all cells using the

‘scanpy.pp.highly_variable_genes‘ function, with default parameters. These HVGs were used for dimensionality reduction. The data (for HVGs) was scaled to have zero mean and unit variance for each gene using ‘scanpy.pp.scale‘.

2.3.2. *Dimensionality Reduction*

Principal Component Analysis (PCA) was performed on the scaled HVGs using ‘scanpy.tl.pca‘. The number of principal components (PCs) to retain was determined based on an elbow plot of variance explained, retaining the top 30 PCs. Uniform Manifold Approximation and Projection (UMAP) was applied on the selected PCs for 2D visualization of the cellular landscape using ‘scanpy.tl.umap‘.

2.3.3. *Visualization*

UMAP plots were generated using ‘scanpy.pl.umap‘ and cells were colored by life cycle stage, source (lab vs. field), parasite strain, individual patient ID (for field isolates), and days in culture (for lab strains).

2.4. *Cell Stage Annotation and Refinement*

The ”life cycle stage” column from ‘labels.csv‘ was used as the primary annotation. The consistency of these labels was verified by examining the expression of known marker genes for *P. falciparum* stages in the UMAP embeddings.

2.5. *Differential Gene Expression (DGE) Analysis*

2.5.1. *Lab vs. Field Isolate Comparisons*

For each major life cycle stage present in both lab and field isolates (e.g., Ring, Trophozoite, Schizont, early/late Gametocytes if distinguishable), DGE analysis was performed to identify genes that were significantly different between lab-cultured parasites and field isolates. The Wilcoxon rank-sum test was used, correcting for multiple comparisons using the Benjamini-Hochberg FDR. DGE was performed using ‘scanpy.tl.rank_genes_groups‘.

2.5.2. *Stage-Specific Marker Gene Identification*

Separately for lab strains and field isolates, marker genes for each life cycle stage were identified by comparing cells of that stage against all other cells within the same source group (lab or field). This was performed using ‘scanpy.tl.rank_genes_groups‘, specifying each stage as a group against the rest.

2.6. *Developmental Trajectory Inference and Pseudotime Analysis*

2.6.1. *Asexual Cycle Trajectory*

Cells belonging to the asexual intraerythrocytic cycle (Ring, Trophozoite, Schizont) were subset. Trajectory inference was performed using Monocle 3. Trajectories were reconstructed separately for all lab strains combined and all field isolates combined. The trajectories were rooted based on biological knowledge, with the Ring stage as the earliest point. Pseudotime values were assigned to each cell along the inferred trajectory.

2.6.2. *Sexual Development Trajectory*

If sufficient numbers of gametocytes at different developmental stages were present, trajectories for sexual commitment and development were reconstructed, again separately for lab and field isolates. This involved identifying a branch point from the asexual cycle.

2.7. *Comparative Analysis of Developmental Trajectories*

2.7.1. *Trajectory Topology Comparison*

The overall structure (linearity, branching points) of the inferred trajectories between lab and field isolates was compared.

2.7.2. *Gene Expression Dynamics Along Pseudotime*

For genes identified as differentially expressed or as stage markers, their expression was modeled and plotted as a function of pseudotime. These dynamic profiles were compared between lab and field trajectories to identify differences in the timing of gene activation/repression, the magnitude of expression changes, and the persistence of expression. Generalized Additive Models (GAMs) were used to model gene expression along pseudotime, implemented using the ‘mgcv‘ package in R, called from within the Python analysis pipeline.

2.7.3. *Transition Efficiency/Robustness (Qualitative Assessment)*

Differences in the ”smoothness” or ”discreteness” of transitions between stages were assessed by examining cell density along pseudotime and the variance of gene expression at transition points.

2.8. *Identification and Characterization of Gene Modules*

2.8.1. *Co-expression Module Detection*

Modules of co-expressed genes were identified along the pseudotime trajectories for both lab and field isolates. Genes were clustered based on their smoothed pseudotemporal expression patterns using hierarchical clustering on fitted profiles.

2.8.2. *Module Characterization*

For each identified gene module, functional enrichment analysis was performed using Gene Ontology (GO) terms, KEGG pathways, and other relevant databases available through PlasmoDB to understand their biological roles. The 'gprofiler2' package in R was used for functional enrichment analysis.

2.8.3. *Comparative Module Analysis*

The composition and pseudotemporal dynamics of gene modules were compared between lab and field isolates. Modules that were conserved, condition-specific, or showed altered timing were identified.

2.9. *Identification of Candidate Master Regulators*

This step focused on identifying genes fitting the specific profile of interest: low overall expression with transient increases prior to major transcriptional shifts.

2.9.1. *Filtering for Low-Expression Genes*

The mean expression for each gene across all cells within a relevant lineage (e.g., asexual cycle cells) was calculated. Genes with mean expression below the 25th percentile of all gene means were selected to focus on generally low-expressed genes.

2.9.2. *Identifying Transient Expression Peaks along Pseudotime*

For each selected low-expression gene, its expression profile along the pseudotime axis was smoothed for both lab and field trajectories using Generalized Additive Models (GAMs). Significant peaks in the smoothed expression profiles were identified. A peak was defined as a local maximum where the expression level was significantly higher than the gene's baseline expression (e.g., >2-fold increase over its own median expression within that trajectory).

2.9.3. *Correlating Peaks with Stage Transitions/Module Activation*

It was determined if the identified expression peaks for candidate regulators occurred immediately prior to known life cycle stage transitions (defined by cell labels and marker genes) and the upregulation of major gene modules. "Immediately prior" was defined as a specific window in pseudotime (e.g., 10% of the total pseudotime range).

2.9.4. *Prioritization of Candidates*

Candidate regulators that are known or predicted transcription factors (e.g., ApiAP2 family members), kinases, phosphatases, or other regulatory proteins were prioritized. This involved mapping gene IDs to functional annotations from PlasmoDB or other relevant databases.

2.9.5. *Comparative Regulator Analysis*

The sets of candidate regulators identified in lab strains versus field isolates were compared. Differences in which regulators were active, their peak timing, or the magnitude of their transient expression were noted.

