UNIVERSITY OF LIVERPOOL

Life703 Research Project

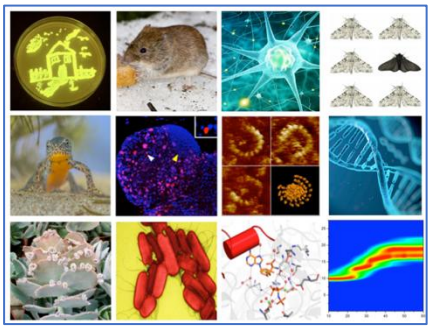
Final Report

Report Title: **The genetics of adaptation: Epigenetic development of a generalist cichlid fish species**

Name: Midhukrishna
Student Number: 201855638

Supervisor Name: Dr. Tarang Mehta

MSc Programme: Bioinformatics



School of Biosciences

GAI Declaration

I **did** use GAI in the preparation of this report.

Tool Details	Description of what used for, and how
Chatgpt 3.0 https://chatgpt.com	I used Chatgpt to structure my document and check for grammatical errors

Lay Summary

Cichlid fish in East Africa's Lake Malawi are known for their incredible variety, with hundreds of species evolving rapidly to fill different ecological roles. This study explored how the early development of one cichlid species, *Astatotilapia calliptera*, is controlled at a molecular level, focusing on changes in how genes are turned on or off. We used a technique called ATAC-seq to examine which parts of the fish's DNA are active at three stages of embryo growth: 3, 7, and 12 days after fertilization. By mapping these active DNA regions, we aimed to understand how they influence the development of traits that help cichlids adapt to their environment. We analysed the data to identify patterns in DNA activity, checked which genes are affected, and studied which proteins, called transcription factors, regulate function. With comparative studies against other cichlid fish development, we can start to better understand how changes in DNA activity contribute to cichlid fish diversity. Understanding these processes not only sheds light on cichlid evolution but also offers insights into how living things develop and adapt, which could have broader applications in biology and conservation.

Abstract

This study investigated the epigenetic regulation of early embryonic development in *Astatotilapia calliptera*, a generalist cichlid species central to the Lake Malawi adaptive radiation. We aimed to characterize chromatin accessibility dynamics at 3, 7, and 12-days post-fertilization (dpf) using ATAC-seq to identify regulatory elements driving developmental plasticity. ATAC-seq was performed on two biological replicates per stage, generating high-quality sequencing data. Reads were processed, aligned to the *A. calliptera* genome, and peaks of open chromatin were identified. Peaks were annotated to genomic features, and enrichment analyses were conducted to assess their distribution and functional significance. We identified distinct chromatin accessibility profiles and enrichment in genomic annotations across stages: activity of conserved noncoding elements (CNEs) and 5' UTRs at 3dpf, elevated promoters at 7dpf, and sustained UTRs/CNEs at 12dpf. With enrichment of embryonic morphogenesis at 3dpf, tissue and organ development at 7dpf, and locomotory behavior at 12dpf, we find that chromatin accessibility shifts from early developmental programs to later functional and behavioral processes, showing how epigenetic regulation guides development in *A. calliptera*. Differential chromatin accessibility was evaluated to detect stage-specific regulatory changes, identifying many differentially active peaks in early comparisons (e.g., 78,675 in 3dpf vs. 7dpf) versus fewer (eg., 10341 in 7dpf vs 12dpf) in later ones. Many of the differentially accessible regions correspond to developmental genes, with down-regulated GO terms such as head development and eye morphogenesis linked to genes including *dnmt1*, *otx2a*, *otx2b*, and *pbx2*, highlighting their roles in early embryonic development. These results suggest that dynamic chromatin accessibility underpins gene regulation during early development, potentially contributing to the phenotypic diversity of cichlids. This work provides a foundation for understanding the molecular basis of adaptive radiation in cichlids, highlighting the role of epigenetic mechanisms in evolutionary diversification.

1. Introduction

East African cichlid fishes are an important model for studying adaptive radiation due to their rapid speciation and extensive phenotypic diversification across Lakes Malawi, Victoria, and Tanganyika (Kocher, 2004). Lake Malawi alone hosts approximately 500–860 species that diverged within the last 800,000 years, displaying a remarkable array of morphologies, behaviours, and ecological adaptations (Kocher, 2004; Brawand *et al.*, 2014; Malinsky *et al.*, 2018). This rapid evolutionary divergence makes cichlids an ideal system for investigating the genetic and molecular mechanisms driving adaptation and speciation (Svardal, Salzburger and Malinsky, 2021). Factors such as genetic variation, hybridization, and environmental pressures have been implicated in this diversity, yet the contribution of epigenetic mechanisms remains underexplored (Svardal, Salzburger and Malinsky, 2021).

Astatotilapia calliptera, a riverine generalist cichlid inhabiting diverse environments in and around Lake Malawi, holds a key phylogenetic position within the Lake Malawi radiation, sharing significant genetic ancestry with the rock-dwelling mbuna clade (Malinsky *et al.*, 2018). Its ecological versatility, thriving in both riverine and lacustrine habitats, suggests a high degree of developmental and adaptive plasticity, making it a critical species for studying the genetic basis of cichlid diversification (Parsons *et al.*, 2017). Hypotheses suggest that *A. calliptera* may resemble the ancestral progenitor of the Lake Malawi radiation or act as a sympatric contributor to its genetic diversity (Svardal, Salzburger and Malinsky, 2021). Elucidating the molecular mechanisms governing its early development could provide insights into the processes facilitating rapid evolutionary divergence in cichlids.

Epigenetic modifications, particularly changes in chromatin accessibility, play a pivotal role in regulating gene expression during development, influencing phenotypic outcomes without altering the underlying DNA sequence. This has been demonstrated in vertebrates, where ATAC-seq analyses have revealed dynamic chromatin landscapes across various developmental stages (Louise Smith, Mok and Münsterberg, 2022). Open chromatin regions, accessible to transcription factors (TFs) and other regulatory proteins orchestrate stage-specific gene expression, shaping developmental pathways (Buenrostro *et al.*, 2013). In cichlids,

epigenetic mechanisms are hypothesized to enhance the plasticity that enables adaptation to diverse ecological niches, yet their role in early embryonic development remains poorly characterized (Mehta *et al.*, 2023; Tetrault *et al.*, 2023). The Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) offers a powerful approach to map open chromatin regions across the genome, providing insights into the regulatory landscape (Li *et al.*, 2019; Bentsen *et al.*, 2020). Recent studies in related cichlid species, such as Nile tilapia, have demonstrated the efficacy of ATAC-seq in identifying regulatory elements associated with adaptive traits, underscoring its potential for studying developmental epigenetics (Mehta *et al.*, 2023).

Despite the evolutionary significance of *A. calliptera*, the epigenetic dynamics of its early embryonic development remain largely uncharacterized. This study aimed to profile chromatin accessibility at three critical developmental stages - 3, 7, and 12-days post-fertilization (dpf) - to elucidate the regulatory mechanisms underpinning developmental plasticity in this species. We hypothesized that differential chromatin accessibility and transcription factor binding patterns drive stage-specific gene regulation, contributing to the molecular foundation of phenotypic diversity in cichlids. As ATAC-seq allows you to study peak annotation, differential accessibility analysis, and TF footprinting, we sought to map the epigenetic landscape and identify regulatory elements associated with developmental genes, providing a framework for understanding the molecular drivers of cichlid adaptive radiation.

2. Methods

2.1 Sample collection and ATAC-seq data generation

Embryos of *Astatotilapia calliptera* were collected at three developmental stages: 3 days post-fertilization (3dpf), 7dpf, and 12dpf) with two biological replicates per stage. ATAC-seq libraries were prepared following established protocols (Buenrostro *et al.*, 2013), generating an average of 46 million 50 bp paired-end reads per sample and 8 million reads per naked DNA control using the NovaSeq 6000 platform (Illumina).

2.2 Read processing and peak calling

Raw ATAC-seq reads were trimmed using Trim Galore (v0.5.0) (Krueger, 2015) and quality checked with FastQC (v0.11.9) (Simon Andrews *et al.*, 2010) using default parameters. Trimmed reads were aligned to the *A. calliptera* fAstCal1.2 genome using Bowtie2 (v2.2.6) (Langmead and Salzberg, 2012), and output in BAM format using SAMtools (v1.9) (Li *et al.*, 2009). Mitochondrial reads, identified via BLAST (v2.3.0) (Camacho *et al.*, 2009), were removed using SAMtools. BAM files were sorted, and duplicated reads marked using Sambamba (v0.6.5) (Tarasov *et al.*, 2015). Non-primary, unmapped, and low-quality reads were filtered with SAM tools, retaining properly paired reads. BAM files were converted to tag-align files using Bedtools (v2.30.0) (Quinlan and Hall, 2010), and Tn5 shifting was performed prior to peak calling. Peaks were identified using MACS2 (v2.1.1) (Zhang *et al.*, 2008).

2.3 Peak annotation and visualization

Narrow peaks were annotated by mapping to genomic features using Bedtools intersect. Peak counts per annotation were collated for each stage (3dpf, 7dpf, 12dpf) using a custom bash script (collatepeakfeature.sh). The distribution of peaks was plotted using ggplot2 in R (v4.4.2) (Wickham, 2016).

2.4 Peak enrichment analysis

Enrichment of peaks in annotated genomic regions was assessed using the Genomic Association Tester (GAT, v1.0) (Heger *et al.*, 2013) with a false discovery rate (FDR) correction using the Benjamini-Hochberg method (FDR < 0.05). Input files included collated peak files, annotation files and chromosome sizes. Enrichment results were plotted using ggplot2 in R (v4.4.2).

2.5 Functional enrichment analysis

Gene Ontology (GO) enrichment analysis was performed on genes with narrow peaks overlapping 5 kb promoter regions using the g:GOst module of g:Profiler (Raudvere *et al.*, 2019) with a significance threshold of adjusted p-value < 0.05 (Benjamini-Hochberg correction). Ensembl gene IDs (ENSACLG) were extracted and used as input for each stage. Enriched GO terms were plotted using ggplot2 in R (v4.4.2).

2.6 Differential chromatin accessibility analysis

Differential chromatin accessibility across developmental stages (3dpf, 7dpf, and 12dpf) in *Astatotilapia calliptera* was analyzed to identify stage-specific open chromatin regions. BAM files from two biological replicates per stage, processed as described in section 2.2, were used as input. Read counts were quantified in 150 bp genomic windows using csaw (v1.20.0) (Lun and Smyth, 2015). Pairwise comparisons (3dpf vs. 7dpf, 3dpf vs. 12dpf, 7dpf vs. 12dpf) were performed using edgeR (v4.0.0) (Chen *et al.*, 2025) with a false discovery rate (FDR) correction (Benjamini-Hochberg, FDR < 0.05). Significant differentially accessible regions were converted to GenomicRanges format using rtracklayer (v1.46.0) (Lawrence, Gentleman and Carey, 2009) and overlapped with a 5 kb promoter BED file from the *A. calliptera* fAstCal1.2 genome annotation to assign gene IDs. Results were saved as TSV files, and unique gene lists for each comparison were exported. These gene lists were analyzed for Gene Ontology (GO) enrichment using the g:GOst module of g:Profiler with a significance threshold of adjusted p-value < 0.05 (Benjamini-Hochberg correction).

3. Results and Discussion

3.1 Chromatin Accessibility Distribution Across Genomic Features

ATAC-seq was performed to assess chromatin accessibility across three developmental stages (3dpf, 7dpf, and 12dpf) in *Astatotilapia calliptera*. Peaks were annotated to genomic features to understand the regulatory landscape during embryogenesis.

3.1.1 Genomic Feature Distribution of Accessible Peaks

Annotated peaks were categorized based on their location in exonic, intronic, promoter, intergenic, conserved noncoding elements (CNEs), and UTR regions. A pooled analysis across all stages showed that intronic regions accounted for 40% of peaks, followed by exons (20%), intergenic (18%), promoters (10%),and UTRs (5%) (Figure 1).

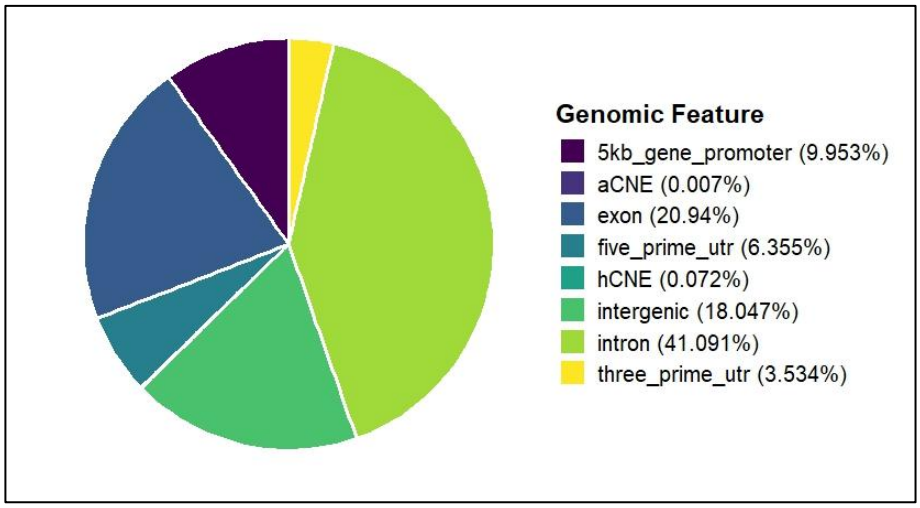


Figure 1. Distribution of ATAC-seq peaks across genomic features in *A. calliptera* across all stages.

