**Contents**

|  |  |  |
| --- | --- | --- |
| **No** | **Title** | **Page no.** |
| 1. | Acronyms | 2 |
| 2. | Meeting Summary | 3 |
| 3. | Planning | 5 |
| 4. | Reading articles | 7 |
| 5. | Methods and Errors encountered | 14 |
| 6. | Databases and Links | 17 |
| 7. | Codes used | 18 |
| 8. | Results and Statistical Analysis | 19 |
| 9. | Thoughts on the Project | 35 |

**Acronyms**

ATAC-seq – Assay for Transposase-Accessible Chromatin using sequencing

TF – Transcription Factor

HPC – High Performance Computing

QC – Quality Control

TSS – Transcription Start Site

UTR – Untranslated Region

CNE / hCNE – Conserved Noncoding Element / highly Conserved Noncoding Element

GO – Gene Ontology

FDR – False Discovery Rate

GAT – Genomic Association Tester

MA plot – log ratio (M) versus mean average (A) plot

g:Profiler – Gene functional profiling tool

STQ – File format for sequencing reads with quality scores

BAM – Binary Alignment Map

CSV / TSV – Comma-Separated Values / Tab-Separated Values file formats

DNA – Deoxyribonucleic Acid

dpf – Days Post Fertilization

**Meeting summary**

**June 12, 2025**

* Discussed about results from GAT.
* Check the GAT results again as it may not be installed in HPC and the final output is wrong
* Don’t move onto enrichment using G:profiler without completing GAT.
* Focus on Biological processes in the G:profiler results
* Note: GAT was not installed in HPC. Manually installed in own laptop for further anlalysis

**June 26, 2025**

* Discussed the progress with graphs from GAT.
* Reviewed enrichment plots prepared:
* Shows high activity in promoters, exons, and conserved non-coding elements (CNEs), with variations across stages
* Adjust graph scales - remove log10 for better visualization of FDR
* Focus on Biological processes (BP) in G:profiler results.
* Use Venn diagrams for term comparisons.

**July 17, 2025**

* Discussed the progress of gene ontology enrichment step and comparisons across stages.
* Use Venn diagrams for finding overlaps
* Send raw lists/tables before plotting final visuals

**August 14, 2025**

* First draft of final report was mailed before the meeting. Discussed on how to improve the draft and missing parts:
* Results discussion: Focused on locomotive behavior genes (56 involved)
* Pick examples based on intersections, uniqueness, or prior studies rather than all genes.
* Use EdgeR for differential accessibility (needs replicates, large BAM files; run on cluster to avoid memory issues).
* BAM files with replication and BED file are needed for next step.

**August 28, 2025**

* Discussed the progress on differential accessibility.
* Talked about EdgeR plots (volcano, MA):
* Higher differences in 3 vs. 12 stages; used replicates (not merged files);
* Similar accessibility in close stages (7 vs. 12); high fold changes expected; don't include dispersion plots if it complicates discussion.
* Error in analysis: zero overlaps occur due to wrong BED file
* Next steps: Run GProfiler on differentially accessible regions, extract gene IDs for enrichment

**Planning**

Date: June 2–12, 2025

* Study and understand codes of GAT
* Prepare initial enrichment plots for 3, 7, 12 dpf.
* Set up shared Google Doc for record book.
* Submit preliminary report draft by June 6.

Date: June 13-23, 2025

* Check if GAT is installed on HPC
* Install GAT otherwise and modify codes accordingly
* Refine GAT plots for next meeting (June 26).
* Start GProfiler runs for BP terms.

Date: June 27–July 16, 2025

* Revise plots; check adjusted q-values.
* Find common and unique genes
* Prepare Venn diagrams for next meeting.
* Send raw lists/tables of terms and q-values.

