Final Project

Reka Ivanyi

Introduction

Antimicrobial peptides, or AMPs are essential molecules whose complex evolution is not fully understood. They have been around as long as life itself: naturally occurring in the skin or gut epithelia of many vertebrates, invertebrates, plants, and bacteria. They are an integral part of organisms' innate immune systems, and often act as the main source of protection from pathogens in organisms that do not have an otherwise well-developed adaptive immune system (Bahar & Ren, 2013). In recent years, AMPs have been proven as effective therapies against bacterial and viral infections, and are currently being investigated as targeted therapies for cancer [Jafari et al. (2022), (Hilchie et al., 2011). Most AMPs are less than 100 amino acids in length and are cationic with a net positive charge between 2 and 9 (Chauhan et al., 2021). They have a hydrophobic region, which, along with their alpha-helical shape, allows them to easily penetrate lipid bilayers to cause cell lysis (Tornesello et al., 2020).

Scientists are working to synthesize antimicrobial proteins in a way which minimizes their allergic potential and cytotoxic effects towards healthy cells (Jafari et al., 2022). Understanding the structure of these proteins is essential for modifying their function, and knowing their evolutionary origins helps us understand how small changes in amino acid sequences can lead to varying structures and functions (Tu et al., 2015). Studying the rapid evolution of AMPs could also allow us to prevent bacteria from adapting to be resistant to them, as has been the case with many traditional antibiotics (Lazzaro et al., 2020). Species-specific AMPs have also been shown to play a part in microbiome formation, since the presence of certain AMPs prevents specific bacteria from colonizing a host (Franzenburg et al., 2013).

Though the evolutionary origins of AMPs in vertebrates have not been studied as vigorously as their direct, biomedical applications, they are known to evolve rapidly and have quickly diversifying functions (Castellanos et al., 2023). Studies have been done on the evolution of AMPs in orders and families of vertebrates and invertebrates, but not as much on their diversification among entire phyla such as for all vertebrates (König & Bininda-Emonds, 2011). The quantities of AMPs expressed and their specific sequences have been tracked within genera, as has been done with Drosophila (Hanson et al., 2019).

My goal was to look deeper into the evolution of AMPs throughout a large dataset of 31 different vertebrate species. I identified a diverse array of antimicrobial peptides present or absent throughout the genomes of these organisms and traced the presence of each AMP through different lineages. I focused on just cathelicidins and beta-defensins for now, as they are the two most common types of antimicrobial peptides (Avila, 2017). I hope to apply this research to a larger dataset of vertebrates after refining the methods used to search for and quantify the presence of these types of AMPs within genomes.

Methods

To complete this initial part of our research, we obtained full-genome protein sequences from the Ensembl public database (Harrison et. al, 2024) for 31 vertebrate species from many of the major vertebrate classes as well as one invertebrate species to use as a comparison (**Table 1**). Each proteome, which is a summary of all of the proteins expressed by the entire genome of an organism, was uploaded to Premise, and an ortholog analysis was done on the genomes using Orthofinder (Emms & Kelly, 2019) to combine all homologous, or related, genes among all the species for which data was collected. After completing this analysis, we compiled a comprehensive list of defensins and cathelicidins commonly found in vertebrate species from the Antimicrobial Peptide Database (Wang et al., 2016) and turned them into a FASTA file, which is a format where amino acids are represented by single-letter codes and there are no spaces between the letters or protein names. This file format can be ready by bioinformatic software to search for matches between sequences. I compared the file of the AMP amino acid sequences to the proteomes using the Basic Local Alignment Search Tool, or BLAST, which searches for similarities between two amino acids or nucleotides with the e value, or significance threshold, of 0.00001 (*Blast*, n.d.). For each species, I recorded which AMPs were present in the proteome and how many times their sequences were found. I also plotted an evolutionary tree of all of the species used in the study from the ortholog analysis, and I noted the presence or absence of each AMP on this phylogenetic tree using the tidyverse and ggtree packages (R Core Team, 2024), (Guangchuang, 2022).