This distribution, with intronic regions dominating, suggests potential enhancer activity, while the prominence of promoter peaks indicates active transcription regulation during early development.

3.1.2 Proximity of Accessible Peaks to Transcription Start Sites (TSS)

To identify peaks potentially associated with transcriptional activation, peaks were mapped relative to transcription start sites (TSS), with "near" defined as within ±100 bp of TSSs and "distal" as beyond ±100 bp. A bar plot depicting the distribution of peaks relative to TSSs revealed that peaks were most concentrated within ±100 bp of TSSs at all developmental

stages, with a progressive increase in peak density from 3dpf to 12dpf (Figure 2). Specifically, at 3dpf, an average of 1,857 peaks were near TSSs and by 12dpf, peaks near TSSs further increased to an average of 2,046. These trends indicate stronger promoter activity as development progresses, supporting intensified gene regulation.

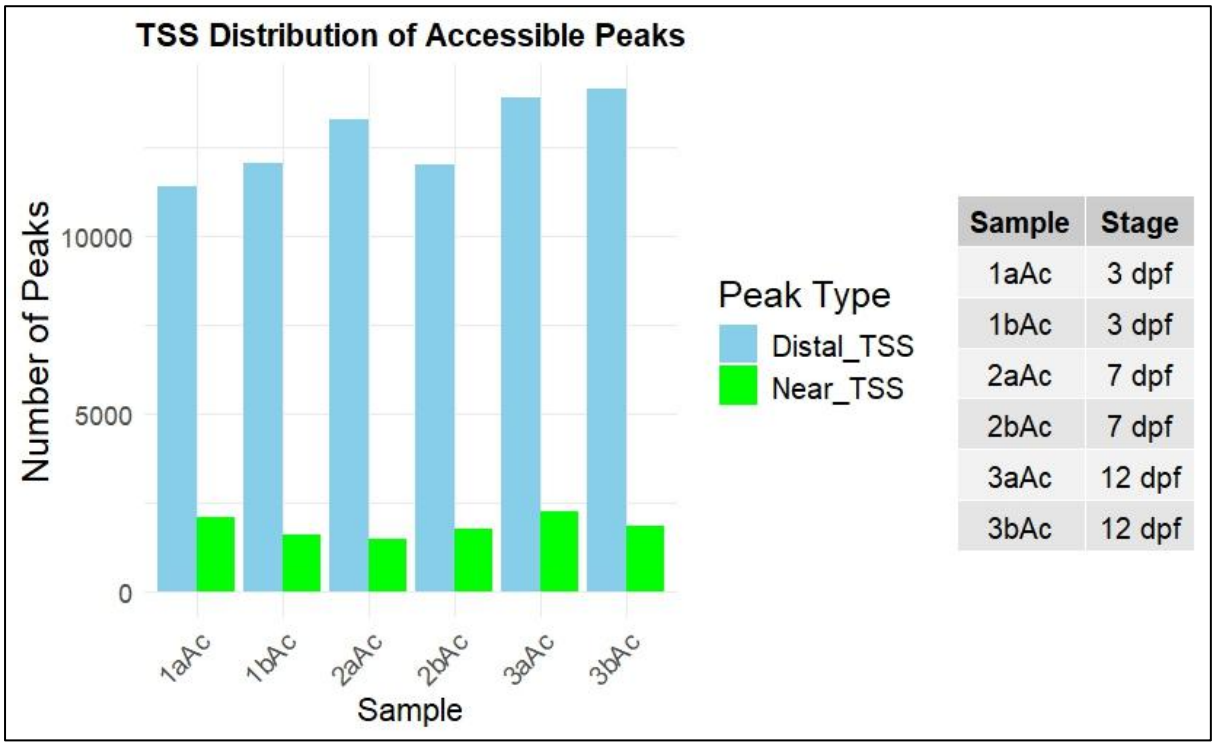


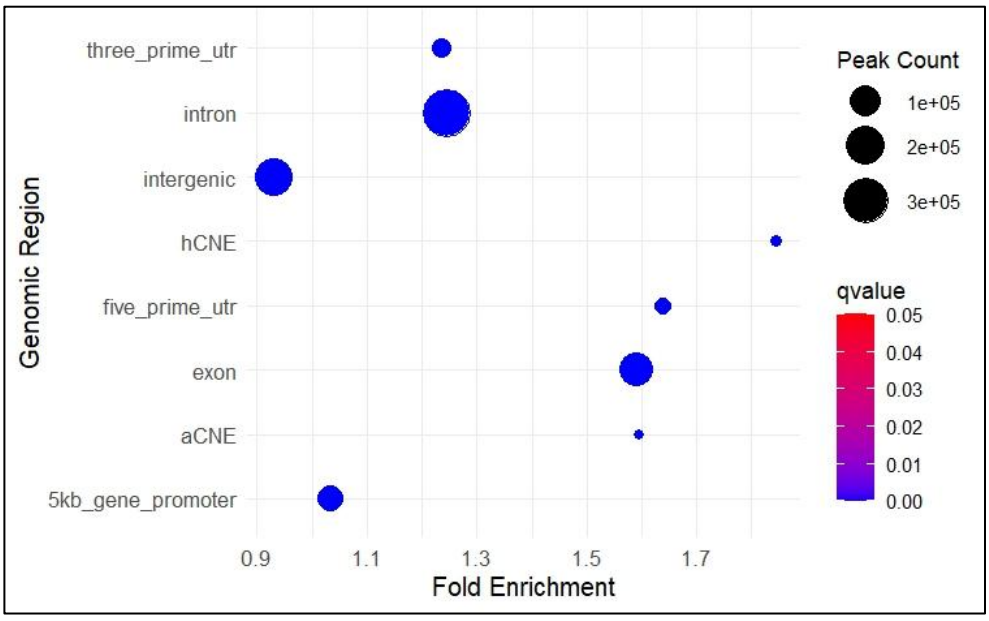
Figure 2: TSS distribution of accessible peaks.

3.2 Stage-Specific Enrichment of Accessible Regions

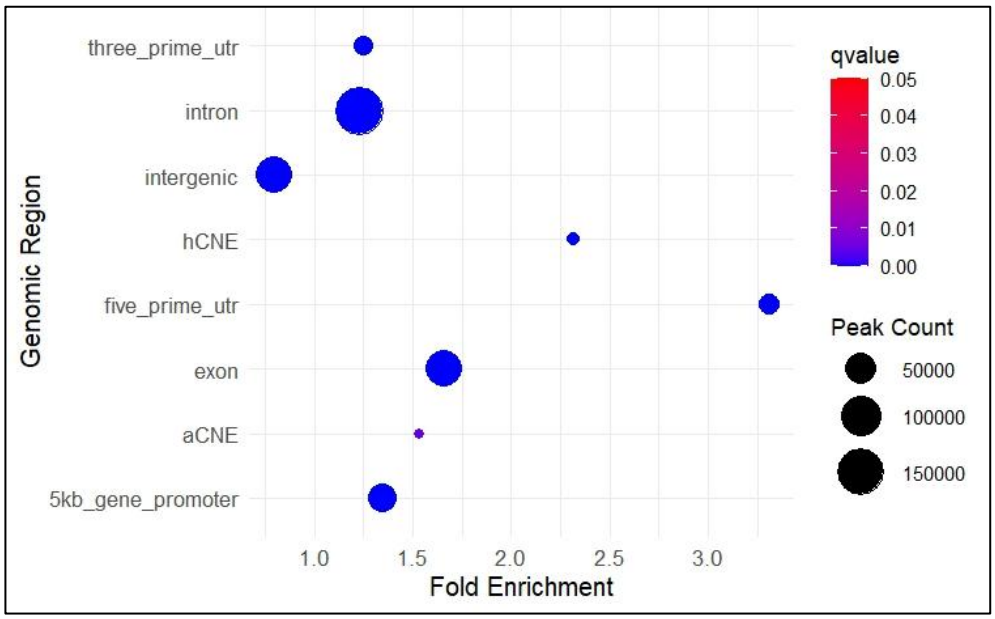
To evaluate regulatory changes during development, Genomic Association Tester (GAT) analysis was used to assess the enrichment of ATAC-seq peaks within functional genomic categories (Figure 3a-c). Fold enrichment of peaks across genomic categories shows a dynamic regulatory landscape, with promoter and UTR-associated activities becoming prominent during mid to late developmental stages. At 3dpf, peaks showed significant activity in highly conserved non-coding elements (hCNEs; fold enrichment = 1.84, q-value = 3.38e-06) and 5' UTRs (fold enrichment = 1.63, q-value = 0.0), with moderate enrichment in introns (fold enrichment = 1.24, q-value = 0.0) and minimal enrichment in promoter regions (fold enrichment = 1.03, q-value = 0.0), indicating a focus on conserved regulatory elements during early embryogenesis. At 7dpf, peak enrichment increased markedly in 5' UTRs (fold enrichment = 3.31, q-value = 0.0), hCNEs (fold enrichment = 2.31, q-value = 4.27e-03), and

exons (fold enrichment = 1.66, q-value = 0.0), with a substantial rise in promoter regions (fold enrichment = 1.34, q-value = 0.0) compared to 3dpf, reflecting heightened transcriptional activity during tissue and organ formation. At 12dpf, enrichment remained high in 5' UTRs (fold enrichment = 2.37, q-value = 0.0) and hCNEs (fold enrichment = 1.95, q-value = 5.56e-06), with promoter regions still prominent (fold enrichment = 1.18, q-value = 0.0) but slightly reduced compared to 7dpf, indicating a transition to more specialized regulatory activity.

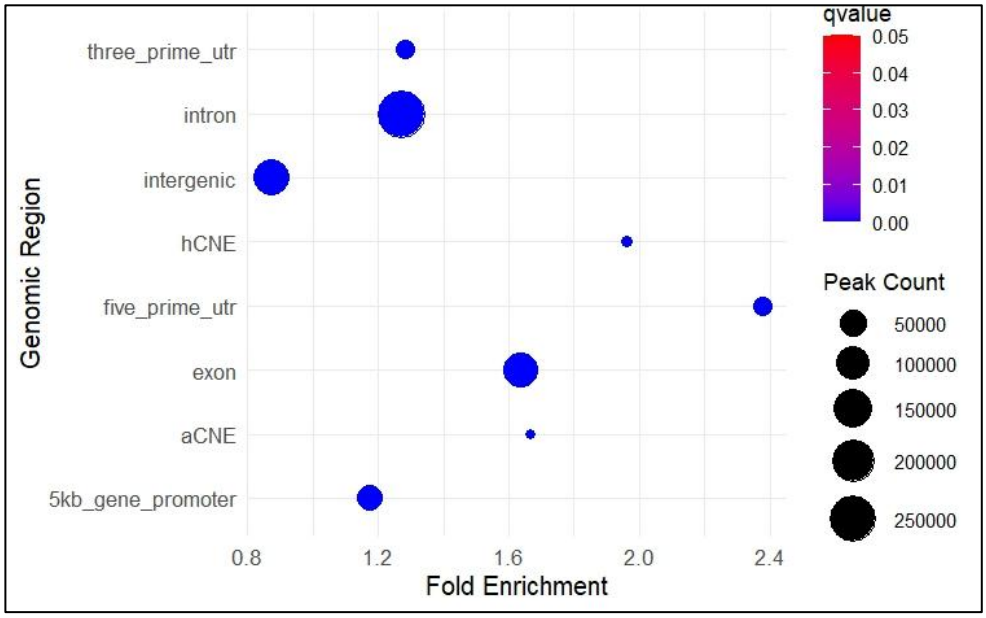
These stage-specific patterns indicate a transition from conserved regulatory elements in early development to promoter-driven regulation in later stages, facilitating morphological and functional diversification in A. calliptera, consistent with regulatory dynamics in teleost embryos, for example the *sox10b* gene, which shows promoter-associated upregulation at 7dpf (Tetrault *et al.*, 2023).



(a)



(b)



(c)

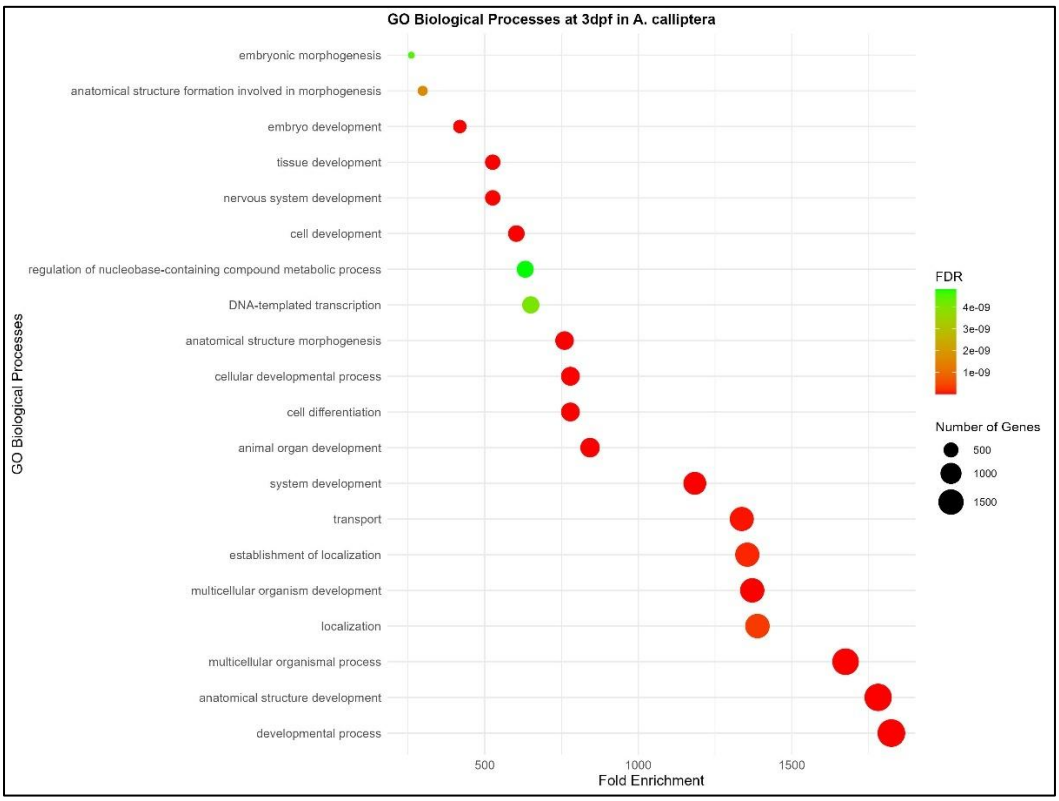
Figure 3. Enrichment of ATAC-seq peaks at (a) 3dpf, (b) 7dpf, and (c) 12dpf.