Date: July 18–August 13, 2025

* Write results draft.
* Focus on BP in G:profiler results.
* Send lists/tables before Venn diagrams to validate overlaps.
* Send pairwise comparison lists (3 vs. 7, etc.).
* Start differential accessibility analysis with EdgeR.

Date: August 15–27, 2025

* Edit the abstract, intro, methods and results section as per comments
* Run EdgeR
* Generate volcano/MA plots
* Finish the final part of results section

Date: August 29–September 3, 2025

* Mail final draft to
* Edit suggested changes

**Reading Articles**

* 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.

Summary: Reviewed the phenomenon of adaptive radiation in East African cichlids, emphasizing their status as a textbook model for rapid speciation and ecological diversification. Documents how hundreds of cichlid species emerged in less than a million years in lakes such as Malawi and Victoria, with ecological factors like habitat specialization and sexual selection contributing to radiation. Also discusses the role of genetic variation in fueling this diversification, while highlighting the need to explore developmental and regulatory mechanisms.

* 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

Summary: Sequenced the genomes of several cichlid species and discovered that gene duplications, regulatory sequence changes, and repeated hybridization events have contributed significantly to their diversity. Their results showed that noncoding regions often exhibit signatures of selection, pointing to regulatory elements as major drivers of adaptive traits. They also noted that cichlid genomes show high levels of shared polymorphism, enabling rapid adaptation.

* 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

Summary: Carried out population genomic analyses of Lake Malawi cichlids and found that species within the lake exhibit extensive hybridization and gene flow. Instead of distinct species boundaries, they observed complex ancestry relationships across lineages, with different species groups exchanging genes. This finding redefined cichlid evolution as a reticulate process rather than a simple branching tree. For *A. calliptera*, the study highlighted its mixed ancestry with both riverine and lacustrine groups, positioning it as a keystone species for understanding the genomic and regulatory basis of radiation.

* 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

Summary: Synthesized genetic studies on East African cichlids and emphasized the roles of hybridization, standing variation, and introgression in driving diversification. They pointed out that while genetic mechanisms are well studied, epigenetic contributions remain poorly understood.

* 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

Summary: Studied the ecology and morphology of *A. calliptera*, showing that it thrives in both rivers and lakes, suggesting remarkable ecological plasticity. They proposed that *A. calliptera* may represent either the ancestral progenitor of the Malawi cichlid radiation or an ongoing contributor to its gene pool through sympatric hybridization. This makes it an ideal species to study, since its generalist nature may reflect ancestral developmental programs.

* 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

Summary: Reviewed applications of ATAC-seq in developmental biology, showing how dynamic chromatin accessibility orchestrates stage-specific transcriptional programs. They discussed examples in vertebrates where ATAC-seq revealed regulatory landscapes controlling early morphogenesis. The article demonstrates the use of ATAC-seq to uncover developmental regulation in embryos.

* 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

Summary: Introduced ATAC-seq as a method to assay open chromatin. They showed that it required fewer cells than DNase-seq and provided high-resolution maps of promoters and enhancers. They also demonstrated that nucleosome positioning could be inferred from fragment sizes.

* 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

Summary: Applied ATAC-seq to Nile tilapia embryos and identified regulatory elements associated with adaptive traits such as pigmentation and morphological development. They demonstrated that accessible chromatin regions overlap with enhancers controlling important adaptive phenotypes. This study validates the relevance of ATAC-seq in cichlid research and shows that regulatory variation has real functional and ecological consequences.

* 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

Summary: Analysed chromatin dynamics in teleost embryos, focusing on transcription factors such as *sox10b* that regulate morphogenesis. They found that chromatin accessibility patterns were tightly linked to the timing of developmental gene activation. Their results provide a model for how regulatory elements guide stage-specific gene expression.

* Li et al. (2019)

Summary: Conducted ATAC-seq in zebrafish and demonstrated stage-specific changes in chromatin accessibility during early embryonic development. Their work showed how accessibility landscapes correlate with major developmental transitions, validating the approach for fish models.