2.10. *Inferred Regulatory Interactions (Simplified)*

2.10.1. *Identifying Potential Downstream Targets*

For each high-confidence candidate regulator, the expression dynamics of other genes in the pseudotime window immediately following the regulator's expression peak were analyzed. Genes whose expression significantly changed (upregulation or downregulation) after a regulator's peak were considered putative targets. Correlation analysis (potentially with a time lag) or differential expression testing between cells "before" and "after" the regulator's peak (within a defined pseudotime window) was used.

2.10.2. *Comparing Putative Regulatory Links*

These inferred regulator-target relationships between lab and field conditions were compared to see if regulatory networks appeared rewired.

All analyses were performed ensuring reproducibility, with code and parameters documented using tools like Jupyter notebooks and GitHub. Statistical significance was assessed using p-values adjusted for multiple comparisons (Benjamini-Hochberg FDR < 0.05).

3. RESULTS

3.1. *Dataset Overview and Quality Control*

The single-cell transcriptomic dataset comprises 45,691 individual *Plasmodium falciparum* cells, profiled across 5,274 genes that were detected in at least three cells. The data originates from two distinct sources: 37,624 cells (82.3%) from laboratory-adapted strains and 8,067 cells (17.7%) from field isolates recovered from four asymptomatic patients in Mali. Following preprocessing, which included log-transformation of normalized expression values, the resulting gene expression matrix exhibited a sparsity of 80.25%.

A summary of the dataset characteristics reveals key differences between the lab and field populations (Table 1). Cells from lab strains showed a higher median number of genes detected per cell (954) and higher median total normalized expression (2093.79) compared to field isolates (802 genes and 1813.51 total expression, respectively). This may reflect differences in cell viability,

RNA integrity, or inherent biological properties between parasites adapted to stable *in vitro* culture versus those from a dynamic *in vivo* environment.

Table 1. Dataset Summary Statistics

Metric	Overall Value	Lab Strains Valu
Number of Cells	45691.00	37624.00
Number of Genes (post-filt)	5274.00	5274.00
Median Genes per Cell	937.00	954.00
Median Norm. Expr. per Cell	2058.59	2093.79

The distribution of cells across the parasite life cycle stages also differed substantially between the two sources (Table 2). The lab-derived dataset contains a comprehensive representation of the asexual intraerythrocytic development cycle (IDC), including early/late rings, trophozoites, and schizonts. In contrast, the field isolates were predominantly composed of sexual stages (gametocytes), with only late ring and early trophozoite stages representing the asexual cycle. Notably, early ring, early schizont, and late schizont stages were entirely absent from the field samples. Furthermore, very mature, late-stage gametocytes (annotated as 'LE') were exclusively found in the field isolates, suggesting that the *in vivo* environment supports progression to stages of sexual development not fully recapitulated in the laboratory culture conditions used here.

Table 2. Cell Distribution by Source and Life Cycle Stage

life_cycle_stage	early ring	early schizont	early trophozoite	gametocyte (developing)	gametocyte (female)	gametocyte (female, late)
Field	0	0	122	3091	3559	1962
Lab	1633	3804	9635	0	0	0
Total	1633	3804	9757	3091	3559	1962

3.2. *The Transcriptional Landscape of Lab and Field Parasites*

To visualize the cellular landscape, we performed dimensionality reduction using UMAP on the 2,000 most highly variable genes. The scree plot in Figure 1 illustrates the variance explained by each principal component, highlighting that the first few components capture most of the variance in the dataset.

The UMAP visualization in Figure 2, where cells are colored by their annotated life cycle stage, shows that cells organize into a distinct, continuous structure reflecting the known progression of parasite development. A large, cyclical arrangement corresponds to the asexual IDC, progressing from ring stages through trophozoites to schizonts. Separate clusters branching off this cycle represent the sexual stages, with clear separation between male and female gametocytes. This organization confirms the high quality of the data and the validity of the provided cell stage labels.

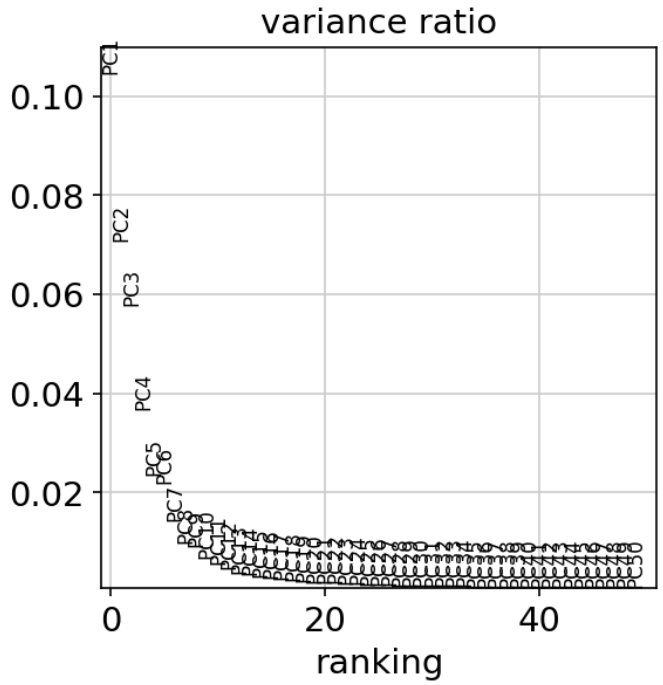

Figure 1. Scree plot showing the variance explained by each principal component. The first few components explain most of the variance in the dataset.

Figure 3 provides a UMAP embedding of the single-cell transcriptomes, colored by sample source (lab strains and field isolates). The distinct clustering patterns reveal differences in cellular composition and transcriptional profiles between lab-adapted parasites and field isolates.

Coloring the same UMAP embedding by source (lab vs. field) as shown in Figure 4 reveals that while cells from both origins co-cluster within shared life cycle stages, there are also regions of the embedding occupied exclusively by one source. As expected from the cell counts, the early ring and schizont regions of the asexual cycle are populated only by lab cells, whereas the clusters corresponding to late-stage gametocytes are populated only by field isolates. This visualization reinforces the distinct cellular compositions of the two datasets and suggests that even within shared stages, subtle transcriptional differences may exist.