3.3 Functional Enrichment of Accessible Genes

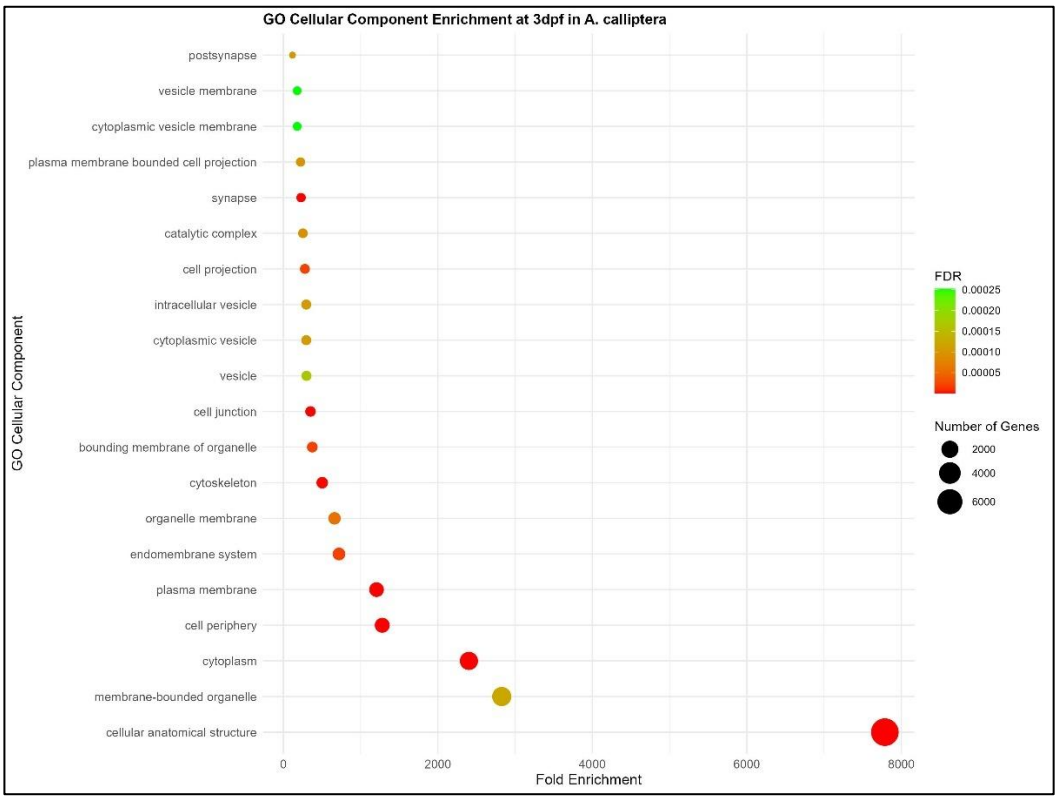
3.3.1 GO Term Enrichment Across Developmental Stages

Genes associated with ATAC-seq peaks were subjected to GO enrichment analysis using g:Profiler to identify active biological processes at each developmental stage. Enrichment is categorized into Biological Process, Molecular Function, and Cellular Component, illustrating stage-specific gene regulatory programs. At 3dpf, enriched GO:BP terms included embryonic

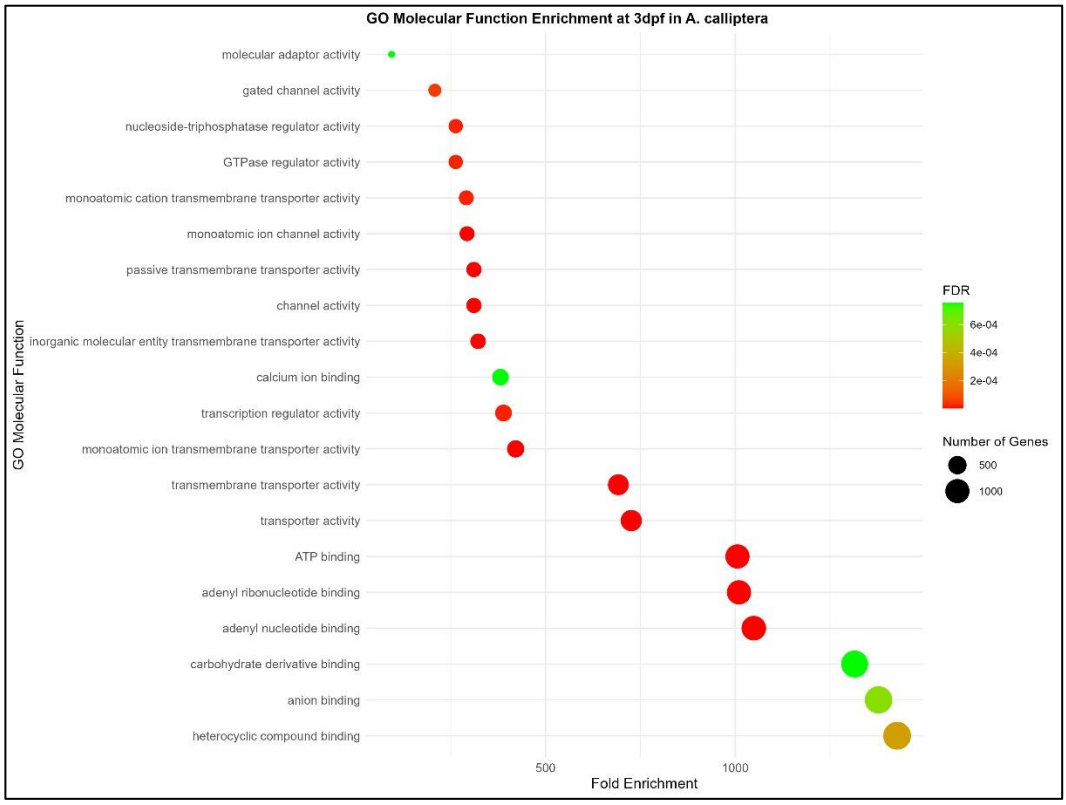
morphogenesis (fold enrichment = 262, p-value = 0), cell differentiation(fold enrichment = 792, p-value = $3.58e^{-20}$), and developmental process(fold enrichment = 1868, p-value = 0), reflecting foundational processes of early embryogenesis (Figure 4a-c). At 7dpf, a significant increase in enriched terms was observed, encompassing tissue development(fold enrichment = 532, p-value = 0), animal organ development (fold enrichment = 860, p-value = 0), and cytoskeletal organization (fold enrichment = 256, p-value = 0), indicating a peak in developmental complexity (Figure 4d-f). At 12dpf, enriched terms shifted toward functional specialization, including locomotory behaviour (fold enrichment = 56, p-value = 0.03), neurodevelopment (fold enrichment = 2816, p-value = 0), and digestive system development (fold enrichment = 55, p-value = 0.03), suggesting preparation for ecological adaptation (Figure 4g-i).



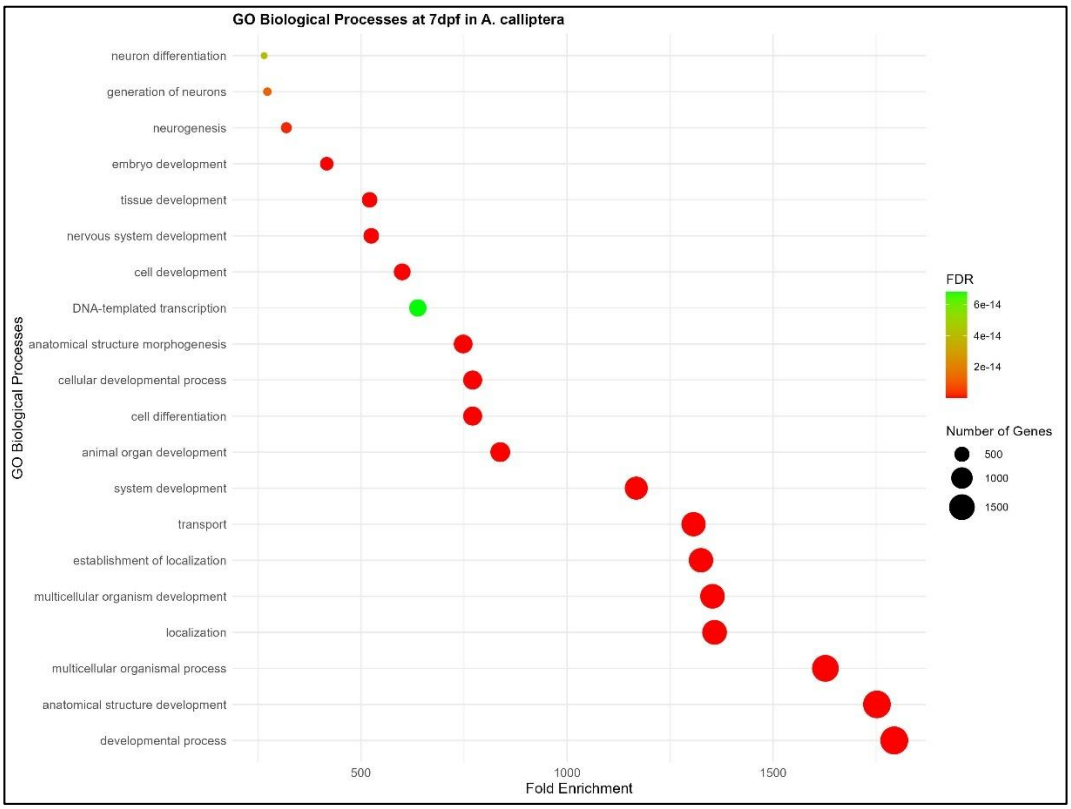
(a)


GO Cellular Component Enrichment at 3dpf in A. calliptera
postsynapse
vesicle membrane
cytoplasmic vesicle membrane
plasma membrane bounded cell projection
synapse
catalytic complex
cell projection
intracellular vesicle
cytoplasmic vesicle
vesicle
cell junction
bounding membrane of organelle
cytoskeleton
organelle membrane
endomembrane system
plasma membrane
cell periphery
cytoplasm
membrane-bounded organelle
cellular anatomical structure
GO Cellular Component
0
2000
4000
6000
8000
Fold Enrichment
FDR
0.00025
0.00020
0.00015
0.00010
0.00005
Number of Genes
2000
4000
6000

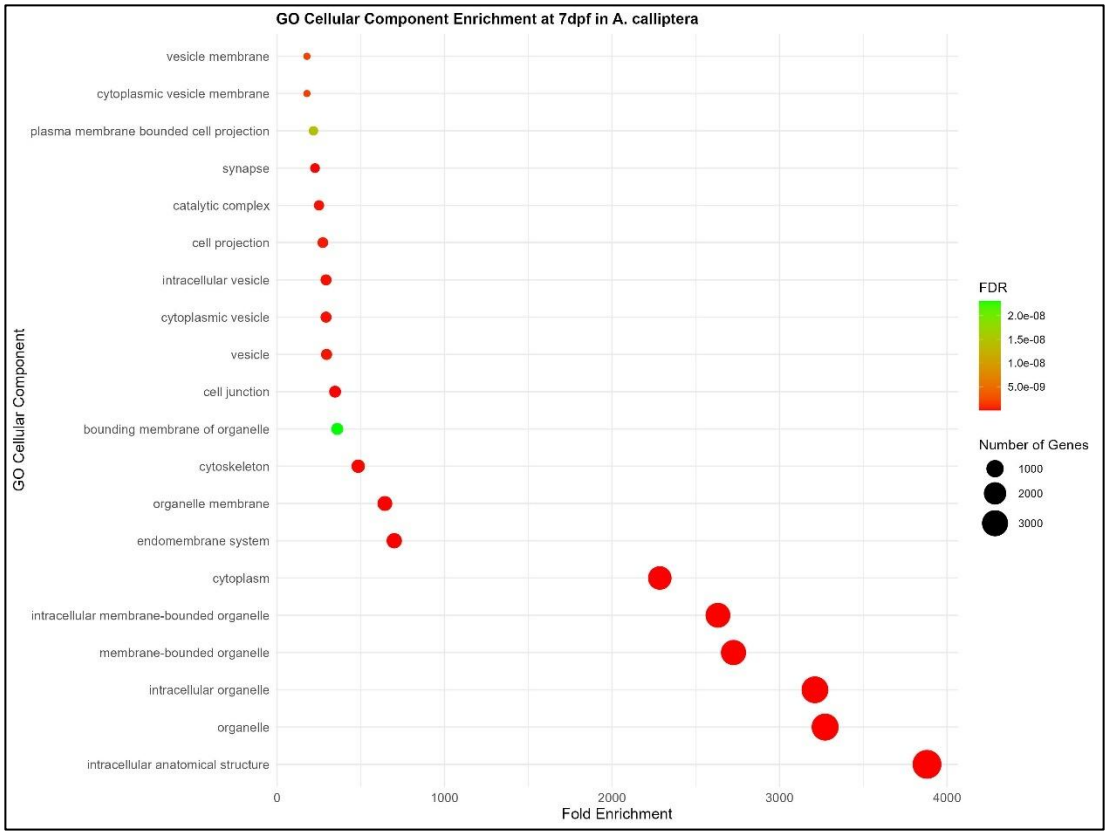
(b)


