

# ENS210 Group #3 Project Report

Berra Sayın Fatih Taşyaran Rana Kalkan

*Instructor: Ogün Adebali*

December 15, 2019

## 1 Introduction

Characterizing a protein family in an evolutionary aspect could give hints about function, evolutionary history, critical residues and domains of this protein family. In order to observe a protein's functionality across animals which live in aquatic and terrestrial environments and also animals like tetrapods who live in intersection of this environments, we chose the protein family **Carbamoyl-phosphate synthase, large subunit (CPS)**.

CPS catalyses the synthesis of carbamoyl phosphate from biocarbonate, ATP and glutamine or ammonia and represents the first committed step in pyrimidine and arginine biosynthesis in prokaryotes and eukaryotes, and in the urea cycle in most terrestrial vertebrates. Most prokaryotes carry one form of CPS that participates in both arginine and pyrimidine biosynthesis, however certain bacteria can have separate forms. Eukaryotes have two different forms of CPS: a mitochondrial enzyme (CPSI) that participates in both arginine biosynthesis and the urea cycle; and a cytosolic enzyme (CPSII) involved in pyrimidine biosynthesis. There is a third form of the enzyme, CPSIII, found in fish, which uses glutamine as a nitrogen source instead of ammonia.

In general, animals who live in different environments have different excretory systems. However, animals who live in the same environment and who are different kind of species can also have different excretory system. By examining the protein family which is one of the families that differ within different excretory systems, we are aiming to find evolutionary history of this protein family and its effect on the phenotypes of different species.

Ammonia is poisonous to all vertebrate, although it is not dangerous in low density and a good solvent in water. Therefore, it is possible to excrete ammonia for animals who live in abundance of water called aquatic animals. In contrary, there is no such feasibility for terrestrial animals so that they have to convert ammonia into urea or uric acid in their livers by reacting ammonia with  $CO_2$  via the urea cycle and excrete urea with consuming more energy due to this process.

As mentioned before, species and their excretory systems are intervening in terms of environments. Whales who are mammals excrete urea in contrary to marine fishes. But some fishes which live in very alkaline environments excrete urea in contrary to marine fishes. Also there is the case of animals like amphibians, which live in both of these two environments, demonstrates three of this nitrogen waste forms. Because of this variety, we want to investigate the evolutionary history of the protein family of **CPS** that has a significant role in this process.

## 2 Method

To be able to investigate this evolutionary history, we acquired sequences of **CPSI**, **CPSII** and **CPSIII** via biological databases such as BLAST, PFAM, InterPro and GenBank for different species that have characteristics we mentioned above. To understand the homology between many species, we also looked at OrthoDB where we can find numerous orthologous sequences. Then, we aligned those sequences via MEGA and R using *ClustalW* algorithm. In order to specifically observe mutated residues and conserved residues we permuted all 3 variants in a pairwise manner which are:

- CPSI(1)
- CPSII(2)
- CPSIII(3)
- CPSI - CPSII(4)
- CPSI - CPSIII(5)
- CPSII - CPSIII(6)
- CPSI - CPSII- CPSIII(7)
- Consensuses of CPSI, CPSII & CPSIII(8)

