

INTRODUCTION

Background:

Elanco is a leader in bioinformatics with global investments in countries such as the US, UK, and China. They work on preventative treatments for animals, developing innovative therapeutics for our pets. To further advance this work, we created a pipeline to automate the mapping of cat genomes (V-Genes) to readily identify the possible genetic variations in the 300+ cats tested.

Motivation:

To avoid decreased drug efficacy and adverse events in animals, understanding conserved and unique regions in antibody genes is a key factor. By selecting genes that are highly conserved across most cat breeds, Elanco can infer the success rate of monoclonal antibody research to work towards universal treatments.

Overall Goal:

To help identify genetic variations among antibody variable genes in cats using both publicly available files and a reference genome. This helps to contribute to inclusivity, efficiency, and safety when developing new treatments for domestic cats.

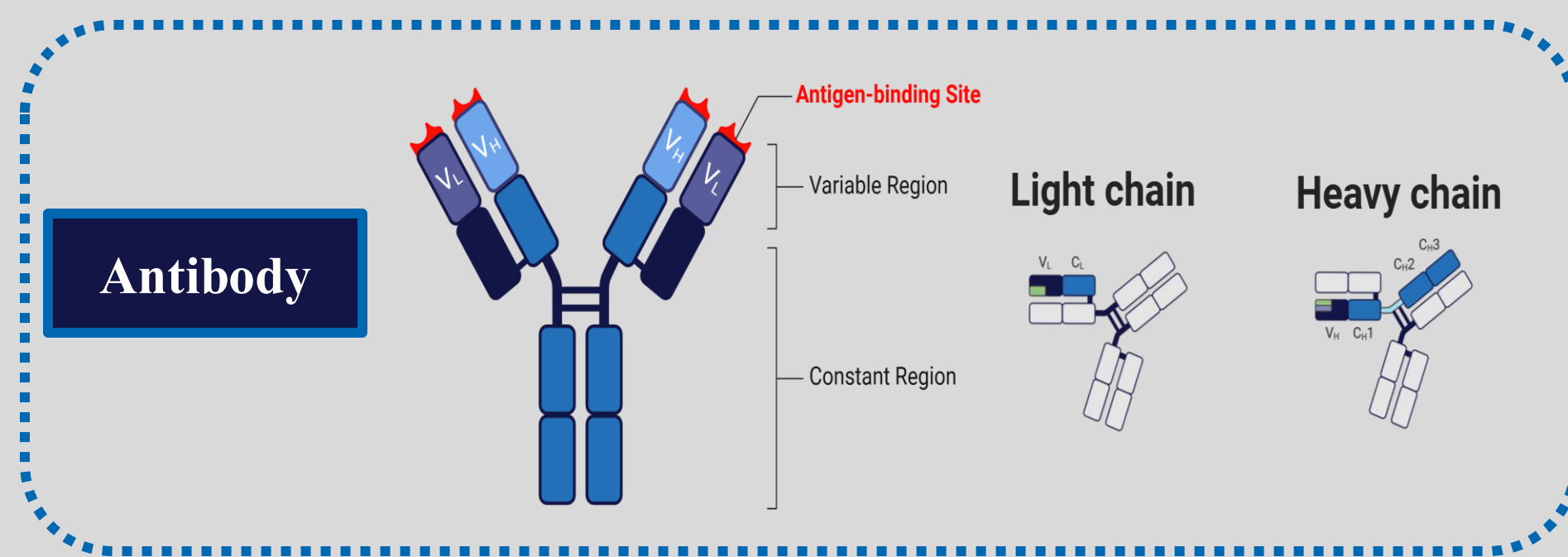


Figure 1. Visual representation of V-genes on antibodies. These regions of interest bind to any harmful microorganisms that enter the body.

METHODS

Steps:

- Using the 99Lives database, a list of cat genes were collected and their FASTQ files were downloaded.
- Files were aligned with a reference genome and indexed to create a document that allows the program to find the variants in the genome.
- The aligned sequence were converted into amino acids and went through Clustal Omega to visualize the sequence alignment of the V-Gene.

Packages Used:

- Biocontainers** - Manages bioinformatics packages to store data
- BWA** - Aligns paired ends of a gene to a reference genome to map similarities
- Seqkit** - Manipulates .fasta and .fastq Files
- Samtools** - Allows you to view and manipulate alignment files

RESULTS

Figure 2.