Results

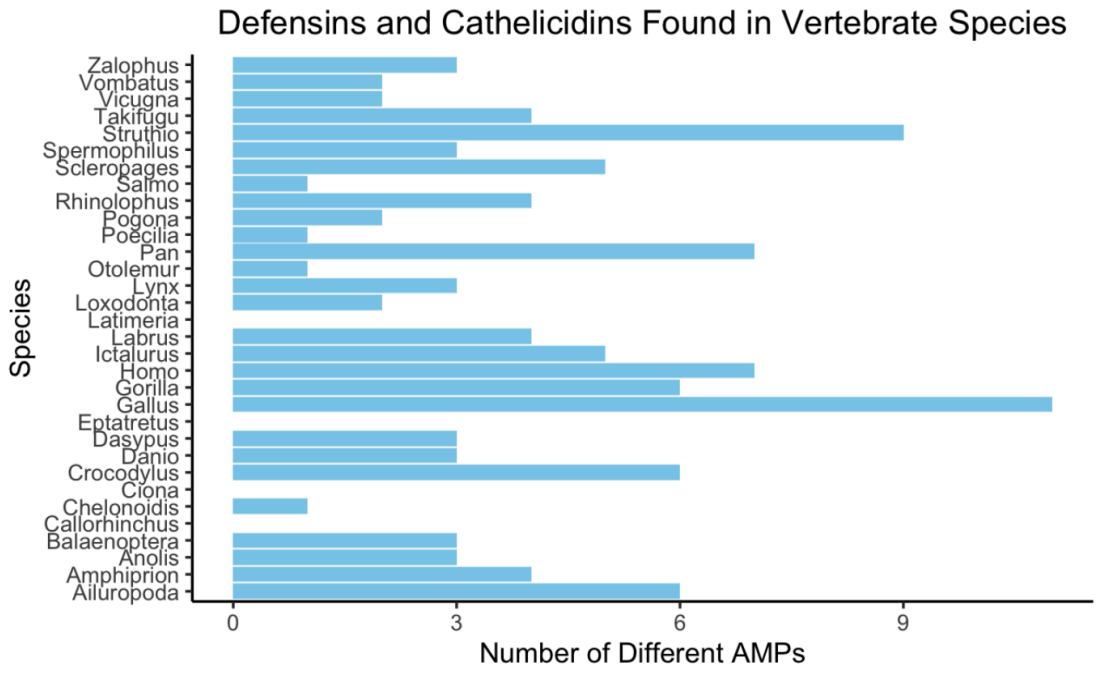

Figure 1: This graph shows the number of defensin and cathelicidin proteins found in the genomes of the list of 31 vertebrates (and one invertebrate) from various orders and classes. Species with no AMPs present are not listed.

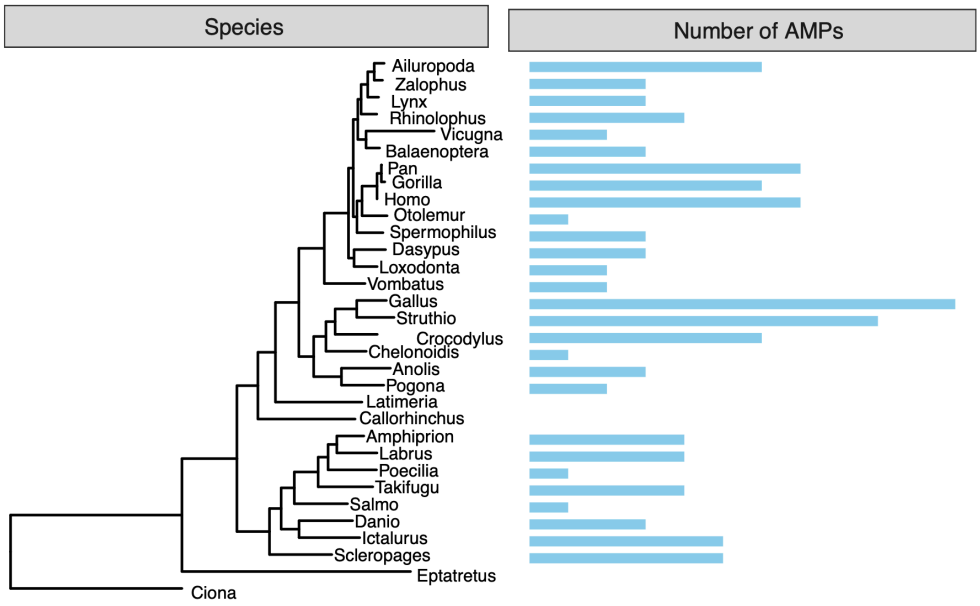

Figure 2: This figure shows how the amount of AMPs found in each species align with their evolutionary relatedness. The tree was generated using an Orthofinder analysis, and *Latimeria* and *Callorhinchus* were put in the wrong place. They should be closer to *Eptatretus*.