* 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

Summary: Studied enhancer signatures in vertebrate embryos and showed that enhancers transition from pluripotency-associated marks to lineage-specific activity as cells differentiate. (Their findings are similar with my observation that early stages in cichlids emphasize conserved noncoding elements, while later stages shift to promoters and tissue-specific processes.)

* 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

Summary: Reviewed how genetic and regulatory mechanisms contribute to morphological traits in cichlids, including jaw shape and coloration. They stressed the importance of regulatory evolution for adaptive diversity.

* 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

Summary: Demonstrated that the DNA methyltransferase *dnmt1* controls jaw morphology in cichlids, directly linking epigenetic regulation to adaptive traits. (use in discussion)

* 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

Summary: showed that *otx2* is regulated by chromatin state in zebrafish, controlling eye and neural development. (use in discussion)

* 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

Summary: Analysed chromatin accessibility in zebrafish optic development and found that homeobox genes such as rx3 are regulated by accessibility changes. My identification of pbx2 in eye morphogenesis GO terms mirrors their findings, suggesting that similar mechanisms govern eye development in cichlids.

* 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

Summary: Studied *klf6a* in zebrafish and showed it controls endothelial cell rearrangement during cardiovascular pruning. (use in discussion)

* 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

Summary: Demonstrated that *hdac6* regulates tubulin deacetylation, cilia formation, and motility. (use in discussion)

* 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

Summary: Found that *sox11a* is essential for eye morphogenesis in zebrafish. (use in discussion)

* 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
* Summary: Reviewed craniofacial development and highlighted regulatory roles for *otx2* and *pbx2*. (use in discussion)

**Methods and errors encountered**

The project can be categorized to the following steps for easier understanding:

1. Pre-project work
2. Peak annotation
3. Enrichment analysis
4. Differential accessibility

**Pre-project work**: Some works that have already been carried out to initiate the project. The ATAC-seq data was generated. These reads have been filtered to remove duplicates and unwanted parts. The peaks were identified using tools like MACS2 and TF sites were characterised.

Note: All raw data files needed for the project is stored in HPC in the location ‘home/hc-storage/ATAC\_Acalliptera’

**Peak annotation**: Includes visualizing the peak annotations using ggplot to map where chromatin peaks are located across the genome. Mainly used R studio to plot bar plot and pie chart to visualize the peaks.

Files used: 1b-i.peakannot.counts2, 1b-ii.tss\_stats.out2.

Errors: Bar plot couldn’t be generated along with supplementary table. Rectified using the function ‘tableGrob’ in R by typing in the reference table separately.

**GAT:** Genomic Association Tester was used to find the enrichment of peaks after annotating them. The details on how to use GAT was studied from website (refer databases page for link). Results were obtained and file was sent to supervisor confirmation. Dotplots were prepared in R studio using the resultant tsv files.

Files used for dotplot: gatnormed\_Ac\_3dpf\_Peak\_AllAnnot\_allIDR\_ChrBgd.tsv, gatnormed\_Ac\_7dpf\_Peak\_AllAnnot\_allIDR\_ChrBgd.tsv, gatnormed\_Ac\_12dpf\_Peak\_AllAnnot\_allIDR\_ChrBgd.tsv

Errors: GAT was not installed on HPC. Manual installation couldn’t be carried out because of restricted access. Rectified by downloading and running the GAT own personal laptop.

**Enrichment analysis:** To know the function of enriched genes obtained from GAT, G:profiler was used. The IDs were extracted and uploaded into the webpage to obtain csv files containing adjusted p values and fold enrichment. The csv files were used as inputs for ggplot. Biological Process, Molecular Function, and Cellular Component plots were made. Mainly focused on the biological plots for further steps. The gene names that may relate to epigenetic modification was identified.