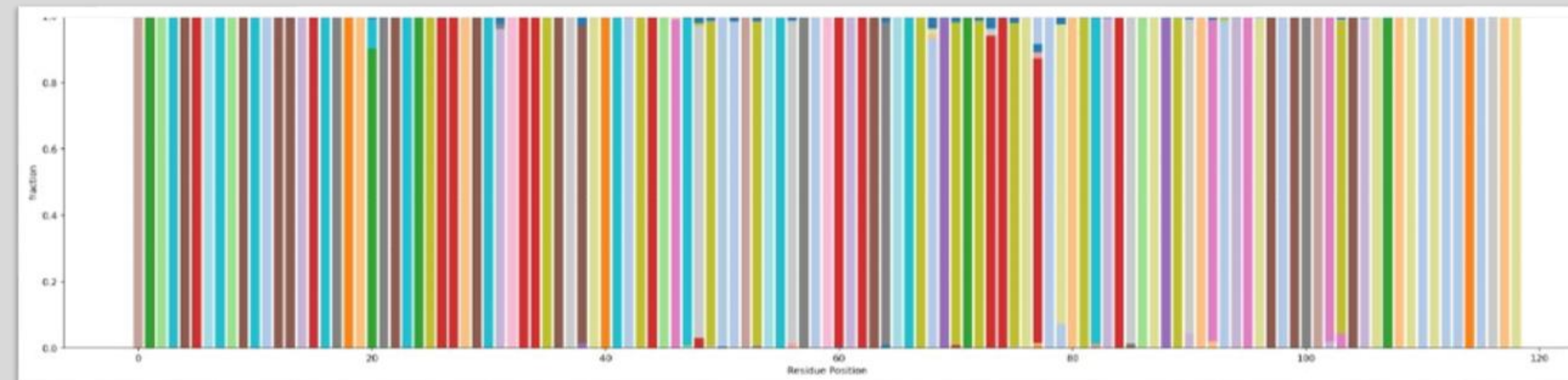
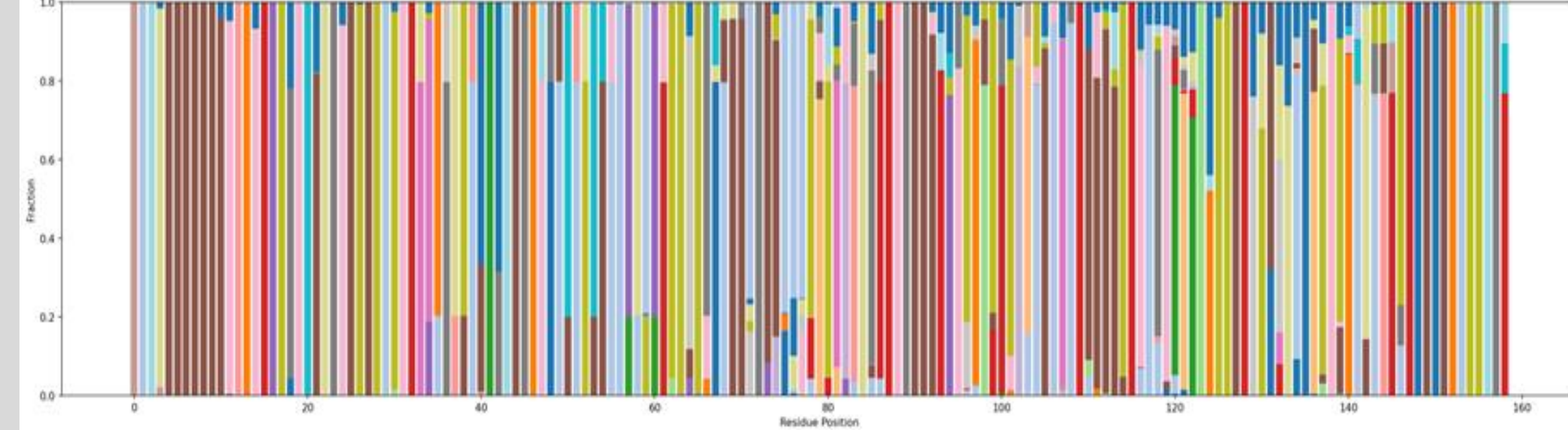
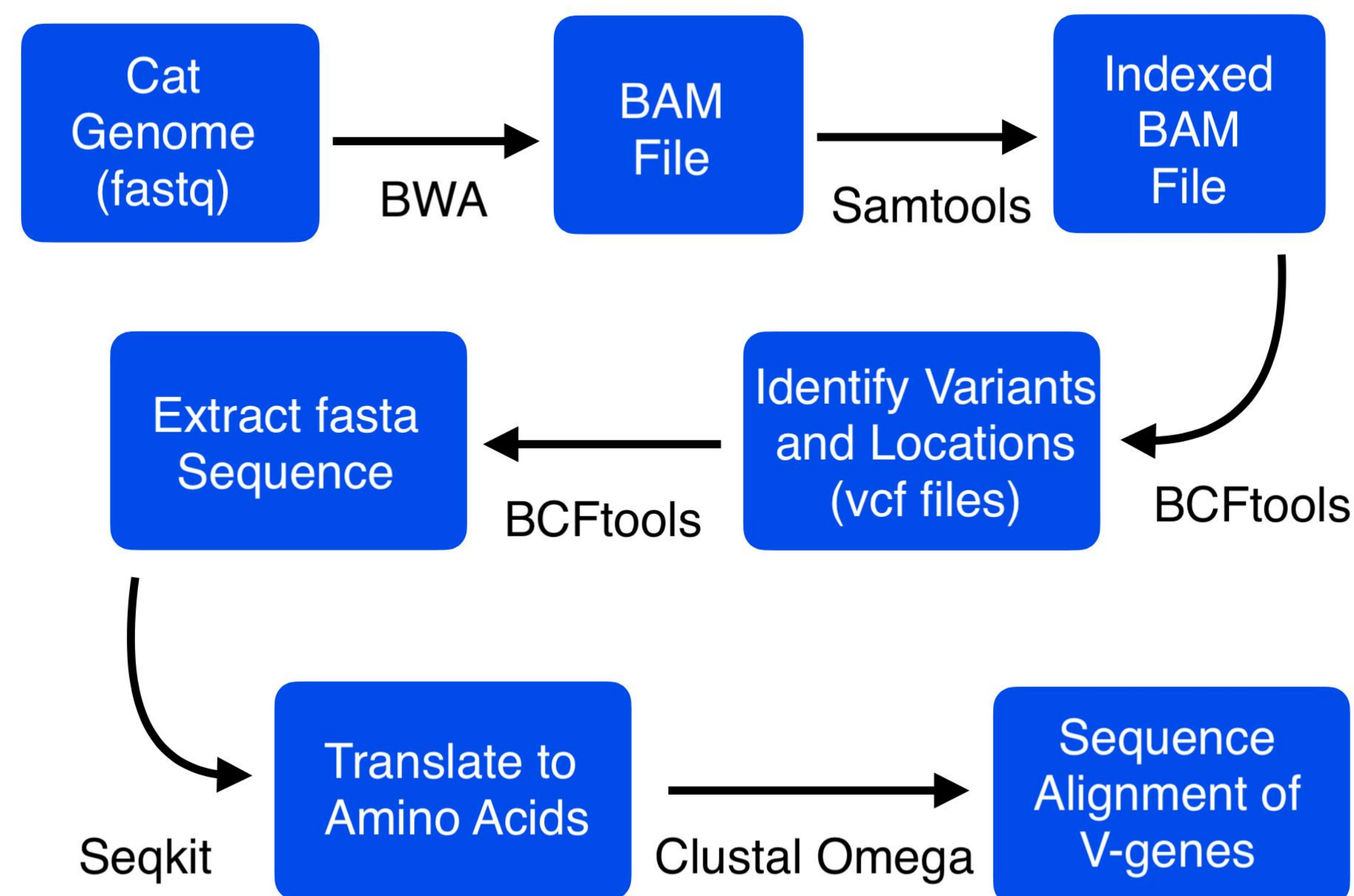