GO Molecular Function Enrichment at 3dpf in A. calliptera
molecular adaptor activity
gated channel activity
nucleoside-triphosphatase regulator activity
GTPase regulator activity
monoatomic cation transmembrane transporter activity
monoatomic ion channel activity
passive transmembrane transporter activity
channel activity
inorganic molecular entity transmembrane transporter activity
calcium ion binding
transcription regulator activity
monoatomic ion transmembrane transporter activity
transmembrane transporter activity
transporter activity
ATP binding
adenyl ribonucleotide binding
adenyl nucleotide binding
carbohydrate derivative binding
anion binding
heterocyclic compound binding
GO Molecular Function
500
1000
Fold Enrichment
FDR
6e-04
4e-04
2e-04
Number of Genes
500
1000

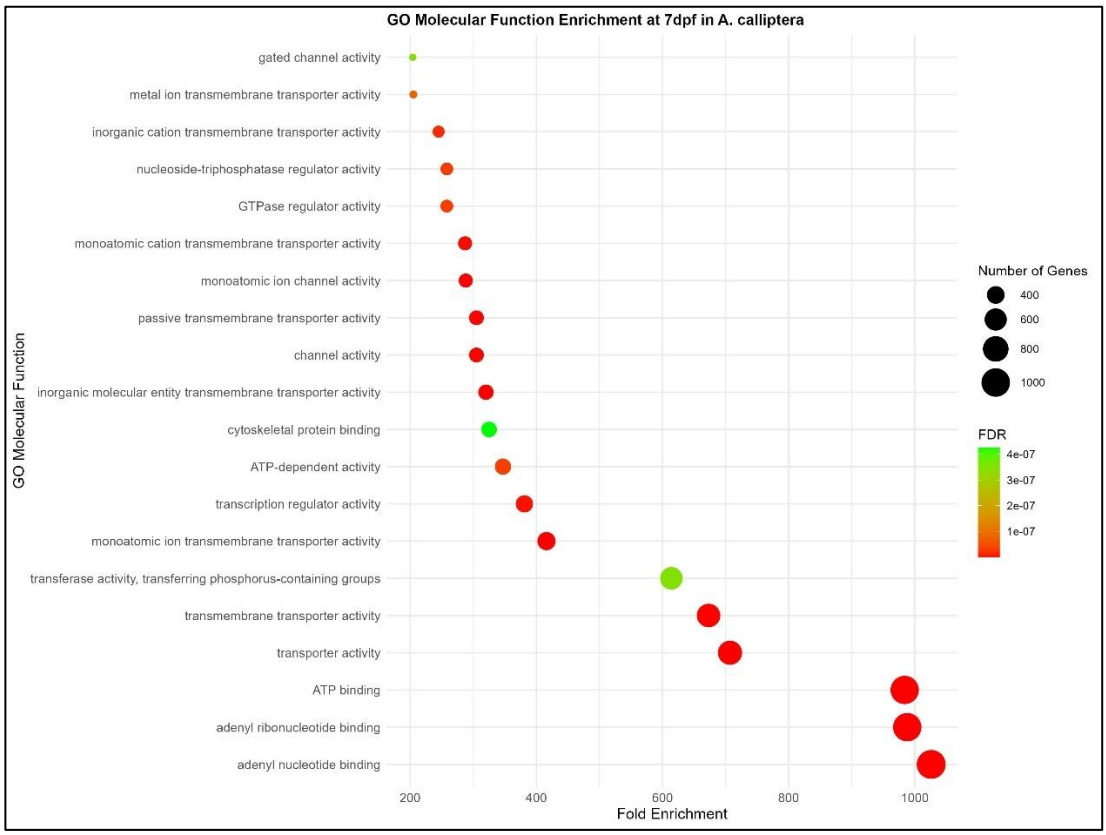
(c)


GO Biological Processes at 7dpf in A. calliptera
neuron differentiation
generation of neurons
neurogenesis
embryo development
tissue development
nervous system development
cell development
DNA-templated transcription
anatomical structure morphogenesis
cellular developmental process
cell differentiation
animal organ development
system development
transport
establishment of localization
multicellular organism development
localization
multicellular organismal process
anatomical structure development
developmental process
GO Biological Processes
500
1000
1500
Fold Enrichment
FDR
6e-14
4e-14
2e-14
Number of Genes
500
1000
1500

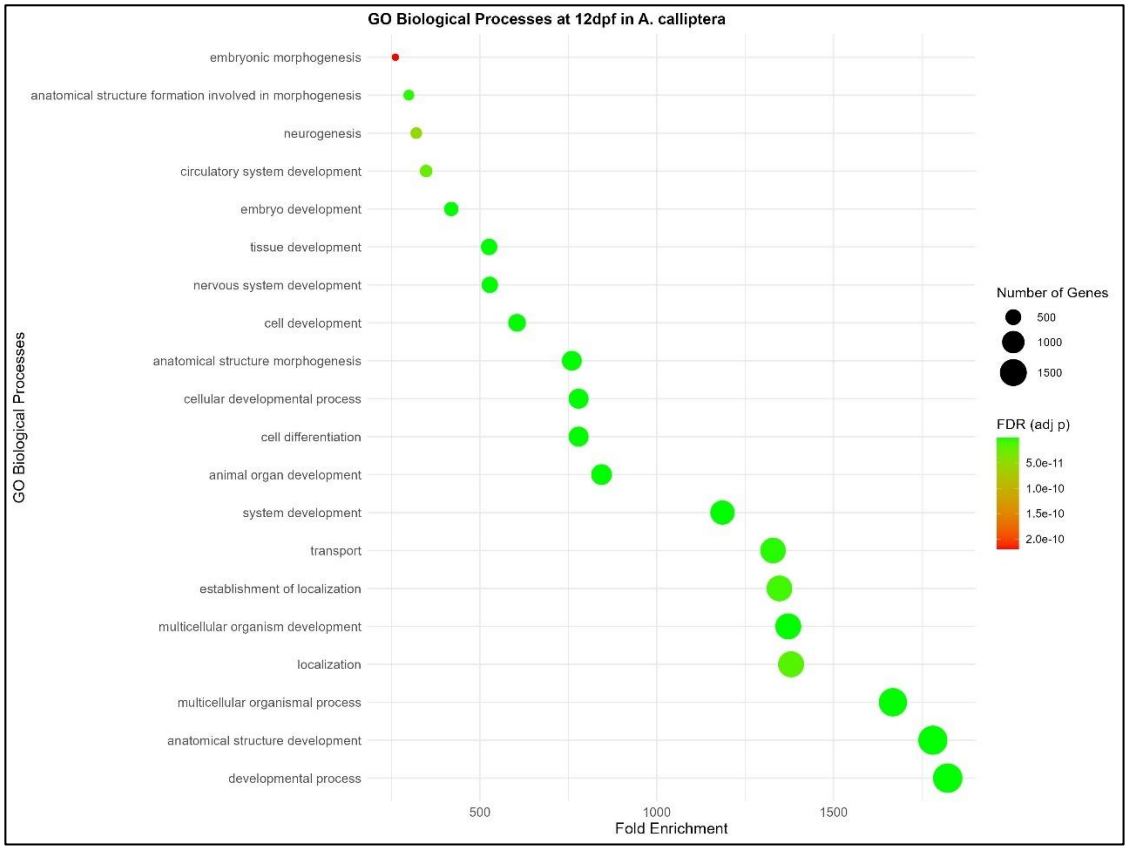
(d)


GO Cellular Component Enrichment at 7dpf in A. calliptera
vesicle membrane
cytoplasmic vesicle membrane
plasma membrane bounded cell projection
synapse
catalytic complex
cell projection
intracellular vesicle
cytoplasmic vesicle
vesicle
cell junction
bounding membrane of organelle
cytoskeleton
organelle membrane
endomembrane system
cytoplasm
intracellular membrane-bounded organelle
membrane-bounded organelle
intracellular organelle
organelle
intracellular anatomical structure
GO Cellular Component
0
1000
2000
3000
4000
Fold Enrichment
FDR
2.0e-08
1.5e-08
1.0e-08
5.0e-09
Number of Genes
1000
2000
3000

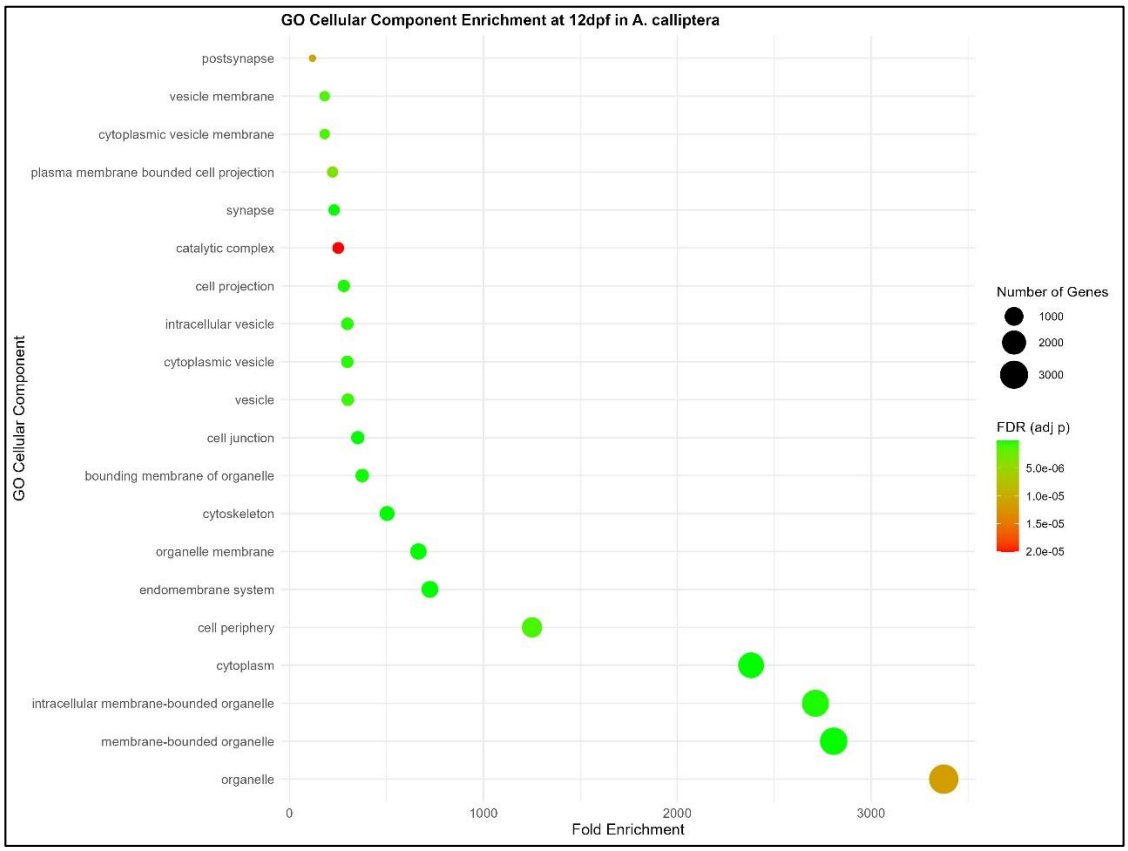
(e)


GO Molecular Function Enrichment at 7dpf in A. calliptera
gated channel activity
metal ion transmembrane transporter activity
inorganic cation transmembrane transporter activity
nucleoside-triphosphatase regulator activity
GTPase regulator activity
monoatomic cation transmembrane transporter activity
monoatomic ion channel activity
passive transmembrane transporter activity
channel activity
inorganic molecular entity transmembrane transporter activity
cytoskeletal protein binding
ATP-dependent activity
transcription regulator activity
monoatomic ion transmembrane transporter activity
transferase activity, transferring phosphorus-containing groups
transmembrane transporter activity
transporter activity
ATP binding
adenyl ribonucleotide binding
adenyl nucleotide binding
GO Molecular Function
200
400
600
800
1000
Fold Enrichment
Number of Genes
400
600
800
1000
FDR
4e-07
3e-07
2e-07
1e-07