3.3. *Stage-Specific Differential Gene Expression Between Lab and Field Isolates*

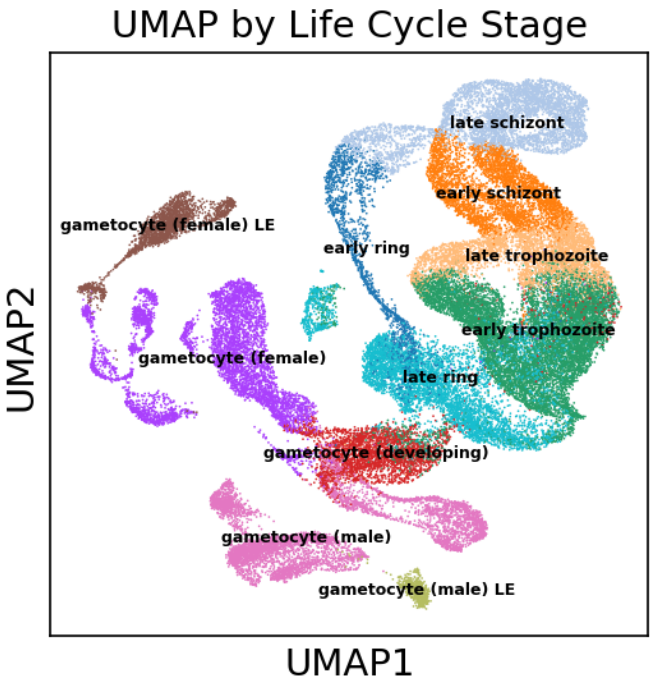

Figure 2. UMAP visualization of single-cell transcriptomes, colored by life cycle stage. The distinct clusters and their arrangement reflect the progression of parasite development, confirming data quality and cell stage annotations.

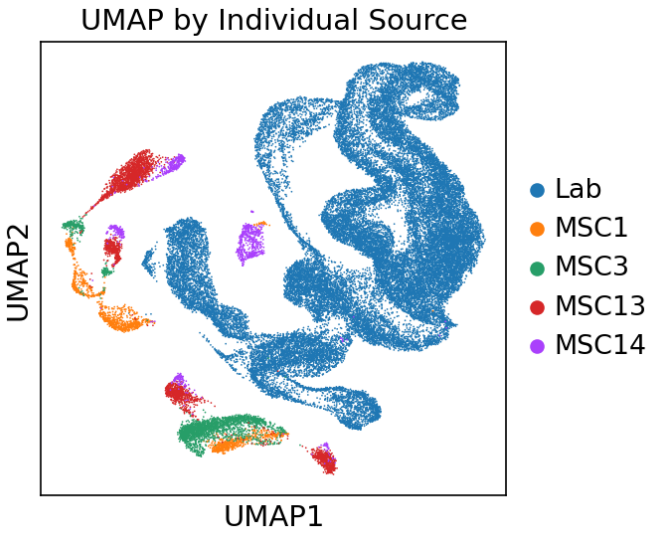

Figure 3. UMAP visualization of single-cell transcriptomes colored by sample source (lab strains and field isolates). The distinct clustering patterns reveal differences in cellular composition and transcriptional profiles between lab-adapted parasites and field isolates.

To systematically investigate the transcriptional differences between lab-adapted and field-isolated parasites, we performed differential gene expression (DGE) analysis within four major life cycle stages present in both datasets: late ring, early trophozoite, female gametocyte, and male gametocyte.

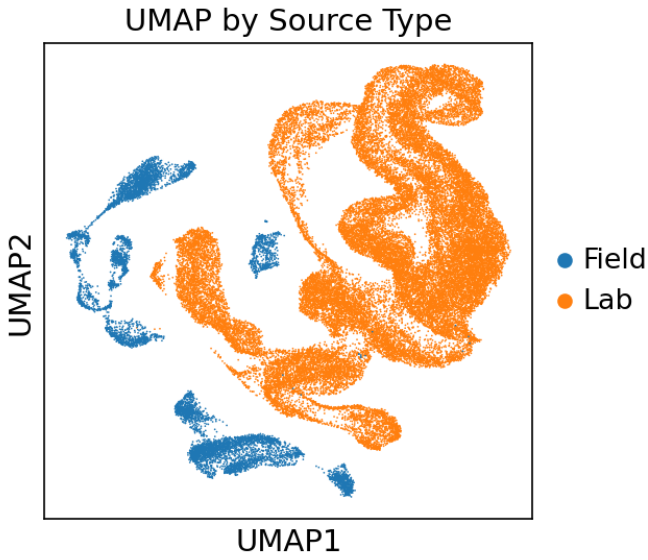

Figure 4. UMAP visualization of single-cell transcriptomes colored by source (lab vs. field), revealing distinct clusters and shared regions, reflecting differences in cellular composition and suggesting potential transcriptional divergence even within shared life cycle stages.

3.3.1. *Late Ring Stage*

In the late ring stage, a comparison between 428 field cells and 5,438 lab cells revealed a vast number of differentially expressed genes. Genes such as `PF3D7-1372200` (log-fold change [LFC] = 4.87) and `PF3D7-0831800` (LFC = 2.33) were significantly upregulated in field isolates (FDR < 0.05). As shown in the volcano plot in Figure 5, there is a strong skew towards upregulation in field isolates, indicating a widespread transcriptional divergence at this early stage of the IDC.

3.3.2. *Early Trophozoite Stage*

The divergence continued into the early trophozoite stage (122 field vs. 9,635 lab cells). Again, a large number of genes were significantly upregulated in the field population, including `PF3D7-1372200` (LFC = 10.44) and `PF3D7-1001500` (LFC = 5.39). The magnitude of these differences suggests that field parasites exhibit a distinct transcriptional program during their primary growth phase, potentially related to adaptation to host factors absent in culture. As shown in Figure 6, the higher log-fold change observed in the field isolates compared to lab strains may be attributed to specific adaptations to the *in vivo* environment, such as nutrient availability or immune responses.