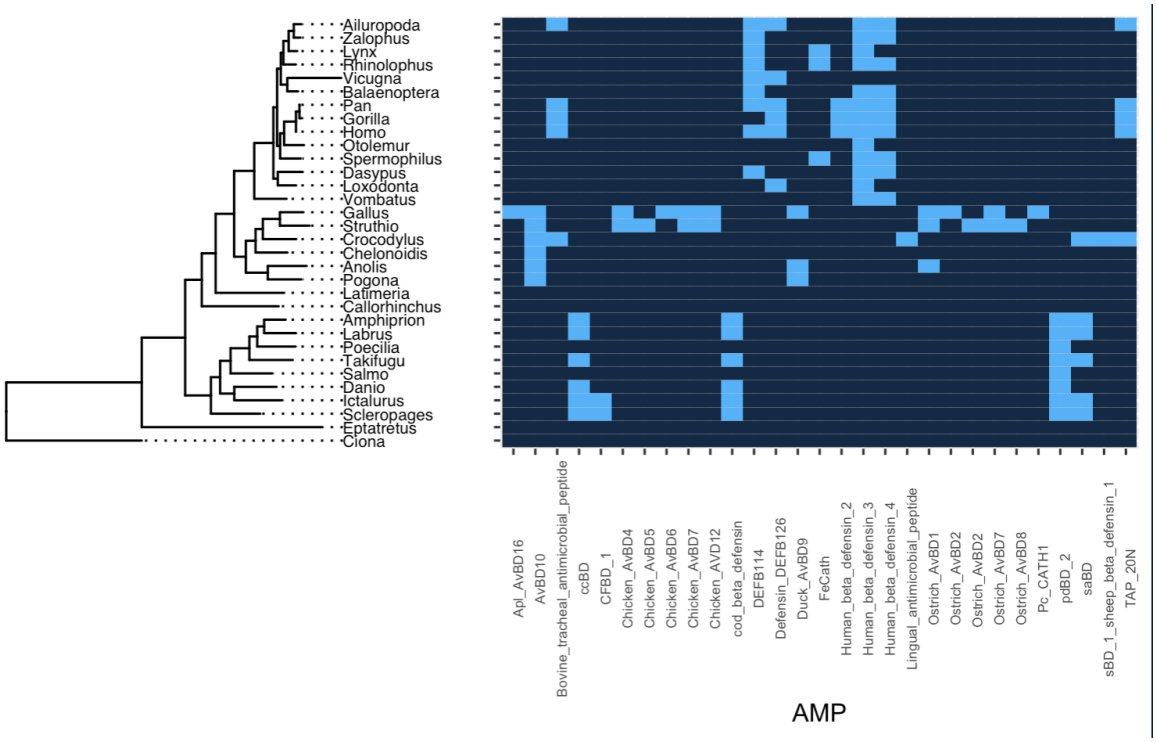

Figure 3: This figure shows the presence or absence of each AMP studied and which species it was found in. The order of the species in the matrix corresponds to their evolutionary relatedness from the tree. The light blue signifies presence and the dark blue signifies absence from the genome.

The chicken and ostrich have the highest number of different AMPs (11 and 9 respectively), followed by the chimpanzee and human (which both have 7). The coelacanth, ghost shark, hagfish, and sea squirt had none of the AMPs that were searched for. There was a significant difference between the numbers of total antimicrobial peptides between the classes of animals studied ($F7 = 5.472$, $p = 0.000751$). The significant differences were between the class Aves and each of the other classes: Chondrichthyes ($p = 0.0045948$), Mammalia ($p = 0.0037877$), Myxini ($p = 0.0045948$), Reptilia ($p = 0.0051045$), Ascidiacea ($p = 0.0045948$), Actinopterygii ($p = 0.0034841$), and Actinistia ($p = 0.0045948$). Human defensins 3 and 4 were found among most of the mammals, while the chicken and ostrich defensins were found in the chicken and ostrich genomes. AvBD10, an avian AMP, was not only found in the birds, but in most of the reptiles as well. This was also true of the duck-derived peptide AvBD9. Most of the fish derived AMPs were found exclusively within fish genomes. Bovine tracheal antimicrobial peptide and another very similar peptide, TAP 20N, were found in the genomes of the crocodile, panda, human, gorilla, and chimpanzee. Only two cathelicidins were found among all the species, with Fe Cath being found in the Canada lynx and Greater horseshoe bat, and pc Cath 1 being found in just the chicken.