Files used:

For extracting Ensembl gene IDs: Ac\_3dpf\_ATAC\_peaks.final.allIDR.narrowPeak.features.gff, Ac\_7dpf\_ATAC\_peaks.final.allIDR.narrowPeak.features.gff, Ac\_12dpf\_ATAC\_peaks.final.allIDR.narrowPeak.features.gff

Resultant .csv files used for further plotting and analysis: gProfiler\_acalliptera\_3dpf.csv, gProfiler\_acalliptera\_7dpf.csv, gProfiler\_acalliptera\_12dpf.csv

Errors: log10 scale was used for FDR during plotting. However, the differences between components could not be identified as logscale was highly similar for all components. Hence it was plotted taking adjusted p-value directly as FDR.

**Differential accessibility:** The analysis till now only showed which genes were enriched at specific stages. Inorder to find how the accessibility of the genes changes over the stages, differential accessibility must be carried out. EdgeR package was installed in R studio prior to this step. The codes and protocol were based on the GitHub repository instructions. BAM files (two replicates per stage) and BED file for mapping was used.

Files used:

6 BAM files (2 per stage) located at – home/hsmidhuk/hc-storage/ATAC\_Acalliptera/3.TFfprint\_SignalTrack/

BED file located at – /home/hsmidhuk/hc-storage/ATAC\_Acalliptera/Acallipteragenome/current\_gtf/astatotilapia\_calliptera/Astatotilapia\_calliptera.fAstCal1.2.101.5kb\_promoters.bed

Errors:

1. EdgeR couldn’t be installed in HPC due to restrictions. Files were downloaded from HPC to own laptop to run the analysis.
2. At first 3 merged BAM files (each merged from two replicates per stage) were used. However, since the glmQLFit function required a minimum of 2 replicates per stage, this approach was discarded. All BAM files (2 replicates per stage) was used as input.
3. The ‘findoverlap’ function couldn’t identify significant peaks. No overlaps were found between peaks and all peaks showed same length during results. Rectified by filtering for significance before overlaps.

**Databases and links**

1. Genomic Association Tester: <https://gat.readthedocs.io/en/latest/installation.html>
2. G:profiler: [g:Profiler – a web server for functional enrichment analysis and conversions of gene lists](https://biit.cs.ut.ee/gprofiler/gost)
3. EdgeR:<https://ivanek.github.io/analysisOfGenomicsDataWithR/12_ATACSeq_html.html#differential-accessibility-analysis>
4. Ensembl Genome Browser (for gene IDs & annotations): <https://www.ensembl.org>

**Codes used**

All codes have been uploaded to GitHub repository. The link to repository is:

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

**Results and statistical analysis**

**Peak Annotation Analysis**

**Objective:** To annotate ATAT-seq peaks across the 3 developmental stages in A. calliptera to determine their genomic distribution and proximity to Transcription start sites (TSS).

**Results:**

The file ‘*1b-i.peakannot.counts2*’ was used to create a pie char that shows how the ATAC-seq peaks are distributed across different parts of the genome.

**A pie chart with different colored circles

AI-generated content may be incorrect.**

Figure 1: Distribution of narrow open chromatin peaks

The pie chart depicts the overall distribution of ATAC-seq peaks across genomic features, pooled across stages. A substantial proportion of peaks are in introns, exons and promoter regions. The high amount of peak in the promoter regions shows that promoter activity plays a central role in early embryonic development.

A bar plot was made to show how many peaks are located near TSSs compared to those that are further away using the file ‘1b-ii.tss\_stats.out2’.

A graph of a number of different types of peaks

AI-generated content may be incorrect.

Figure 2: TSS distribution of accessible peaks

The bar plot shows the distribution of accessible peaks relative to transcription start sites (TSS). Across all three developmental stages, a significant number of peaks are located within ±100 bp of TSS, indicating strong promoter activity.

**Genomic Association Tester (GAT)** was used to test for statistical enrichment ATAC-seq peaks across various genomic features annotated in the ‘.narrow peak’ files. The GAT output provided fold enrichment values and FDR-corrected p-values for each feature category.