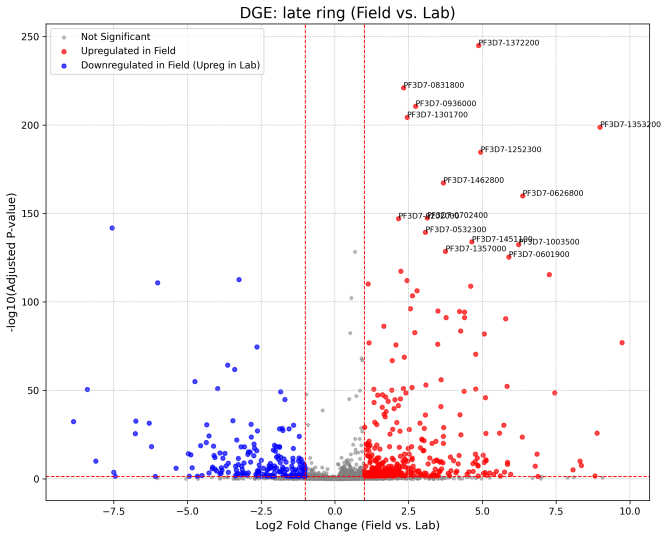

Figure 5. Volcano plot showing differential gene expression between field isolates and lab strains in the late ring stage, revealing a strong skew towards upregulation in field isolates, suggesting a widespread transcriptional divergence at this early stage of the IDC.

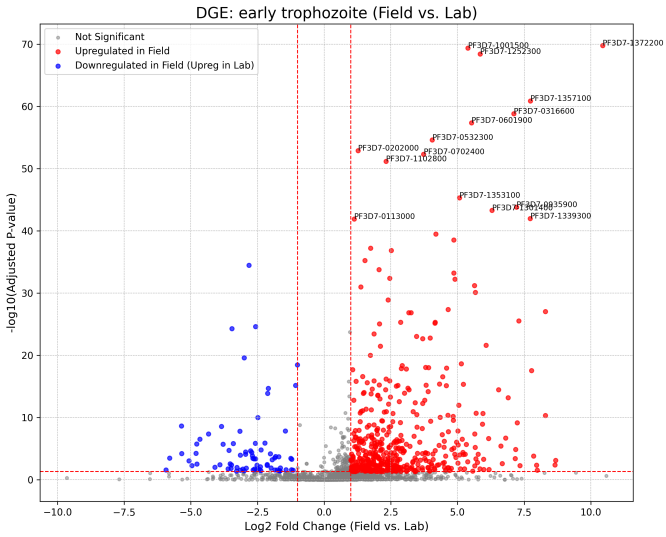

Figure 6. Volcano plot showing differentially expressed genes in early trophozoites comparing field isolates to lab strains. The plot shows a strong skew towards upregulation in field isolates, indicating a widespread transcriptional divergence at this early stage of the intraerythrocytic development cycle.

3.3.3. *Sexual Stages (Gametocytes)*

The most dramatic transcriptional differences were observed in the sexual stages. In female gametocytes (1,656 field vs. 3,903 lab cells), thousands of genes were differentially expressed, with many p-values approaching zero. Top upregulated genes in field isolates included the known male gametocyte marker `PF3D7-1031000` (Pfs25, LFC=1.22) and `PF3D7-1423600` (LFC=5.72). The unexpected upregulation of a male marker in the female cluster from field isolates may suggest either a less distinct separation of sexual lineages *in vivo* or potential misannotation that reflects a different cellular state in field gametocytes. It is possible that the regulatory mechanisms controlling sexual differentiation are less stringent in field isolates, leading to the expression of both male and female markers in the same cells. Figure **??** shows the differentially expressed genes in female gametocytes, highlighting the strong skew towards upregulation in field isolates.

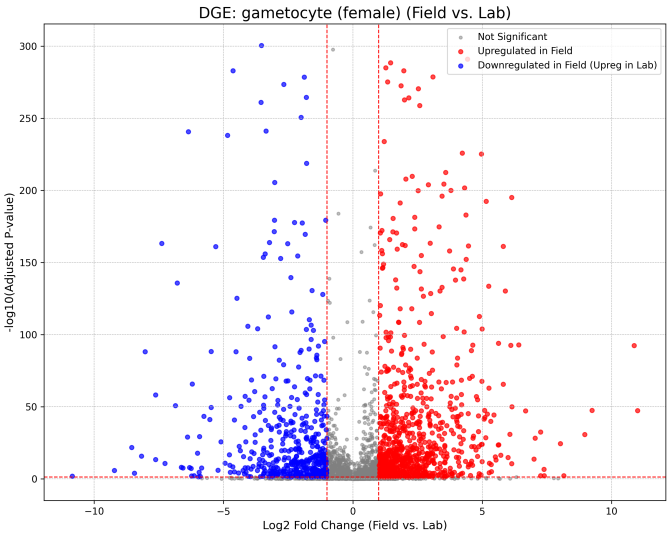

Figure 7. Volcano plot showing differentially expressed genes in female gametocytes between field isolates and lab strains. Red points indicate genes upregulated in field isolates, while blue points indicate genes downregulated in field isolates. The strong skew towards upregulation in field isolates suggests a wide transcriptional divergence in female gametocytes between field and lab conditions.

Similarly, male gametocytes (3,364 field vs. 1,964 lab cells) showed profound differences, with genes like `PF3D7-0205000` (LFC=2.46) and `PF3D7-1201600` (LFC=2.67) being highly upregulated in the field population. These results suggest a fundamental divergence in the sexual development programs between lab and field conditions. The distinct transcriptional profiles observed in gametocytes from field isolates may reflect adaptations to the mosquito vector, which are not necessary for gametocyte development in laboratory cultures.

3.4. *Developmental Trajectories Reveal Altered Progression Dynamics*

To compare the dynamic transcriptional programs controlling stage transitions, we reconstructed developmental trajectories using PAGA and calculated diffusion pseudotime.