Figure 3.



Stacked Bar Plots of all the mapped genes of a particular genome, illustrating the relative frequency of amino acids at each residue position. Solid bars indicate highly conserved residues, while multicolored bars indicate variation of nucleotide sequences.



CONCLUSION + LEARNINGS

- Created a total of about 500+ files of aligned cat data.
- Developed a pipeline to auto-map diverse database files to the reference gene that can be applied to future projects.
- Gained experience with bioinformatics skills through genetic data wrangling and gene alignment.

IMPACT

- This is useful for future research focused on finding causes behind mutations and health risks.
- By tailoring to individual genes, this project allows for direct access to specific gene sequences found in domesticated cats.
- This provided information about V-genes for feline immunogenicity, human therapies, mutations, and diseases that are shared between cats and humans.
- The project created a data processing pipeline to use in further research across multiple scientific fields which can later be adapted to include more genes.

FUTURE GOALS

- Adapt the pipeline for use in other animals relevant to Elanco, such as cattle, swine, and poultry, to support broader health research.
- Continue improving the user interface and scalability to support larger datasets and sequencing projects.
- Sequence and run the pipeline for additional cat individuals and breeds.

REFERENCES

- "Antibody Structure" created in BioRender. (Scientific Image and Illustration Software | BioRender)
- Buckley, R. M., & Lyons, L. A. (2020). Precision/Genomic Medicine for Domestic Cats. *The Veterinary clinics of North America. Small animal practice*, 50(5), 983–990. <https://doi.org/10.1016/j.cvsm.2020.05.005>

ACKNOWLEDGEMENTS

- Teaching Assistant:** Chintu Mane
- Elanco Mentors:** Jingjing Ling, Ahmed Hassan, Varun Chauhan
- Corporate Partners Technical Specialist:** Cai Chen
- Senior Data Scientist:** Doug Crabill