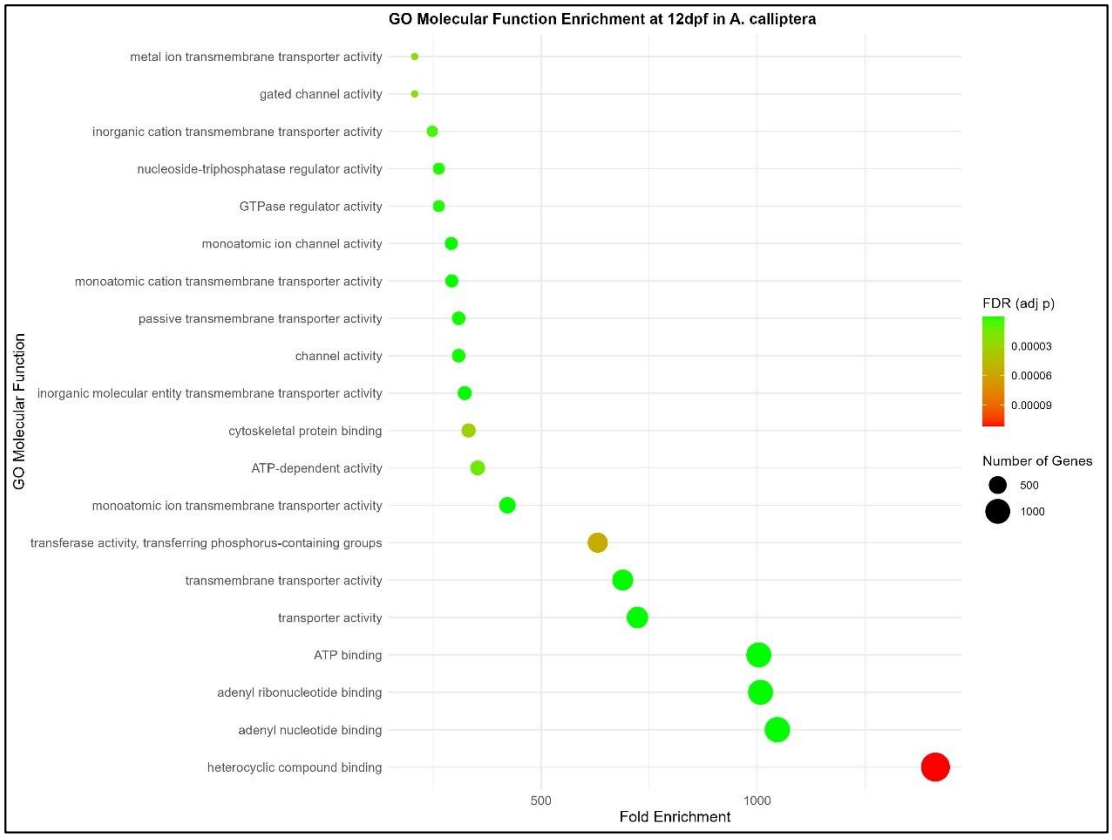
(f)


GO Biological Processes at 12dpf in A. calliptera
embryonic morphogenesis
anatomical structure formation involved in morphogenesis
neurogenesis
circulatory system development
embryo development
tissue development
nervous system development
cell development
anatomical structure morphogenesis
cellular developmental process
cell differentiation
animal organ development
system development
transport
establishment of localization
multicellular organism development
localization
multicellular organismal process
anatomical structure development
developmental process
GO Biological Processes
500
1000
1500
Fold Enrichment
Number of Genes
500
1000
1500
FDR (adj p)
5.0e-11
1.0e-10
1.5e-10
2.0e-10

(g)



(h)



(i)

Figure 4: GO term enrichment of accessible genes at (a-c) 3dpf, (d-f) 7dpf, and (g-i) 12dpf in *A. calliptera*.

3.3.2 Overlap and Specificity of GO Terms

Shared and unique GO:BP terms were analysed across the three developmental stages (3dpf, 7dpf, and 12dpf) to assess variations in biological processes (Fig.5, Fig6a-c). The results showed how GO:BP terms varied across stages. Ten GO:BP terms were shared across all three stages, predominantly involving core developmental processes such as 'embryonic morphogenesis' and 'cell differentiation'. These terms include housekeeping genes such as *gapdh* (glyceraldehyde-3-phosphate dehydrogenase). The number of unique biological process GO terms varied across stages, with 3dpf exhibiting only 1 unique term (amino acid metabolic process, involving *mthfr*, *shmt2* etc.), 7dpf displaying the highest number with 75 unique terms (e.g., actin cytoskeleton organization with *smyd1b*, *klf6a* etc.), and 12dpf showing 4 unique terms (e.g., locomotory behavior with *tardbp*, *hdac6* etc.) (Table 1).

Table 1: Important term names unique in each developmental stage

Stage	Term name	Examples of genes involved
3dpf	amino acid metabolic process	*mthfr, shmt2*
7dpf	Actin cytoskeleton organization	*smyd1b, klf6a*
7dpf	Cytoskeleton organization	*prmt7, ruvbl2*
7dpf	Digestive tract development	*dnmt1, eed*
7dpf	Developmental growth involved in morphogenesis	*katna1, fus*
12dpf	carbohydrate metabolic process	*hdac8, g6pd*
12dpf	locomotory behavior	*tardbp, hdac6*
12dpf	carbohydrate derivative catabolic process	*hk2, pfk*
12dpf	cell migration	*smarca4, phf8*

This reflects the dynamic nature of chromatin accessibility and gene regulation, where 7dpf may represent a peak in developmental complexity involving processes such as cytoskeleton organization, digestive tract development, and morphogenesis, while 12dpf transitions towards functional specialization with processes like locomotory behaviour and cell migration becoming prominent. Among the genes mentioned in Table 1, *Smyd1b* is a histone

methyltransferase essential for myofibril assembly and muscle development in zebrafish, suggesting a potential role in the evolution of muscle-related traits in cichlids (Li *et al.*, 2013). *Klf6a* is a transcription factor that regulates endothelial cell rearrangement during cardiovascular pruning, providing insights into the evolution of circulatory systems (Wen *et al.*, 2021). *Hdac6* regulates tubulin deacetylation and cilia assembly, indicating a contribution to sensory and motility adaptations in cichlids (Łysyganicz *et al.*, 2021).

One of the most striking findings is the enrichment of the GO term locomotory behaviour at 12dpf, absent in the earlier stages. Among the 56 genes associated with this term, such as *tardbp* and *hdac6*, all the genes were also linked to developmental process at 3dpf and 12dpf suggesting that these genes play dual roles in early developmental patterning and later functional specialization, highlighting a transition from broad developmental processes to specific behavioural and functional outcomes as the organism matures.

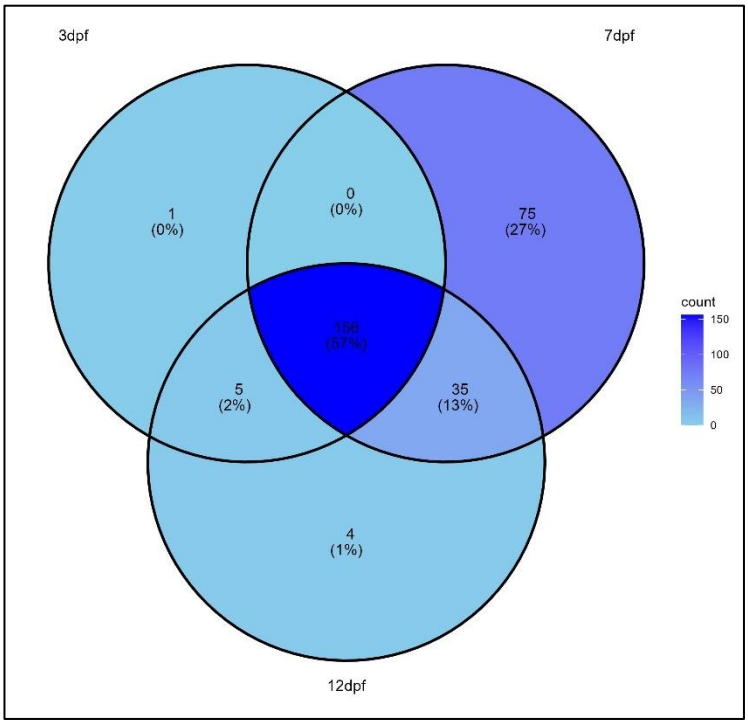
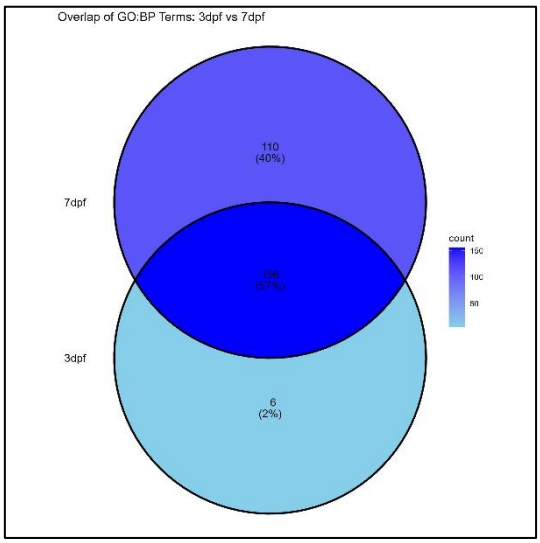
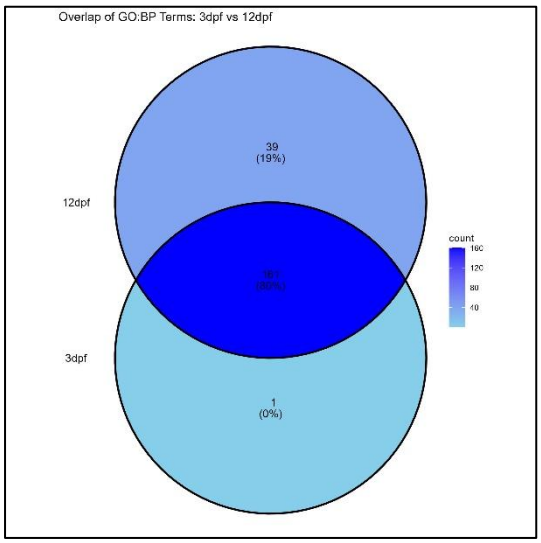


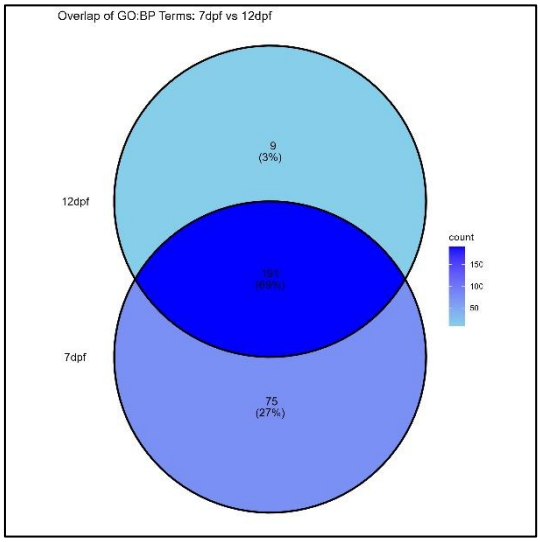
Figure 5: Venn diagram summarizing the number of unique Biological Process GO terms identified at each developmental stage.



(a)



(b)



(c)

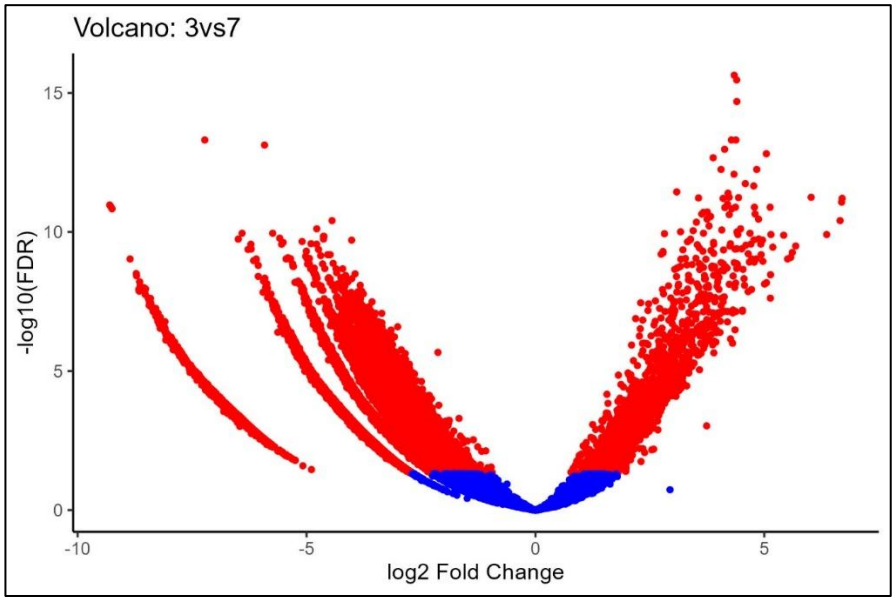
Figure 6: Venn diagrams comparing GO:BP term overlaps between (a) 3 dpf vs. 7 dpf, (b) 7 dpf vs. 12 dpf, and (c) 3 dpf vs. 12 dpf.

The observed progression of enriched GO terms aligns with developmental patterns described in cichlids, where early embryogenesis focuses on axis specification, followed by morphogenetic activity and functional specialization (Kratochwil *et al.*, 2015). The emergence of locomotion-related terms at later embryonic stages in *A. calliptera* may reflect species-specific ecological adaptations essential for habitat exploration and survival, consistent with behavioural diversity in cichlids (Parsons *et al.*, 2017).