3.4.1. *Asexual Development*

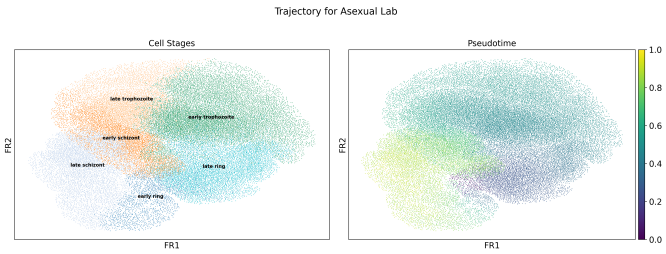

Figure 8. UMAP visualization of the *$P.$ falciparum* asexual development cycle in lab strains, colored by cell stage annotation (left) and diffusion pseudotime (right), showing a clear trajectory from early rings to late schizonts.

For the lab strains, as depicted in Figure **??**, we reconstructed a complete, cyclical trajectory of the IDC, rooted at the 'early ring' stage. The pseudotime progression faithfully recapitulated the known sequence of development from rings to trophozoites and finally to schizonts. The PAGA graph in Figure **??** and Figure **??** confirmed strong connectivity between consecutive stages.

For the field isolates, the asexual trajectory was necessarily incomplete, containing only late rings and early trophozoites. While this segment showed the expected progression, the absence of other stages prevented a full comparison of the cycle's dynamics and efficiency. The fact that we could not reconstruct the full asexual cycle trajectory in the field isolates underscores the challenges of studying these parasites, which are often present at low densities and are difficult to synchronize. This incomplete trajectory is visualized in the PAGA graphs in Figure **??** and Figure **??**.

3.4.2. *Sexual Development*

The sexual development trajectory in lab strains, as shown in Figure **??**, rooted at 'gametocyte (developing)', showed a clear bifurcation into mature male and female lineages. The PAGA graph in Figure **??** confirms this bifurcation.

In contrast, the trajectory for field isolates revealed a more complex picture. Figure **??** shows the UMAP visualization of this trajectory. This trajectory, rooted at 'gametocyte (female)', not only separated male and female populations but also extended into the 'LE' (late-stage) gametocyte populations, which were absent in the lab data. The PAGA analysis (Figure **??** and Figure **??**)

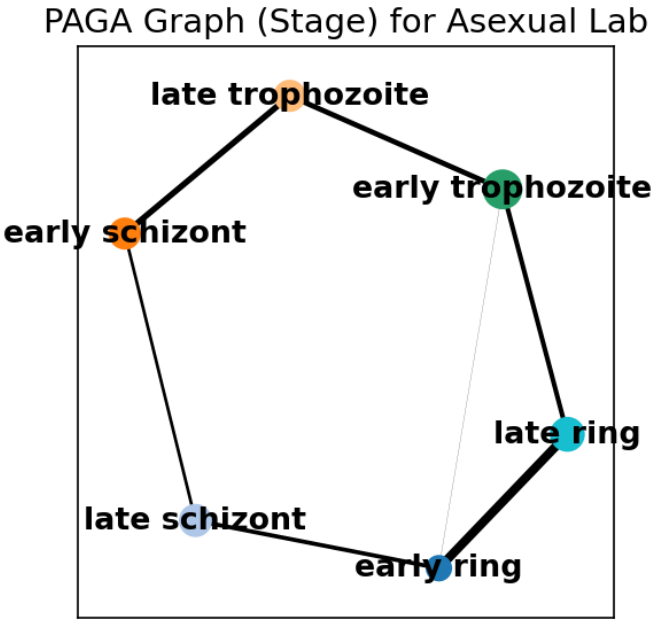

Figure 9. PAGA graph of the asexual development cycle in lab strains, showing connectivity between consecutive stages, and confirming the trajectory from rings to trophozoites and finally to schizonts.

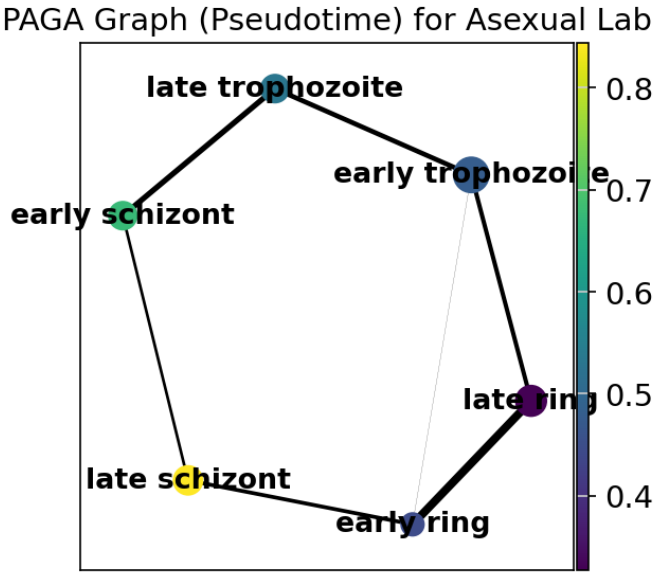

Figure 10. PAGA graph for the asexual development cycle of lab strains, demonstrating strong connectivity between consecutive stages and confirming pseudotime progression from rings to trophozoites and schizonts.

showed strong connections from the main female and male clusters to their respective 'LE' counterparts, confirming that these represent a further step in maturation occurring *in vivo*. This provides clear evidence that the

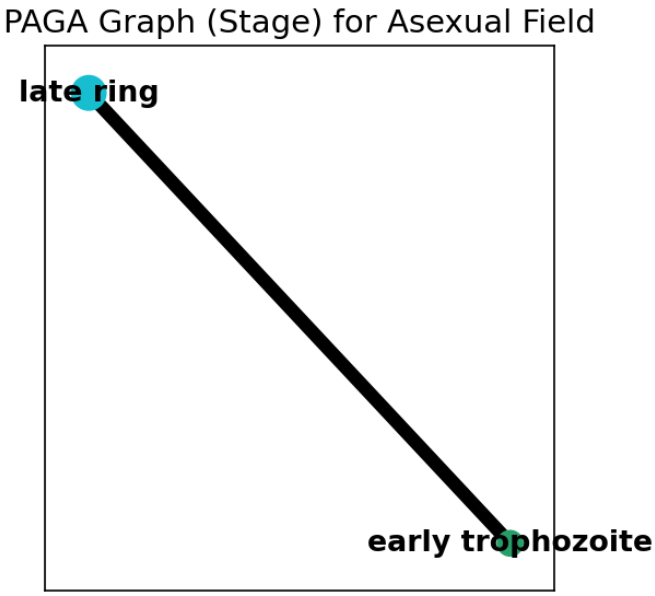

Figure 11. PAGA graph of the asexual trajectory for field isolates showing connectivity between late ring and early trophozoite stages, revealing the incomplete asexual cycle captured in field samples.