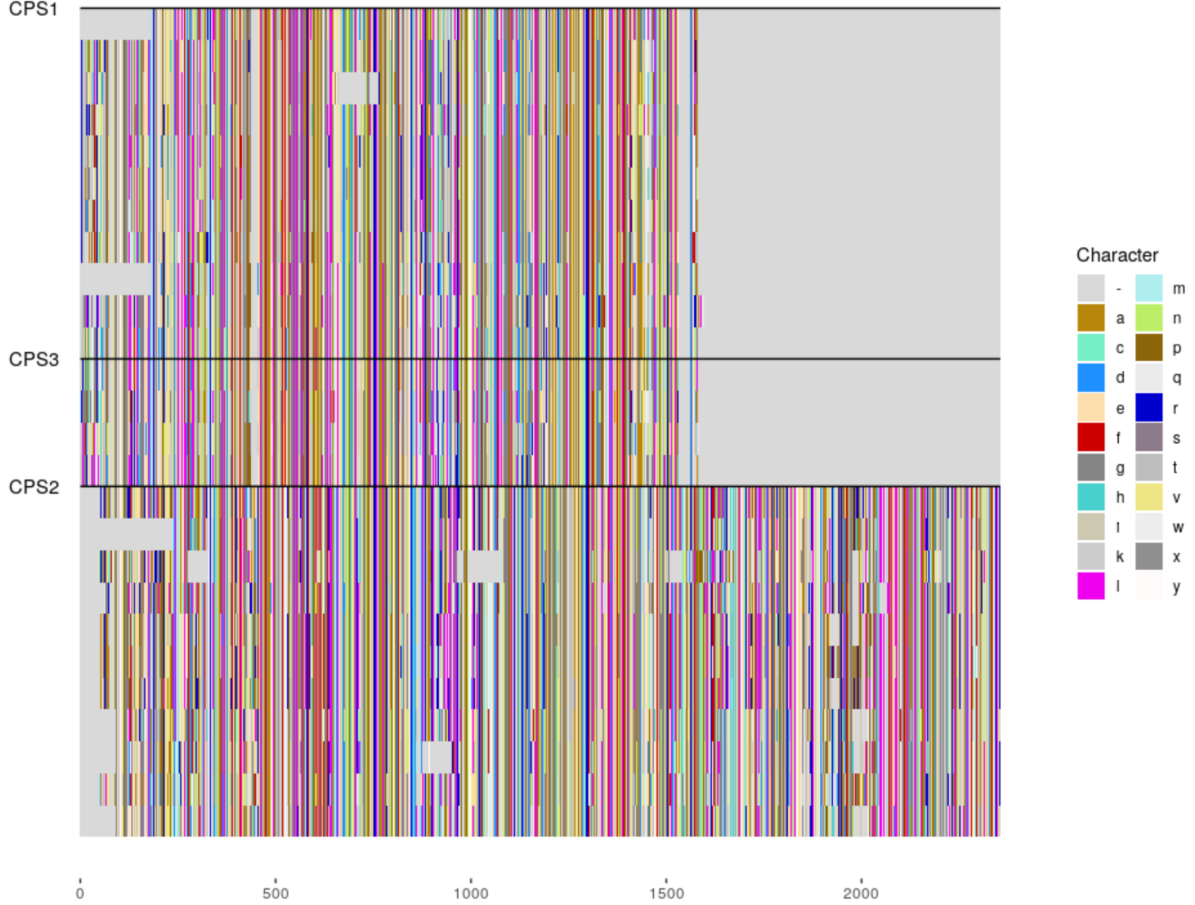


Figure 1: Visual Interpretation of Multiple Sequence Alignments

We aligned mentioned 8 combinations of sequences and we looked at their *Heatmap of Distances*, *Visual Interpretation*, *Conserved Residue positions* and *it's overlapping over domains* and *Pylogenetic Tree*. To find the critical positions, we also grouped the sources of this as sequences as *Amniotes*, *Amphibians* and *Aquatics*.

First we aligned protein sequences of a small samples (11 species from CPSI, 11 species from CPSII and 4 species from CPSIII) from distinct taxonomic groups in Jawed vertebrates which are Birds, Alligators and others, Turtles, Lizards, Mammals, Amphibians, Coelacanth, Bony Fishes, Cartilaginous fishes, Fig 1. We observed that CPSII has longer sequence than CPSI and CPSIII which are almost has same length of sequence. Then we looked at their conserved residues Fig 2.

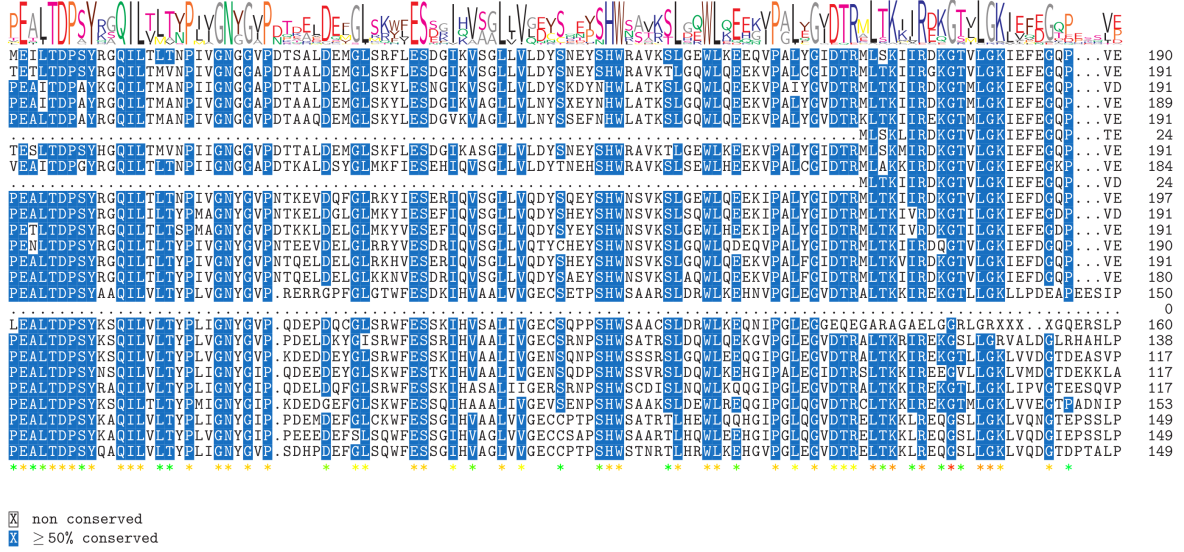


Figure 2: Conserved Residues Among CPSI-II-III

After calculating conserved residues and visual interpretations, we draw a heatmap of distance between aligned sequences for the species in the small subset to specifically see which of them closer to each other. In this interpretation we found out *Human*, *Platypus* and *Koala* are closer to each other than other animals. Fig 3 We also observed this situation in phylogenetic tree we draw and also we observed relatedness among Rooster and Crocodile. Which increased our confidence level in our work since these two species are also closer in the evolutionary tree. Fig 4