Discussion

The goal of this study was to generate more information about the evolution of antimicrobial proteins in the phylum chordata and see whether there are lineage-specific AMPs present in vertebrate genomes. We found clear differences between the numbers and types of defensins and cathelicidins expressed in the different vertebrate lineages (**Fig 3**). Human beta defensin orthologs were found in most of the mammal species, but occurred less frequently in other vertebrates. The chicken and ostrich derived AMPs, on the other hand, were found mostly in the ostrich and chicken, the two species of bird examined in the study. This was expected, but it shows that the analysis was done correctly. The AMPs that were originally derived from fish species were found most commonly in the fish species; another expected conclusion, but a sign of successful analysis. Pdbs 2 was one of these peptides, and it is known to protect against bacterial infections caused by aquatic bacteria, suggesting that this peptide has become specialized to protect fish against the bacteria they encounter most frequently (Chen et al., 2013).

The chicken and ostrich genomes contained the most antimicrobial proteins, but that could be because a lot of the reference AMPs used were derived from chickens and ostriches. Humans and other mammals were found to have the second-highest variety of AMPs, and that could be because most of the other defensins and cathelicidins that weren't avian-derived were derived from mammals. Studies also suggest that mammals could have more types of AMPs because they encounter more microbes in their natural habitats (Tu et al., 2015). The more different types of microbes that an organism encounters, the more types of AMPs it is likely to develop. The clear differences in AMP types of various species and also the variation between AMP numbers suggest that AMPs do in fact evolve to fit the ecological niche and environmental conditions of species (Tu et al., 2015).

It has been studied that immune genes in various vertebrate species have evolved significantly over time, and that this diversification or loss in genes is driven by the need for specific biological functions. Genes for defensins and cathelicidins have been found to evolve by duplication events in avian and primate species, resulting in a diverse array of AMPs that can protect against a wide variety of bacteria, viruses, and fungi (Van Dijk et al., 2023). Genes similar to those that produce cathelicidins have been found in jawless fish as well, which suggests that avian and primate cathelicidins, and potentially other AMPs, may originate from a common ancestor shared between cyclostomes and gnathostomes (Van Dijk et al., 2023). Our results further confirm these findings.

In our study, the AMPs examined were not found in the Sea squirt, the Hagfish, the West Indian ocean coelacanth, or the Australian ghost shark. The sea squirt isn't a vertebrate, so it makes sense that the vertebrate-derived AMPs were not found in its genome. The hagfish is one of the earliest evolved vertebrates, so it seems like the most common vertebrate AMPs like cathelicidins and defensins evolved later than the Eptatretus genus. The coelacanth and ghost shark are likely different enough from the other fish species and not similar enough to the rest of the vertebrate classes to have the same AMPs as any of the other species.

There are many future directions in which we can take this study. We are planning on using a much larger dataset of vertebrate genomes to analyze for a comprehensive list of common antimicrobial proteins. We will create a list of AMPs that encompasses all the known cathelicidins and beta defensins, as well as other AMPs that are more commonly found in fish. We will also refine the methods used for the analysis of the BLAST data, as the process will have to be more streamlined when working with a lot more species. There is currently an ascertainment bias present in our study, as we can only find the AMPs that we are specifically looking for. If we search for more avian derived proteins for example, we will find more AMPs present in avian species. This could be skewing the numbers of different AMPs we found for each species, and the reduction of this bias will be considered in future studies.