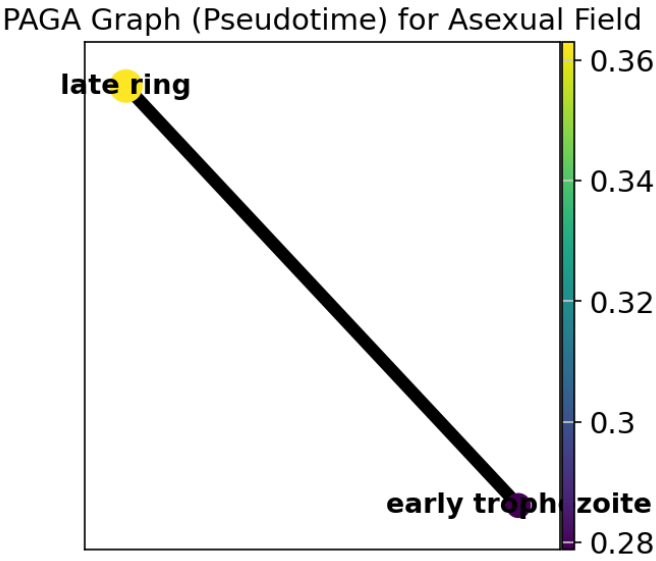

Figure 12. PAGA graph of the asexual development trajectory for field isolates, showing the connection between late ring and early trophozoite stages, colored by pseudotime. This incomplete trajectory reflects the absence of early ring and schizont stages in the field samples, preventing a full comparison of the asexual cycle dynamics.

developmental pathway of gametocytes is extended or altered in the host environment compared to standard laboratory culture. The presence of late-stage gametocytes in the field isolates but not in the lab strains

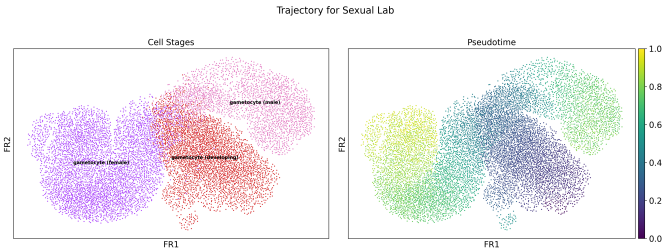

Figure 13. FA plot of the sexual development trajectory for lab strains, colored by cell stage (left) and diffusion pseudotime (right). This trajectory, rooted at 'gametocyte (developing)', shows a clear bifurcation into mature male and female lineages.

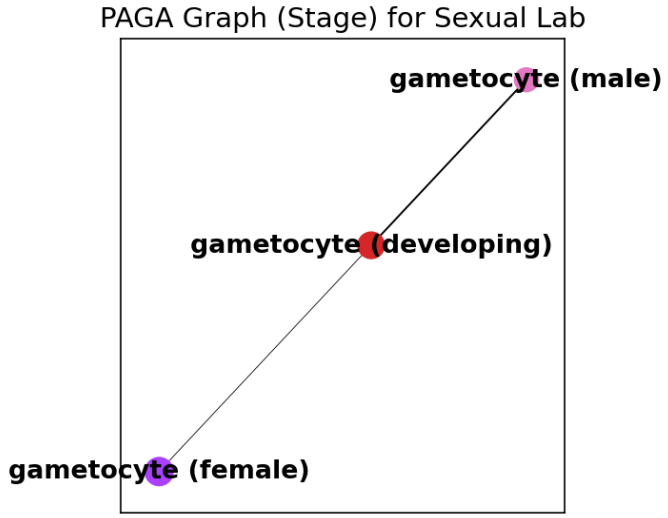

Figure 14. PAGA graph for sexual development in lab strains, showing connectivity between developing gametocytes and differentiated male and female gametocytes, recapitulating the expected progression of sexual development.

suggests that the *in vivo* environment provides specific signals or conditions that are necessary for the completion of gametocyte maturation.

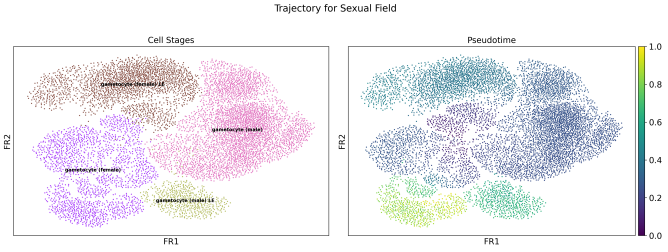

Figure 15. UMAP visualization of the sexual development trajectory for field isolates, colored by cell stage annotation (left) and diffusion pseudotime (right). The trajectory extends to late-stage gametocytes, which are absent in lab data, suggesting an altered developmental pathway **in vivo**.

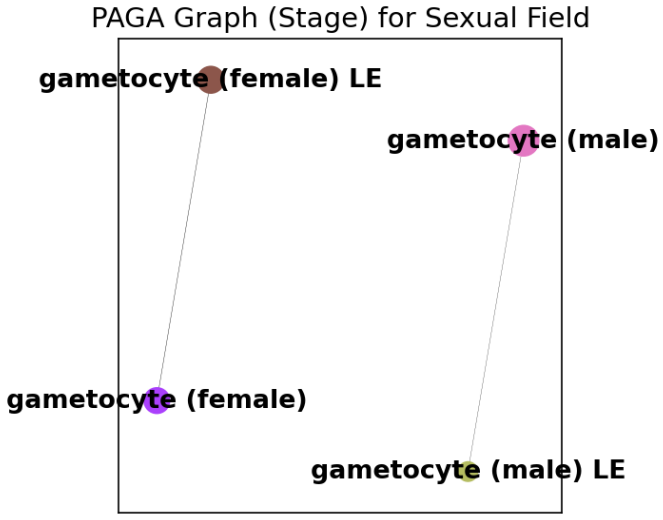

Figure 16. PAGA graph of sexual development in field isolates reveals connections between developing gametocytes and late-stage gametocytes (LE), suggesting a developmental pathway extension not observed in lab strains.