The .tsv files were used to plot peak enrichment at the 3rd, 7th and 12th dpf.

A graph with blue dots

AI-generated content may be incorrect.

(a)

A graph with blue dots

AI-generated content may be incorrect.

(b)

A graph with blue dots

AI-generated content may be incorrect.

(c)

Figure 3: Peak enrichment at (a) 3dpf, (b) 7dpf and (c) 12dpf in *A. calliptera*

* Peak enrichment at 3dpf shows highest enrichment at hCNE(highly conserved non-coding element), five prime utr and aCNE. Highest peaks were found in introns and intergenic regions.
* The 7dpf showed the highest enrichment at five prime utr followed by hCNE and exon. Compared to the 3dpf, the enrichment and number of peaks associated with promoter was higher in 7dpf.
* 12dpf showed an overall lower amount of peak enrichment than 7dpf but more than that of 3dpf. The promoter gene is still enriched higher than that of 3dpf. 5’utr and hCNE showed the highest fold enrichment.

**Functional Enrichment**

**Objective:** To identify biological processes, molecular functions, and cellular components associated with accessible chromatin regions by performing Gene Ontology (GO) enrichment analysis on genes linked to ATAC-seq peaks.

**Results**

G:profiler was used for enrichment analysis. The ENSAC IDs of each stage was retrieved from respective .gff files. The IDs were uploaded into g:profiler. The .csv files containing data on enriched genes were used for further analysis.

A graph with red and yellow dots

AI-generated content may be incorrect.

\A graph with red dots

AI-generated content may be incorrect.

A graph with red and yellow dots

AI-generated content may be incorrect.

Note: The logscale shows no differences between each process, switch to normal p-value for better results

A graph showing a number of red dots

AI-generated content may be incorrect.

(a)

A graph with many colored dots

AI-generated content may be incorrect.

(b)

A graph with red and green dots

AI-generated content may be incorrect.

(c)

Figure 4: Enrichment of accessible genes at 3dpf of *A. calliptera* (a) Biological process enrichment, (b) Cellular component enrichment, (c) Molecular function enrichment

A graph with red dots

AI-generated content may be incorrect.

(a)

A graph with red dots

AI-generated content may be incorrect.

(b)

A graph with red and green dots

AI-generated content may be incorrect.

(c)

Figure 5: Enrichment of accessible genes at 7dpf of *A. calliptera* (a) Biological process enrichment, (b) Cellular component enrichment, (c) Molecular function enrichment

A graph with green dots

AI-generated content may be incorrect.

(a)

A graph with green and yellow dots

AI-generated content may be incorrect.

(b)

A graph showing a number of dots

AI-generated content may be incorrect.

(c)

Figure 6: Enrichment of accessible genes at 12dpf of *A. calliptera* (a) Biological process enrichment, (b) Cellular component enrichment, (c) Molecular function enrichment

A blue and black circle with white text

AI-generated content may be incorrect.

Figure 7: Overlap of GO gene terms across the three developmental stages

The major points that can be inferred from the enrichment plots and venn diagram are included below:

* The GO enrichment results revealed that each developmental stage shows different patterns of active biological processes, suggesting that chromatin accessibility and gene regulation change over time in a stage-specific manner.
* At 3 dpf, enriched biological processes were mostly related to early development, such as embryonic morphogenesis, axis formation, and cell differentiation.
* At 7 dpf, several GO terms were related to growth, tissue formation, organ development, and cytoskeletal organization, indicating that this stage involves rapid morphogenesis and structural complexity. This stage showed the highest number of enriched GO terms, suggesting increased regulatory activity.
* At 12 dpf, enriched terms included locomotion, neurodevelopment, and digestive system development, highlighting preparation for more independent behavior and organ function.
* The Venn diagram shows that some GO terms are shared between stages, while others are stage-specific. This highlights both core developmental processes and stage-specific regulatory changes that may shape morphological diversity.
* Several GO terms (e.g., cytoskeleton organization, morphogenesis, locomotory behavior) are directly linked to morphological changes, which play an important role in the adaptive radiation of East African cichlids.
* 7dpf has a major impact on the development of *A. calliptera*.
* The regulation of genes is higher in 7 & 12 dpf.