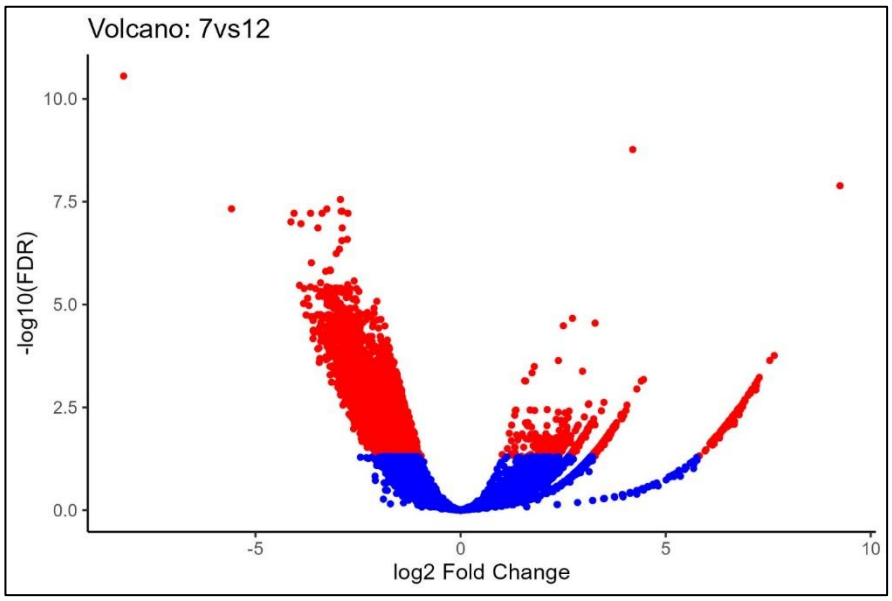
3.4 Differential Expression Analysis

3.4.1 Volcano and MA Plots of Differential Peaks

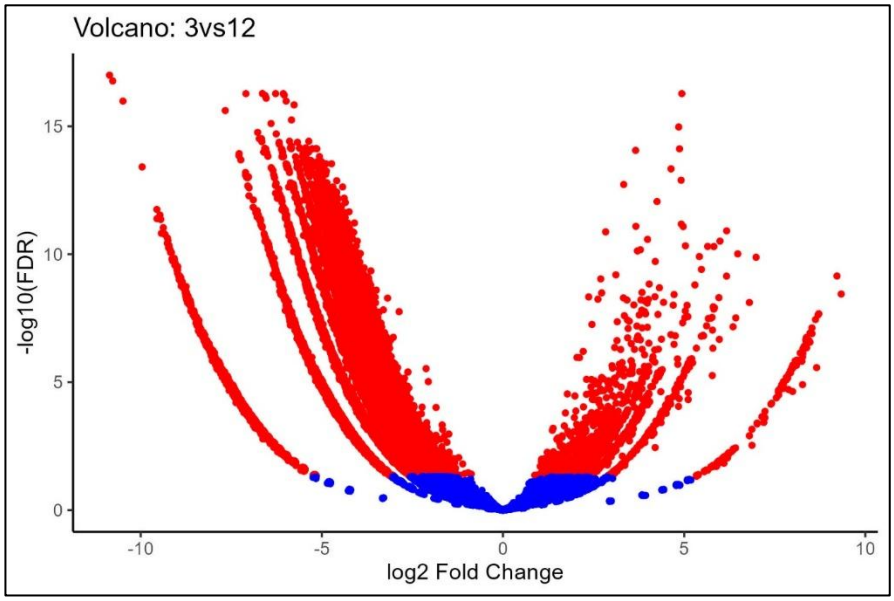
Differential chromatin accessibility analysis was performed to identify regions with significant changes in ATAC-seq peak intensity across pairwise comparisons of developmental stages: 3dpf vs. 7dpf, 3dpf vs. 12dpf, and 7dpf vs. 12dpf, focusing on gene promoters to associate with gene activity based on enrichments found in previous analyses. The analysis utilized 150 bp windows, with two biological replicates per stage, and filtered out low-count windows to ensure robust statistical power. Volcano plots were generated to visualize the $\log_2$ fold change against the false discovery rate (FDR). In the 3vs7 comparison (Fig. 7a), 78,675 peaks were significantly differentially accessible (FDR < 0.05), with 23,691 peaks overlapping promoter regions, corresponding to 8,687 unique genes. The 3vs12 (Fig. 7c) comparison identified 76,959 significant peaks, with 22,468 overlapping promoters, linked to 9,315 unique genes. In contrast, the 7vs12 (Fig. 7b) comparison showed fewer significant peaks (10,341), with 2,696 overlapping promoters, associated with 2,208 unique genes. These results indicate a higher degree of regulatory divergence in earlier developmental transitions.



(a)



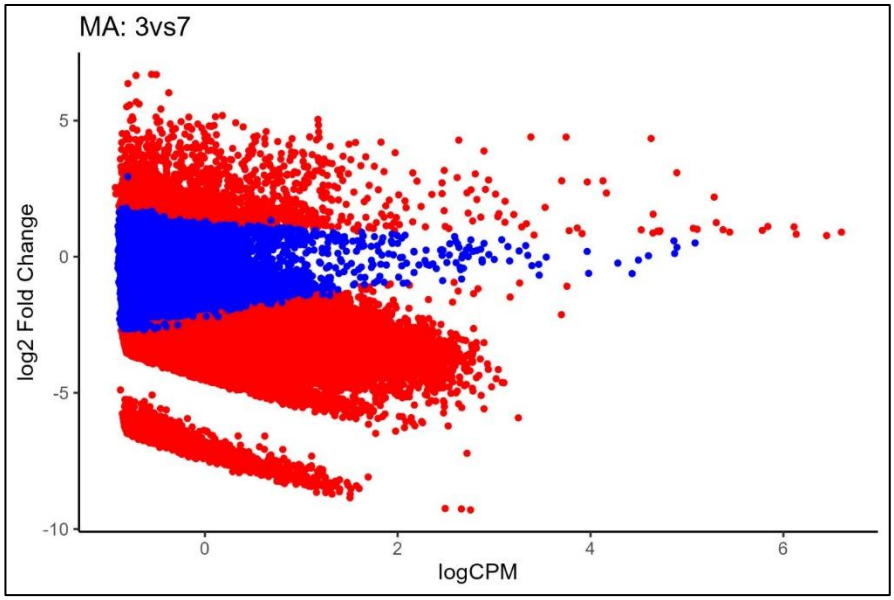
(b)



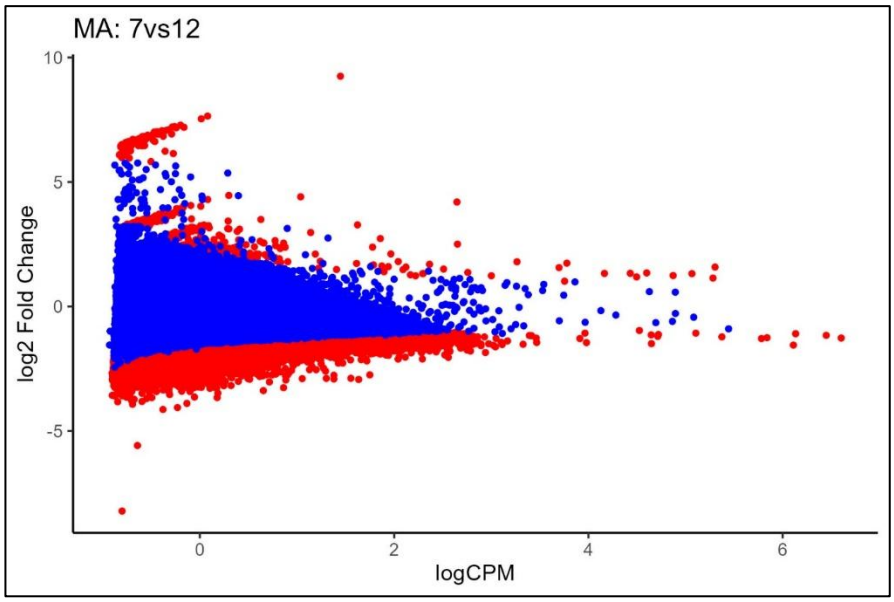
(c)

Figure 7: Volcano plot of differential peaks (a) 3vs7dpf (b) 7vs12dpf and (c) 3vs12dpf

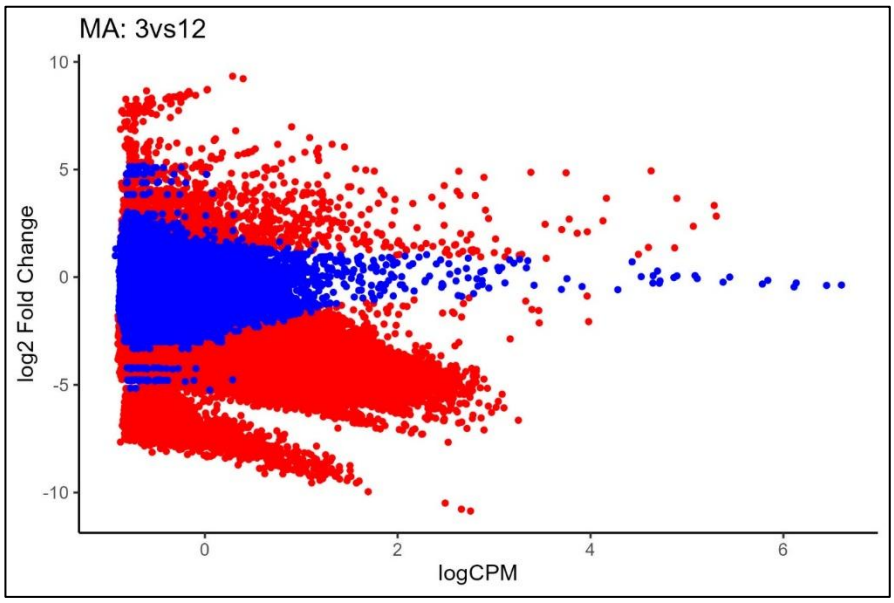
MA plots, plotting $\log_2$ fold change against mean accessibility, revealed a larger spread in negative fold changes in the 3vs7 (Fig. 8a) and 3vs12 (Fig. 8c) comparisons, indicating more pronounced decreases in accessibility at later stages. The 7vs12 (Fig. 8b) comparison showed a more balanced distribution, suggesting a stabilization of chromatin accessibility as development progresses.


MA: 3vs7
log2 Fold Change
5
0
-5
-10
0
2
4
6
logCPM

(a)


MA: 7vs12
log2 Fold Change
10
5
0
-5
0
2
4
6
logCPM

(b)



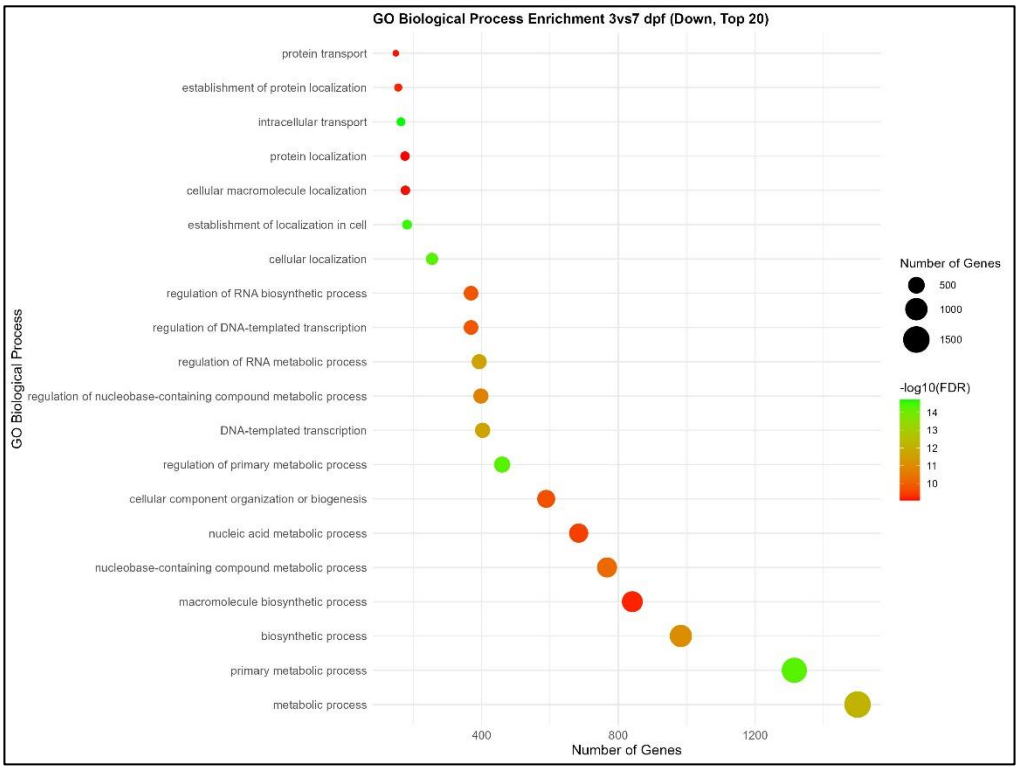
(c)

Figure 8: MA plot of differential expression analysis (a) 3vs7dpf (b) 7vs12dpf and (c) 3vs12dpf