References

Avila, E. E. (2017). Functions of Antimicrobial Peptides in Vertebrates. *Current Protein & Peptide Science*, *18*(11). https://doi.org/10.2174/1389203717666160813162629

Bahar, A., & Ren, D. (2013). Antimicrobial Peptides. *Pharmaceuticals*, *6*(12), 1543–1575. https://doi.org/10.3390/ph6121543

Blast: Basic local alignment search tool. (n.d.). https://blast.ncbi.nlm.nih.gov/Blast.cgi

Castellanos, F. X., Moreno-Santillán, D., Hughes, G. M., Paulat, N. S., Sipperly, N., Brown, A. M., Martin, K. R., Poterewicz, G. M., Lim, M. C. W., Russell, A. L., Moore, M. S., Johnson, M. G., Corthals, A. P., Ray, D. A., & Dávalos, L. M. (2023). The evolution of antimicrobial peptides in Chiroptera. *Frontiers in Immunology*, *14*, 1250229. https://doi.org/10.3389/fimmu.2023.1250229

Chauhan, S., Dhawan, D. K., Saini, A., & Preet, S. (2021). Antimicrobial peptides against colorectal cancer-a focused review. *Pharmacological Research*, *167*, 105529. https://doi.org/10.1016/j.phrs.2021.105529

Chen, Y., Zhao, H., Zhang, X., Luo, H., Xue, X., Li, Z., & Yao, B. (2013). Identification, expression and bioactivity of Paramisgurnus dabryanus -defensin that might be involved in immune defense against bacterial infection. *Fish & Shellfish Immunology*, *35*(2), 399–406. https://doi.org/10.1016/j.fsi.2013.04.049

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Franzenburg, S., Walter, J., Künzel, S., Wang, J., Baines, J. F., Bosch, T. C. G., & Fraune, S. (2013). Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proceedings of the National Academy of Sciences*, *110*(39). https://doi.org/10.1073/pnas.1304960110

Guangchuang, Y. (2022). Chapman; Hall/CRC. doi:10.1201/9781003279242

Hanson, M. A., Lemaitre, B., & Unckless, R. L. (2019). Dynamic Evolution of Antimicrobial Peptides Underscores Trade-Offs Between Immunity and Ecological Fitness. *Frontiers in Immunology*, *10*, 2620. https://doi.org/10.3389/fimmu.2019.02620

Hilchie, A. L., Doucette, C. D., Pinto, D. M., Patrzykat, A., Douglas, S., & Hoskin, D. W. (2011). Pleurocidin-family cationic antimicrobial peptides are cytolytic for breast carcinoma cells and prevent growth of tumor xenografts. *Breast Cancer Research*, *13*(5), R102. https://doi.org/10.1186/bcr3043

Jafari, A., Babajani, A., Sarrami Forooshani, R., Yazdani, M., & Rezaei-Tavirani, M. (2022). Clinical Applications and Anticancer Effects of Antimicrobial Peptides: From Bench to Bedside. *Frontiers in Oncology*, *12*, 819563. https://doi.org/10.3389/fonc.2022.819563

König, E., & Bininda-Emonds, O. R. P. (2011). Evidence for convergent evolution in the antimicrobial peptide system in anuran amphibians. *Peptides*, *32*(1), 20–25. https://doi.org/10.1016/j.peptides.2010.10.009

Lazzaro, B. P., Zasloff, M., & Rolff, J. (2020). Antimicrobial peptides: Application informed by evolution. *Science*, *368*(6490), eaau5480. https://doi.org/10.1126/science.aau5480

R Core Team. (2024). *R: A language and environment for statistical computing_. R foundation for statistical computing*. %3Chttps://www.R-project.org/%3E

Tornesello, A. L., Borrelli, A., Buonaguro, L., Buonaguro, F. M., & Tornesello, M. L. (2020). Antimicrobial Peptides as Anticancer Agents: Functional Properties and Biological Activities. *Molecules*, *25*(12), 2850. https://doi.org/10.3390/molecules25122850

Tu, J., Li, D., Li, Q., Zhang, L., Zhu, Q., Gaur, U., Fan, X., Xu, H., Yao, Y., Zhao, X., & Yang, M. (2015). Molecular Evolutionary Analysis of -Defensin Peptides in Vertebrates. *Evolutionary Bioinformatics*, *11*, EBO.S25580. https://doi.org/10.4137/EBO.S25580

Van Dijk, A., Guabiraba, R., Bailleul, G., Schouler, C., Haagsman, H. P., & Lalmanach, A.-C. (2023). Evolutionary diversification of defensins and cathelicidins in birds and primates. *Molecular Immunology*, *157*, 53–69. https://doi.org/10.1016/j.molimm.2023.03.011

Wang, G., Li, X., & Wang, Z. (2016). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research*, *44*(D1), D1087–D1093. https://doi.org/10.1093/nar/gkv1278