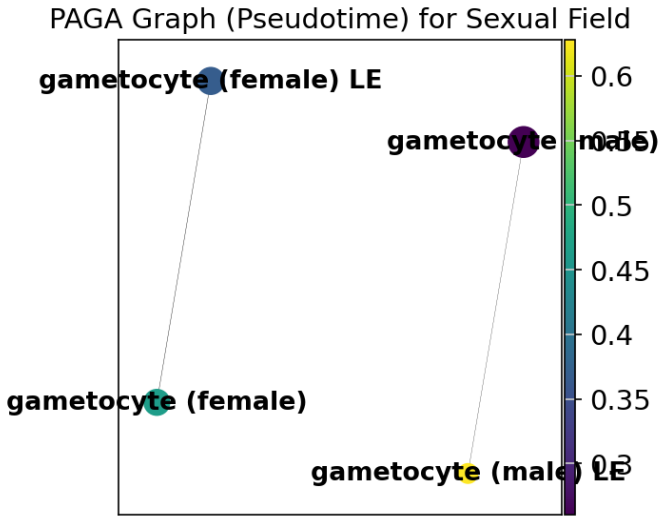

Figure 17. PAGA graph of field isolates showing the sexual development trajectory, which extends into late-stage gametocytes ('LE') not observed in lab strains, suggesting an altered or extended developmental pathway in the host environment.

3.5. *Identification of Candidate Master Regulators and Inferred Regulatory Networks*

To identify potential master regulators driving stage transitions, we developed a pipeline to find low-expression genes with transient expression peaks that precede the activation of large co-expressed gene modules. We applied this analysis to each of the four reconstructed trajectories.

3.5.1. *Asexual Trajectories*

In the **Asexual Lab** trajectory, we identified 17 gene modules and a remarkable 3,428 putative regulatory links. A large number of candidate regulators, such as `PF3D7-0210800` and `PF3D7-0605900`, were predicted to peak at the very beginning of the trajectory (pseudotime bin 0), preceding the activation of modules in bin 1. This suggests a major, coordinated wave of regulatory activity initiates the IDC in the lab setting. Figure **??** shows an example of this regulatory interaction.

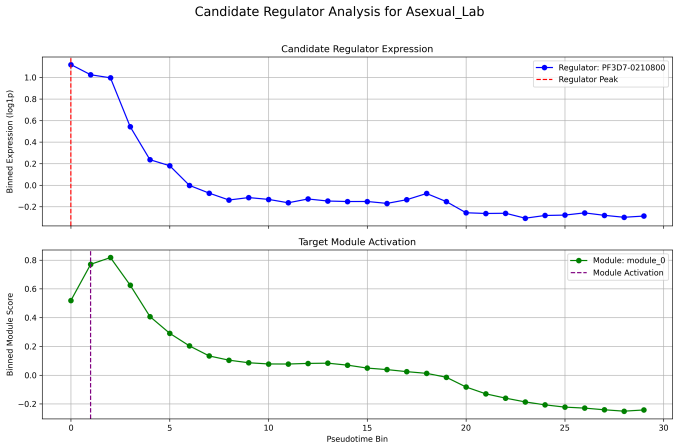

Figure 18. Inferred regulatory interaction in the asexual lab trajectory, showing the expression of candidate regulator 'PF3D7-0210800' peaking before the activation of target module 0, suggesting a regulatory cascade at the onset of asexual development.

In the **Asexual Field** trajectory, we found 13 modules but only 945 putative links. The dynamics appeared different, with top candidate regulators like `PF3D7-0532500` peaking later in the observed trajectory segment (e.g., bin 5), suggesting that the regulatory triggers in field parasites may be timed differently or that the key initiating events occur in the unobserved early ring stage. The reduced number of regulatory links in the field isolates compared to the lab strains may reflect the limited number of stages captured in the field data or differences in the complexity of the regulatory networks operating in the *in vivo* environment. Figure **??** shows the expression pattern of a candidate regulator and its target gene module.

3.5.2. *Sexual Trajectories*

The **Sexual Lab** trajectory yielded 13 gene modules and 404 putative regulatory links. Similar to the asexual lab trajectory, many candidate regulators (e.g.,

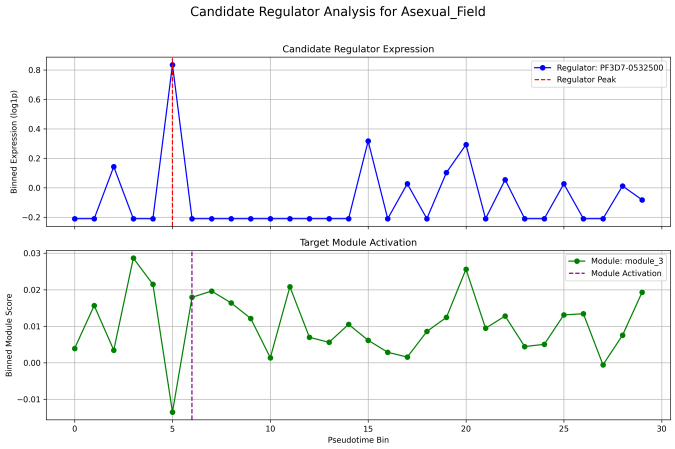

Figure 19. Expression pattern of the candidate regulator PF3D7-0532500 and its target gene module in the asexual trajectory of field isolates, showing a peak of expression in the regulator preceding the activation of its target module, suggesting a regulatory relationship that controls stage transitions during parasite development.

PF3D7-0201500, PF3D7-0203700) were found to peak at the earliest pseudotime bins, suggesting a regulatory cascade at the onset of sexual commitment. An example of such a regulatory link is shown in Figure **??**.