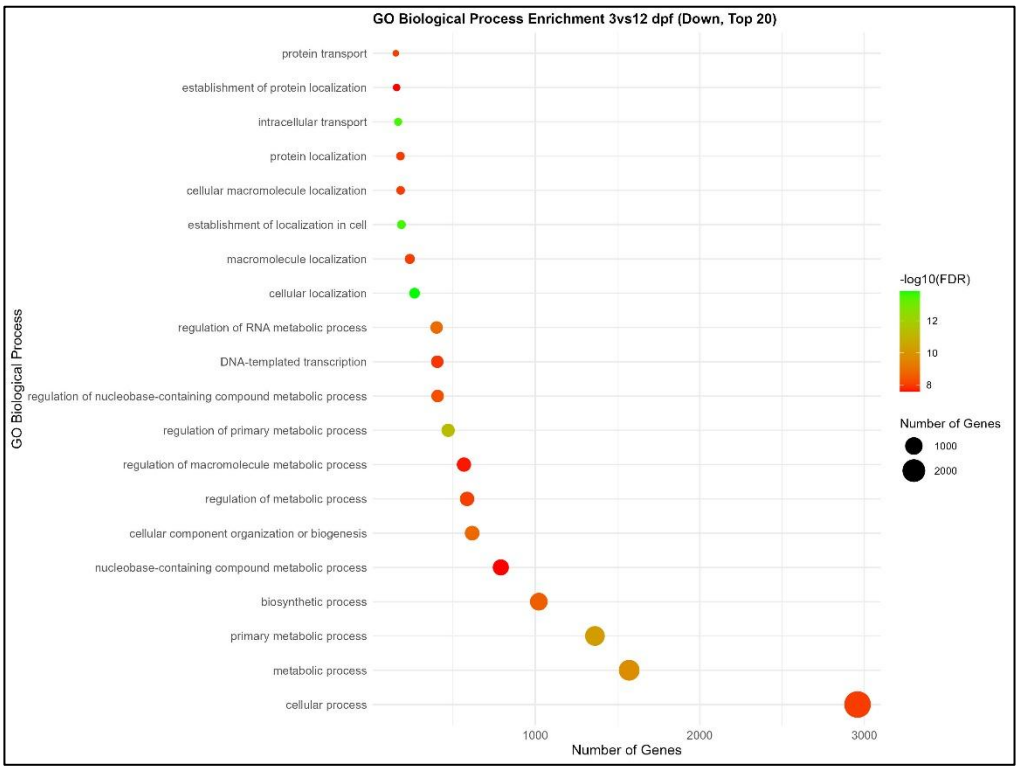
The volcano and MA plots reveal a dynamic epigenetic landscape in *Astatotilapia calliptera*, with the highest number of differentially accessible peaks in the 3vs7 and 3vs12 comparisons, suggesting extensive chromatin remodelling during early embryogenesis. The predominance of promoter-associated peaks, particularly in the 3vs7 and 3vs12 comparisons, indicates that promoter accessibility plays a critical role in activating developmental genes, consistent with studies in teleost fish where chromatin dynamics drive early developmental transitions (Tetrault *et al.*, 2023). The reduced number of significant peaks in the 7vs12 comparison suggests a stabilization of the chromatin landscape as the embryo progresses toward functional maturation, aligning with findings in zebrafish where later stages exhibit targeted regulatory changes (Bogdanović *et al.*, 2012). The high proportion of peaks with increased accessibility at 7dpf and 12dpf in the 3vs7 and 3vs12 comparisons, respectively, points to an intensification of gene regulation as tissues and organs form, supporting the hypothesis that epigenetic mechanisms underpin the rapid developmental progression in cichlids.

3.4.2 GO Term Enrichment of Differentially Accessible Peaks

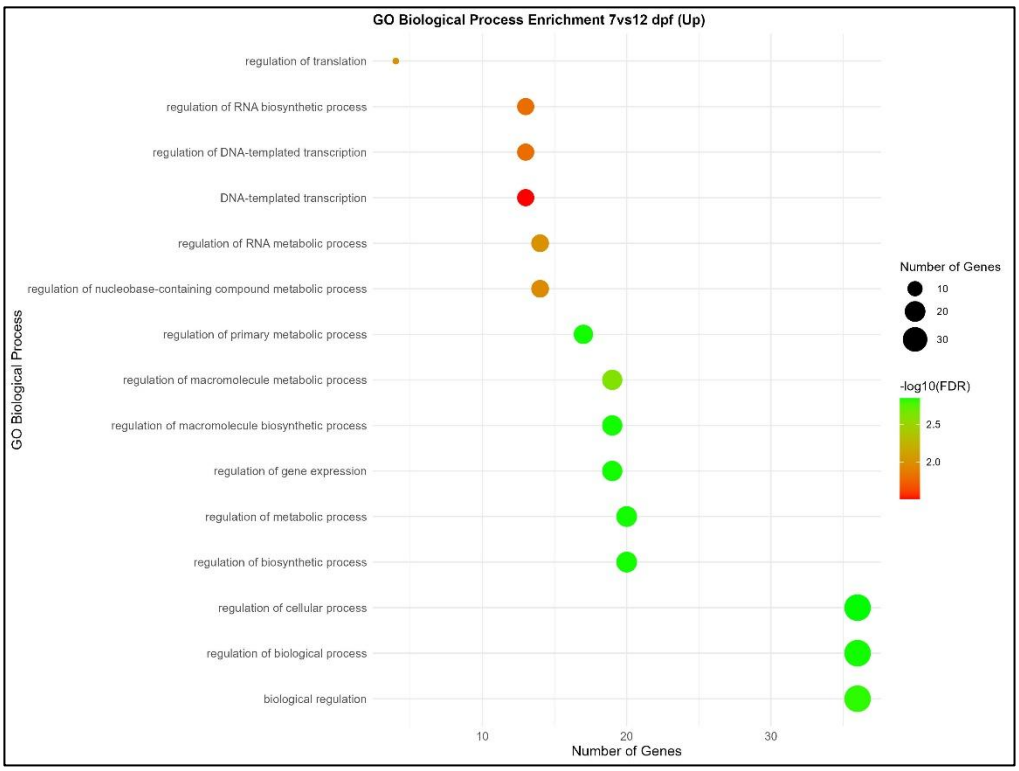
Gene Ontology (GO) Biological Process (BP) term enrichment analysis was conducted on genes linked to promoter-overlapping differentially accessible ATAC-seq peaks. Peaks were separated into up- and down-regulated categories for each stage and significant terms (p-value < 0.05) were visualized in Figure 9. Differential GO enrichment analysis revealed a striking difference in the number of significant terms across pairwise comparisons. Both up- and down-regulated peaks in the 7vs12 comparison exhibited only 15 significant GO:BP terms, indicating limited chromatin accessibility changes between these later stages. In contrast, up-regulated peaks in the 3vs7 and 3vs12 comparisons did not yield any significant terms, suggesting that promoter accessibility in these earlier transitions is predominantly associated with down-regulated or repressive processes. These observations highlight that the transition between 7dpf and 12dpf is less dynamic at the promoter level compared to earlier transitions and emphasize that 7dpf represents a peak in regulatory activity.



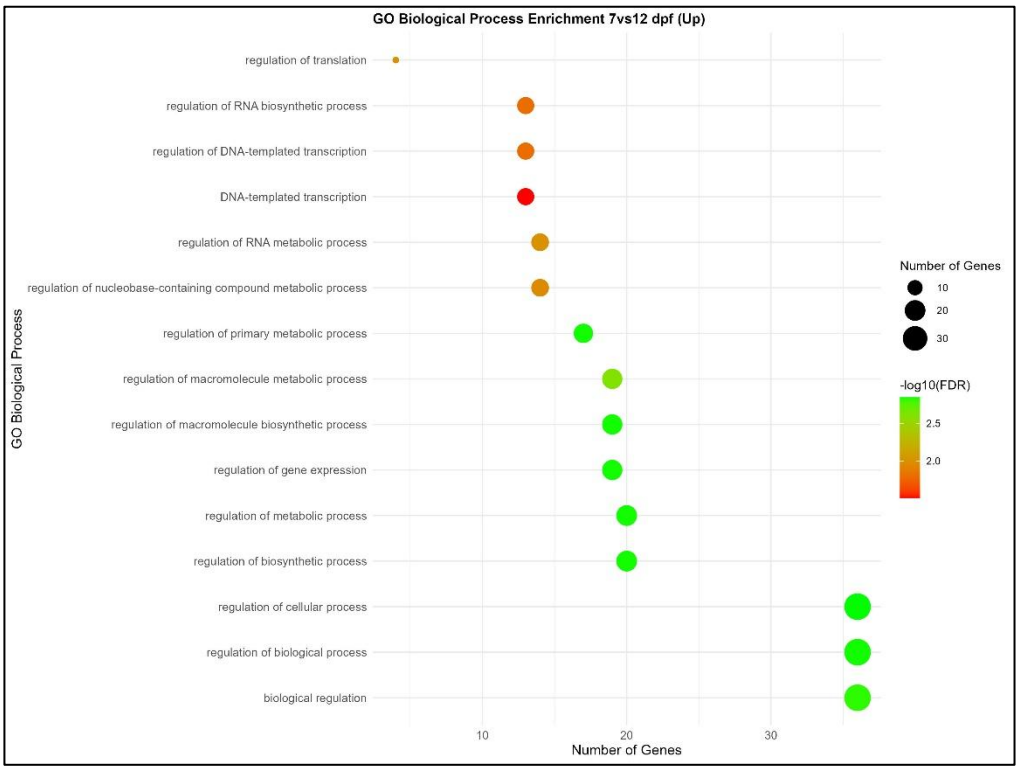
(a)


GO Biological Process Enrichment 3vs12 dpf (Down, Top 20)
protein transport
establishment of protein localization
intracellular transport
protein localization
cellular macromolecule localization
establishment of localization in cell
macromolecule localization
cellular localization
regulation of RNA metabolic process
DNA-templated transcription
regulation of nucleobase-containing compound metabolic process
regulation of primary metabolic process
regulation of macromolecule metabolic process
regulation of metabolic process
cellular component organization or biogenesis
nucleobase-containing compound metabolic process
biosynthetic process
primary metabolic process
metabolic process
cellular process
GO Biological Process
1000
2000
3000
Number of Genes
-log10(FDR)
12
10
8
Number of Genes
1000
2000

(b)


GO Biological Process Enrichment 7vs12 dpf (Up)
regulation of translation
regulation of RNA biosynthetic process
regulation of DNA-templated transcription
DNA-templated transcription
regulation of RNA metabolic process
regulation of nucleobase-containing compound metabolic process
regulation of primary metabolic process
regulation of macromolecule metabolic process
regulation of macromolecule biosynthetic process
regulation of gene expression
regulation of metabolic process
regulation of biosynthetic process
regulation of cellular process
regulation of biological process
biological regulation
GO Biological Process
10
20
30
Number of Genes
Number of Genes
10
20
30
-log10(FDR)
2.5
2.0

(c)



(d)

Figure 9: GO term enrichment of differential accessibility of (a) 3vs7dpf down, (b) 3vs12dpf down, (c) 7vs12dpf up and (d) 7vs12dpf down.

The common and unique terms across all the stages were compared. Common terms across all three comparisons included GO:BP terms such as "regulation of primary metabolic process" and "developmental process". Notable unique terms across the stages primarily associated with 7dpf that may relate to epigenetic modifications included "head development" and "eye morphogenesis" . The "head development" term included genes such as dnmt1, *otx2a*, *otx2b*, and *pbx2*, associated with cranial structure formation (Mork and Crump, 2015). The "eye morphogenesis" term included *dnmt1*, *pbx2*, and *sox11a*, linked to eye structure development (Pillai-Kastoori *et al.*, 2014). The faceted bar plot (Fig. 10) shows the number of genes associated with head and eye development GO terms for the differential chromatin accessibility comparisons, with significant genes detected only in the 3vs7 down- and 3vs12 down-regulated peaks.

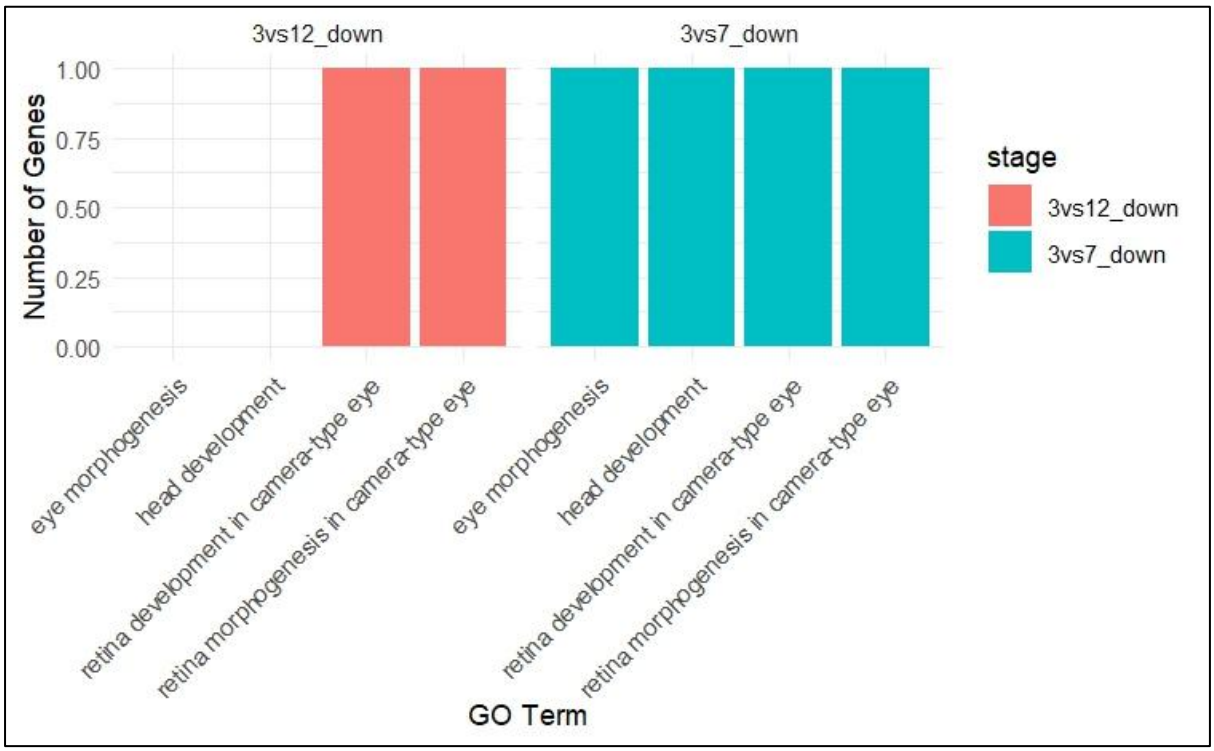


Figure 10: Stage-specific enrichment of selected GO:BP terms related to head and eye development in *Astatotilapia calliptera*.