A diagram of a blue circle

AI-generated content may be incorrect.

Figure 8: Overlap of GO:BP terms across three stages

A blue and white circle diagram

AI-generated content may be incorrect. A blue and white circle with black text

AI-generated content may be incorrect.

A blue and white circle with black text

AI-generated content may be incorrect.

Figure 9: Overlap of BP go terms between each stage

**Differential chromatin accessibility analysis**

Pairwise comparisons of chromatin accessibility revealed major differences between early and later stages. The comparison of 3 vs 7 dpf identified ~78,675 significantly differentially accessible peaks, with 23,691 overlapping promoters and corresponding to ~8,687 unique genes. Similarly, the 3 vs 12 dpf comparison produced ~76,959 significant peaks, with 22,468 promoter-overlapping peaks linked to ~9,315 unique genes. In contrast, the 7 vs 12 dpf comparison identified only ~10,341 significant peaks, with 2,696 overlapping promoters and associated with ~2,208 unique genes.

These results suggest that most regulatory remodeling occurs between 3 dpf and later stages, while changes between 7 and 12 dpf are relatively modest, reflecting stabilization of the chromatin landscape.

A red and blue dotted diagram

AI-generated content may be incorrect. A red and blue dotted diagram

AI-generated content may be incorrect.

(a) (b)

A diagram of a red and blue triangle

AI-generated content may be incorrect.

(c)

Figure 10: MA plot representing differential peals

A diagram of a graph

AI-generated content may be incorrect. A diagram of a graph

AI-generated content may be incorrect.

(a) (b)

A diagram of a graph

AI-generated content may be incorrect.

(c)

Figure 11: Volcano plot of differential peaks

Volcano and MA plots confirmed that the most pronounced changes occurred in early transitions, with more balanced accessibility at later stages. 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 graph with different colored dots

AI-generated content may be incorrect.

(a)

A graph with different colored dots

AI-generated content may be incorrect.

(b)

A graph with different colored dots

AI-generated content may be incorrect.

(c)

A graph with many colored dots

AI-generated content may be incorrect.

(d)

Figure 12: Functional enrichment of GO:BP terms across stages

A screenshot of a graph

AI-generated content may be incorrect.

Figure 13: Faceted bar plot across stages

Note: Remove the empty facet parts when putting in final report

The faceted bar plot (Fig. 13) 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.

**Thoughts on the project**

This project was a unique experience, and I truly enjoyed it. Working with data is something I love, and this project was built around handling lots of it, which made it exciting. I’m very thankful to my supervisor, Dr. Tarang Mehta, for his support and advice whenever I got stuck, especially when I hit roadblocks like installing GAT or fixing a BED file error. Debugging code at every step was tough because I only started learning coding when I began my master’s course. It was challenging to troubleshoot issues like the zero overlaps or adjust scales for plots, but by the end, solving these problems felt oddly satisfying, and I grew to love those challenges. I learned so many new techniques, like ATAC-seq, GProfiler, and EdgeR, which improved my skills in bioinformatics. Exploring new topics, from peak annotation to differential accessibility, was fascinating and opened my eyes to how data can reveal biological stories, like cichlid development. The whole process, from start to finish, was amazing, even with setbacks like HPC issues or re-running analyses. Each step taught me something new, and I felt proud seeing my plots and results come together, like the volcano plots showing differences across 3, 7, and 12 dpf stages. This project made me more confident in coding and research, and I’m excited to keep learning in this field.