

SUMMER INTERNSHIP SUMMARY REPORT

Name: Dhruvi Sheth

Roll No.: 22294

Department: Biological Sciences

Institution: Indian Institute of Science Education and Research Bhopal

Professor: Dr. Nagarjun Vijay

Affiliation: Assistant Professor, Dept. of Biological Sciences, IISER Bhopal

Internship Topic: *Learning Bash Scripting - Intermediate*

Internship Duration: May to July 2025

GitHub Repository: <https://github.com/dssheth/bash-scripting-2025>

Overview

This internship focused on practical learning and application of **Bash scripting** in the context of **bioinformatics**, involving both independent learning and contribution to an ongoing research project.

Key Activities & Contributions

1. Learning Shell Scripting:

- a. Studied the book *Learning the Bash Shell (3rd Edition)* and made the notes.
- b. Created weekly videos explaining the content interactively with live terminal demonstrations.
- c. Prepared and submitted chapter-wise notes aligned with the video explanations.

2. Daily Practice & Documentation

- a. Maintained a GitHub repository with folder for every week, notes, and bash projects containing:
 - Daily files containing Markdown summaries of topics learned.
 - Corresponding `.sh` files with example scripts and command usage.
 - Linked videos, notes, and mini bash projects made.

3. Research Papers

1. *"Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss?"* - read and made a review report of it.
 - Understood the concept of orthologs and reciprocal best hits (RBHs).
 - Replicated the method with the BLAST tool on genomes of Gallus gallus (chicken) and Homo sapiens (human) and got the RBHs along the orthologs. While other tools are yet to be used and studied.
2. *"A fully-automated method discovers loss of mouse-lethal and human-monogenic disease genes in 58 mammals"*
3. *"Evidence for the loss of plasminogen receptor KT gene in chicken"*
 - Read both papers, to understand the pipelines used and compared to study **gene loss** mechanisms.
 - For gene **GPRC6A** in bovines, the results differ using the pipelines mentioned.
 - Currently, using bioinformatics tools and also developing tools to find the gene loss or presence.
4. *"The unexpected loss of the 'hunger hormone' ghrelin in true passerines: a game changer in migration physiology"* - to understand the pipeline for gene loss.

4. Mini Projects

- Designed and implemented bash scripting to create small projects useful in the real world along with practicing and understanding the commands and concepts.
 1. Created a random password generator
 2. An automatic files and folder creator specifically for web development.
 3. Modified the previous project with an interactive CLI app that prompts the user to create each file and folder.
- The github link for these projects:
https://github.com/dssheth/bash-scripting-2025/tree/main/Mini_bash_projects

5. Bioinformatics Tools (Lab Project):

- Gained hands-on experience with various tools:
 - **Databases:** NCBI, Ensembl
 - **Alignment Tools:**
 - BLAST suite (blastn, tblastn, tblastx)
 - EMBOSS (Needle, Matcher), Clustal Omega, Expasy Translate
 - And many other tools.
- These were used as part of a contributing to the ongoing lab project investigating the **loss of the GPRC6A gene in bovines**, under the guidance of the Professor and a PhD researcher.

6. Integrated Tool Development (Bash + Python + HTML):

- Developed a custom bioinformatics tool for:
 - Search the CDS sequences for start and stop codons in all six reading frames and highlight them (using **Bash**).
 - Translating sequences in all six frames (using **Python**).
 - Displaying results in an **HTML** interface with interlinked pages for navigation.
 - Supporting **multi-file input** by the bash script.
- The tool is functional and extendable, and currently under further development.
- The github link for the tool is : <https://github.com/dssheth/codon-highlighter>

7. Results of Tools and Further work

- Till now, the pre-mature stop codons are found in the CDS sequence of bovines with the tool made and the online alignment tools.
- The further work involves:
 - Checking the gene alignment with reference human gene and achieving complete exons of the gene in bovines.
 - Checking correct exon annotation, reading frame and exon phase in the bovine gene to conclude the presence/loss of the gene.