The GO terms "head development" and "eye morphogenesis" indicate that chromatin accessibility regulates genes shaping early cranial and eye structures, critical for cichlid diversification. For "head development," *dnmt1*, *otx2a*, and *otx2b* suggest epigenetic control of jaw and facial morphology. In Lake Malawi cichlids, *dnmt1* drives DNA methylation, influencing morphological traits like jaw shape (Vernaz *et al.*, 2021). In zebrafish, *otx2* is regulated by DNA methylation and histone modifications during retinal and neural development, suggesting *otx2a* and *otx2b* control cichlid head structures (Seritrakul and Gross, 2019). For "eye morphogenesis," *dnmt1*, *pbx2*, and *sox11a* indicate epigenetic regulation of eye structures. In zebrafish, chromatin accessibility regulates homeobox genes like *rx3* during optic cup formation, implying *pbx2*, a homeobox gene, may have a similar role in cichlid eye development (Buono *et al.*, 2021). These epigenetic mechanisms likely drive heritable changes in head and eye morphology, supporting cichlid adaptation to diverse ecological niches. The results suggest major epigenetic changes occur early stages.

Conclusions

This study characterized the epigenetic dynamics of early embryonic development in *Astatotilapia calliptera*, a key species in the Lake Malawi cichlid radiation, using ATAC-seq to profile chromatin accessibility at 3, 7, and 12-days post-fertilization (dpf). By mapping open chromatin regions, annotating peaks to genomic features, and performing enrichment analyses, we identified distinct regulatory landscapes that shift across developmental stages. Analysis revealed unique chromatin accessibility patterns at each stage. Enrichment analyses indicated stage-specific regulatory shifts: elevated highly conserved non-coding elements (hCNEs) and 5' UTRs at 3dpf, increased promoter activity at 7dpf, and sustained hCNEs/UTRs at 12dpf. Gene Ontology (GO) terms highlighted embryonic morphogenesis at 3dpf, tissue and organ formation at 7dpf, and locomotory behaviour at 12dpf. Differential accessibility analysis showed extensive changes in early stages but fewer in later stages, with down-regulated terms such as head development and eye morphogenesis linked to genes like *dnmt1*, *otx2a*, *otx2b*, and *pbx2*. These findings suggest that dynamic chromatin accessibility drives gene regulation during early development, likely contributing to cichlid phenotypic diversity. This research establishes a critical framework for exploring the molecular mechanisms of adaptive radiation in cichlids, underscoring the significance of epigenetic regulation in evolutionary diversification.

Acknowledgments

I express my sincere gratitude to my supervisor, Dr. Tarang Mehta, for his expert guidance and support throughout this project. I am deeply thankful to my parents and brother for their unwavering encouragement and motivation. I also acknowledge the valuable insights provided by other teachers at the University of Liverpool's School of Biosciences, which enriched my learning experience. Finally, I extend my appreciation to my friends for their constant support and camaraderie during this research endeavour.

References

Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C. and Braun, T. (2020) 'ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation', *Nature communications*, 11(1), p. 4267. Available at: 10.1038/s41467-020-18035-1

Bogdanović, O., Fernandez-Miñán, A., Tena, J.J., Calle-Mustienes, E.d.l., Hidalgo, C., Kruysbergen, I.v., Heeringen, S.J.v., Veenstra, G.J.C. and Gómez-Skarmeta, J.L. (2012) 'Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis', *Genome Research*, 22(10) Available at: 10.1101/gr.134833.111

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W. and Bezault, E. (2014) 'The genomic substrate for adaptive radiation in African cichlid fish', *Nature*, 513(7518), pp. 375-381. Available at: 10.1038/nature13726

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J. (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nature methods*, 10(12), pp. 1213-1218. Available at: 10.1038/nmeth.2688

Buono, L., Corbacho, J., Naranjo, S., Almuedo-Castillo, M., Moreno-Marmol, T., de la Cerda, B., Sanabria-Reinoso, E., Polvillo, R., Díaz-Corrales, F.-J., Bogdanovic, O., Bovolenta, P., Martínez-Morales, J.-R., Buono, L., Corbacho, J., Naranjo, S., Almuedo-Castillo, M., Moreno-Marmol, T., de la Cerda, B., Sanabria-Reinoso, E., Polvillo, R., Díaz-Corrales, F.-J., Bogdanovic, O., Bovolenta, P. and Martínez-Morales, J.-R. (2021) 'Analysis of gene network bifurcation during optic cup morphogenesis in zebrafish', *Nature Communications 2021 12:1*, 12(1) Available at: 10.1038/s41467-021-24169-7

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) 'BLAST+: architecture and applications', *BMC bioinformatics*, 10(1), p. 421. Available at: 10.1186/1471-2105-10-421

Chen, Y., Chen, L., Lun, Aaron T.L., Baldoni, Pedro L. and Smyth, Gordon K. (2025) 'edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets', *Nucleic Acids Research*, 53(2) Available at: 10.1093/nar/gkaf018

Heger, A., Webber, C., Goodson, M., Ponting, C.P. and Lunter, G. (2013) 'GAT: a simulation framework for testing the association of genomic intervals', *Bioinformatics*, 29(16), pp. 2046-2048. Available at: 10.1093/bioinformatics/btt343

Kocher, T.D. (2004) 'Adaptive evolution and explosive speciation: the cichlid fish model', *Nature Reviews Genetics*, 5(4), pp. 288-298. Available at: 10.1038/nrg1316

Kratochwil, C.F., Sefton, M.M., Meyer, A., Kratochwil, C.F., Sefton, M.M. and Meyer, A. (2015) 'Embryonic and larval development in the Midas cichlid fish species flock (Amphilophus spp.): a new evo-devo model for the investigation of adaptive novelties and species differences', *BMC Developmental Biology 2015 15:1*, 15(1), p. 12. Available at: 10.1186/s12861-015-0061-1

Krueger, F. (2015) 'Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data', *Babraham Institute*.

Langmead, B. and Salzberg, S.L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature methods*, 9(4), pp. 357-359. Available at: 10.1038/nmeth.1923

Lawrence, M., Gentleman, R. and Carey, V. (2009) 'rtracklayer: an R package for interfacing with genome browsers', *Bioinformatics*, 25(14) Available at: 10.1093/bioinformatics/btp328

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G.P.D.P. (2009) 'The sequence alignment/map format and SAMtools', *bioinformatics*, 25(16), pp. 2078-2079. Available at: 10.1093/bioinformatics/btp352

Li, H., Zhong, Y., Wang, Z., Gao, J., Xu, J., Chu, W., Zhang, J., Fang, S. and Du, S.J. (2013) 'Smyd1b is required for skeletal and cardiac muscle function in zebrafish', *Molecular Biology of the Cell*, 24(22) Available at: 10.1091/mbc.e13-06-0352

Li, Z., Schulz, M.H., Look, T., Begemann, M., Zenke, M. and Costa, I.G. (2019) 'Identification of transcription factor binding sites using ATAC-seq', *Genome biology*, 20(1), p. 45. Available at: 10.1186/s13059-019-1642-2

Louise Smith, E., Mok, G.F. and Münsterberg, A. (2022) 'Investigating chromatin accessibility during development and differentiation by ATAC-sequencing to guide the identification of cis-regulatory elements', *Biochemical Society Transactions*, 50(3), pp. 1167-1177. Available at: 10.1042/BST20210834

Lun, A.T.L. and Smyth, G.K. (2015) 'From reads to regions: a Bioconductor workflow to detect differential binding in ChIP-seq data', *F1000Research*, 4 Available at: 10.12688/f1000research.7016.2

Łysyganicz, P.K., Pooranachandran, N., Liu, X., Adamson, K.I., Zielonka, K., Elworthy, S., van Eeden, F.J., Grierson, A.J. and Malicki, J.J. (2021) 'Frontiers | Loss of Deacetylation Enzymes Hdac6 and Sirt2 Promotes Acetylation of Cytoplasmic Tubulin, but Suppresses Axonemal Acetylation in Zebrafish Cilia', *Frontiers in Cell and Developmental Biology*, 9 Available at: 10.3389/fcell.2021.676214

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F. and Durbin, R. (2018) 'Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow', *Nature ecology & evolution*, 2(12), pp. 1940-1955. Available at: 10.1038/s41559-018-0717-x

Mehta, T.K., Man, A., Ciezarek, A., Ranson, K., Penman, D., Di-Palma, F. and Haerty, W. (2023) 'Chromatin accessibility in gill tissue identifies candidate genes and loci associated with aquaculture relevant traits in tilapia', *Genomics*, 115(4), p. 110633. Available at: 10.1016/j.ygeno.2023.110633

Mork, L. and Crump, G. (2015) 'Zebrafish Craniofacial Development: A Window into Early Patterning', *Current Topics in Developmental Biology*, 115 Available at: 10.1016/bs.ctdb.2015.07.001

Parsons, P.J., Bridle, J.R., Rüber, L. and Genner, M.J. (2017) 'Evolutionary divergence in life history traits among populations of the Lake Malawi cichlid fish Astatotilapia calliptera', *Ecology and evolution*, 7(20), pp. 8488-8506. Available at: 10.1002/ece3.3311

Pillai-Kastoori, L., Wen, W., Wilson, S.G., Strachan, E., Lo-Castro, A., Fichera, M., Musumeci, S.A., Lehmann, O.J. and Morris, A.C. (2014) 'Sox11 Is Required to Maintain Proper Levels of Hedgehog Signaling during Vertebrate Ocular Morphogenesis', *PLOS Genetics*, 10(7) Available at: 10.1371/journal.pgen.1004491

Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841-842. Available at: 10.1093/bioinformatics/btq033

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. and Vilo, J. (2019) 'g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)', *Nucleic acids research*, 47(W1), pp. W191-W198. Available at: 10.1093/nar/gkz369

Seritrakul, P. and Gross, J.M. (2019) 'Genetic and epigenetic control of retinal development in zebrafish', *Current Opinion in Neurobiology*, 59, pp. 120-127. Available at: 10.1016/j.conb.2019.05.008

Simon Andrews, F., Segonds-Pichon, A., Biggins, L., Krueger, C. and Wingett, S. (2010) 'FastQC: a quality control tool for high throughput sequence data'.

Svardal, H., Salzburger, W. and Malinsky, M. (2021) 'Genetic variation and hybridization in evolutionary radiations of cichlid fishes', *Annual Review of Animal Biosciences*, 9(1), pp. 55-79. Available at: 10.1146/annurev-animal-061220-023129

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. and Prins, P. (2015) 'Sambamba: fast processing of NGS alignment formats', *Bioinformatics*, 31(12), pp. 2032-2034. Available at: 10.1093/bioinformatics/btv098

Tetrault, E., Swenson, J., Aaronson, B., Marcho, C. and Albertson, R.C. (2023) 'The transcriptional state and chromatin landscape of cichlid jaw shape variation across species and environments', *Molecular ecology*, 32(14), pp. 3922-3941. Available at: 10.1111/mec.16975

Vernaz, G., Malinsky, M., Svardal, H., Du, M., Tyers, A.M., Santos, M.E., Durbin, R., Genner, M.J., Turner, G.F. and Miska, E.A. (2021) 'Mapping epigenetic divergence in the massive radiation of Lake Malawi cichlid fishes', *Nature Communications*, 12(1), p. 5870. Available at: 10.1038/s41467-021-26166-2

Wen, L., Zhang, T., Wang, J., Jin, X., Rouf, M.A., Luo, D., Zhu, Y., Lei, D., Gregersen, H., Wang, Y. and Wang, G. (2021) 'The blood flow-klf6a-tagln2 axis drives vessel pruning in zebrafish by regulating endothelial cell rearrangement and actin cytoskeleton dynamics', *PLOS Genetics*, 17(7) Available at: 10.1371/journal.pgen.1009690

Wickham, H. (2016) 'Data analysis', in *ggplot2: elegant graphics for data analysis*. Springer, pp. 189-201.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M. and Li, W. (2008) 'Model-based analysis of ChIP-Seq (MACS)', *Genome biology*, 9(9), p. R137. Available at: 10.1186/gb-2008-9-9-r137

Appendices

Please find all the codes and supplementary materials attached to the GitHub repository at:

https://github.com/Midhu-krishna/A.calliptera-ATAC-analysis.git