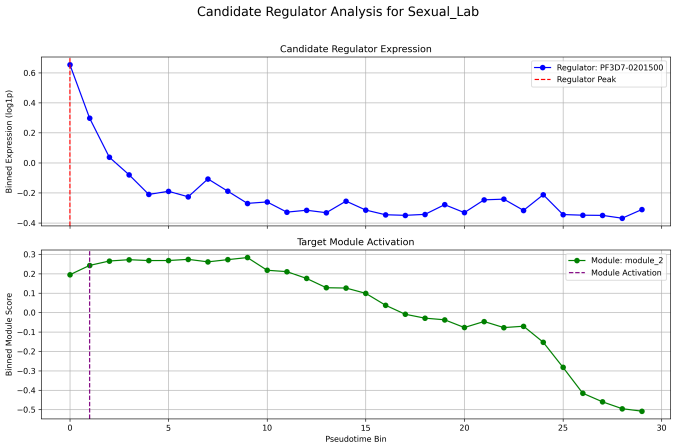

Figure 20. Inferred regulatory link in lab-derived gametocytes. The expression of the candidate regulator 'PF3D7-0201500' peaks before the activation of target module 2, suggesting a regulatory relationship at the onset of sexual commitment.

The **Sexual Field** trajectory presented a much more active regulatory landscape, with 18 modules and 1,917 putative links. This four-fold increase in predicted regulatory interactions compared to the lab sexual trajectory suggests a more complex or fine-tuned regulatory program governs gametocytogenesis *in vivo*. Candidate regulators like PF3D7-0103500 were predicted to act later in the trajectory (e.g., peaking at bin 17 before module activation at bin 18), potentially controlling the transition to the late-stage 'LE' gametocytes unique to the field environment. The increased complexity of the regulatory networks in the field isolates suggests that the *in vivo* environment imposes additional regulatory constraints on gametocyte development, potentially related to interactions with the host immune system or the mosquito vector. Figure **??** shows an example of this inferred regulatory interaction.

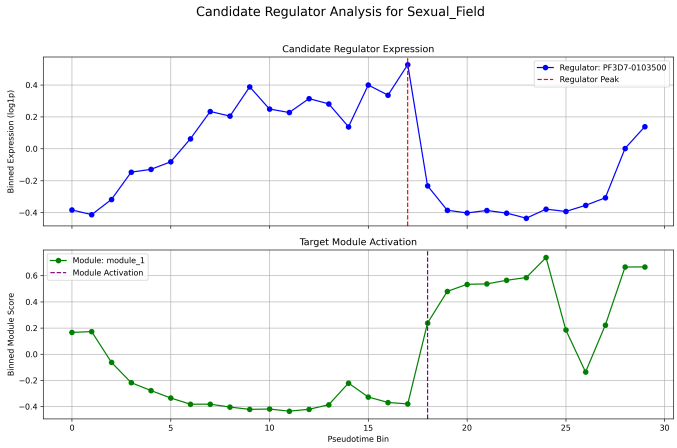

Figure 21. Inferred regulatory interaction in field isolates during sexual development. The expression of the candidate regulator 'PF3D7-0103500' peaks before the activation of a target gene module, suggesting a regulatory relationship in the **in vivo** environment.

3.6. *Summary of Results*

In summary, these results point to significant rewiring of the regulatory networks controlling parasite development. Field isolates not only exhibit profoundly different transcriptional profiles at static stages but also appear to utilize altered developmental trajectories governed by a distinct set of timed regulatory events, particularly during sexual development. The candidate regulators identified here provide a rich set of hypotheses for future experimental validation to uncover the mechanisms of parasite adaptation to the human host. The differences observed between lab-adapted strains and field isolates highlight the importance of studying parasites in their natural environment to fully understand the complexities of malaria biology. These findings have implications for the development of new interventions that target parasite development and transmission, as well as for the interpretation of experimental results obtained from laboratory studies.

4. CONCLUSIONS

4.1. *Conclusions*

This study addressed the critical problem of how laboratory adaptation alters *Plasmodium falciparum* biology, potentially compromising the relevance of research findings for natural infections. We hypothesized that adaptation to lab culture leads to alterations in the timing, path, or efficiency of developmental processes. To address this, we performed a comprehensive comparative analysis of dynamic transcriptional programs and regulatory networks governing stage transitions in lab-adapted strains and field isolates.

Using a single-cell RNA sequencing dataset of 45,691 cells, we reconstructed developmental trajectories, performed differential gene expression analysis, and identified co-expression modules and candidate regulators in both lab and field parasites. Our analysis revealed substantial differences in transcriptional profiles, developmental trajectories, and regulatory networks between lab and field parasites, particularly during sexual development.

Specifically, we observed distinct expression patterns in late ring and early trophozoite stages, with a strong skew towards upregulation in field isolates, indicating a widespread transcriptional divergence at these early stages of the IDC. The most dramatic differences were observed in the sexual stages, where field isolates exhibited distinct transcriptional profiles, including the unexpected co-expression of male and female markers and the presence of late-stage gametocytes absent in lab strains.

Developmental trajectory analysis revealed that while lab strains exhibited a complete, cyclical trajectory of the IDC and a clear bifurcation into mature male and female gametocyte lineages, field isolates showed an incomplete asexual trajectory and a more complex sexual trajectory extending into late-stage gametocytes unique to the *in vivo* environment.

Finally, we identified a unique set of candidate master regulators and inferred regulatory interactions in field isolates, suggesting adaptation to *in vivo* conditions alters developmental control and fate determination. We identified a larger number of regulatory links during the sexual trajectory in field isolates, suggesting more complex regulatory mechanisms.

From these results, we have learned that laboratory adaptation significantly rewires the regulatory networks controlling parasite development. Field isolates not only exhibit profoundly different transcriptional profiles at static stages but also appear to utilize altered developmental trajectories governed by a distinct set of timed regulatory events, particularly during sexual development. These findings emphasize the importance of studying field isolates to fully understand *P. falciparum* biology and the molecular mechanisms underlying parasite adaptation to the human host. The candidate regulators identified here provide a rich set of hypotheses for future experimental validation to uncover the mechanisms of parasite adaptation to the human host.