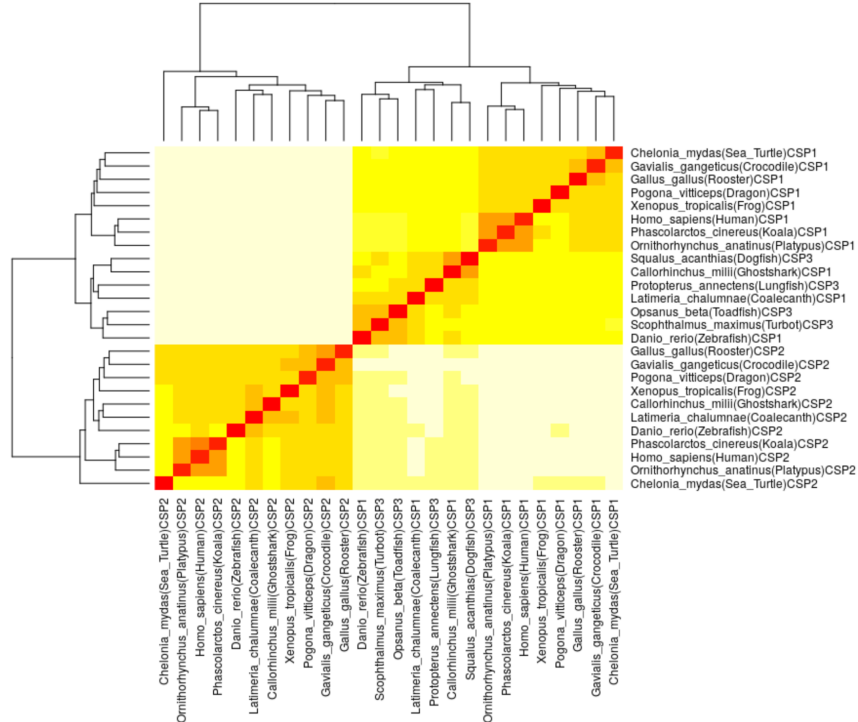


Figure 3: Heat map of Distances of Aligned Sequences Belong to Different Species

### Phylogenetic Tree of CSP Sequences

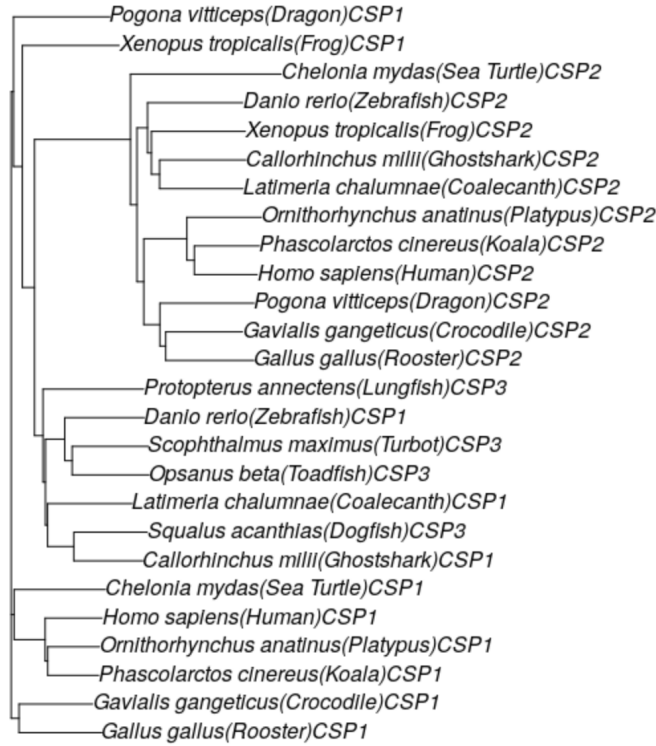


Figure 4: Phylogenetic Tree of Small Dataset

Then we expanded our samples and collected 10 protein sequences from each group of Jawed vertebrates as we mentioned. Then we had 110 sequences from CPSI and CPSII and 4 sequences from CPSIII since we found a few samples in databanks. After that we aligned them as a group of CPSI-CPSII, CPSI-CPSIII, CPSII-CPSIII and CPSI-CPSII-CPSIII. Then we calculated conservation scores for each alignments and plotted their graphs Fig 5. We also identified positions of protein domains and then added those regions belonging to distinct domains into our plots in different colors. We observed that CPSase-L-D2 domain which is represented in green region is less conserved compared to others. This domain is Carbamoyl-phosphate synthase L chain, ATP binding domain and it has an ATP-grasp fold.

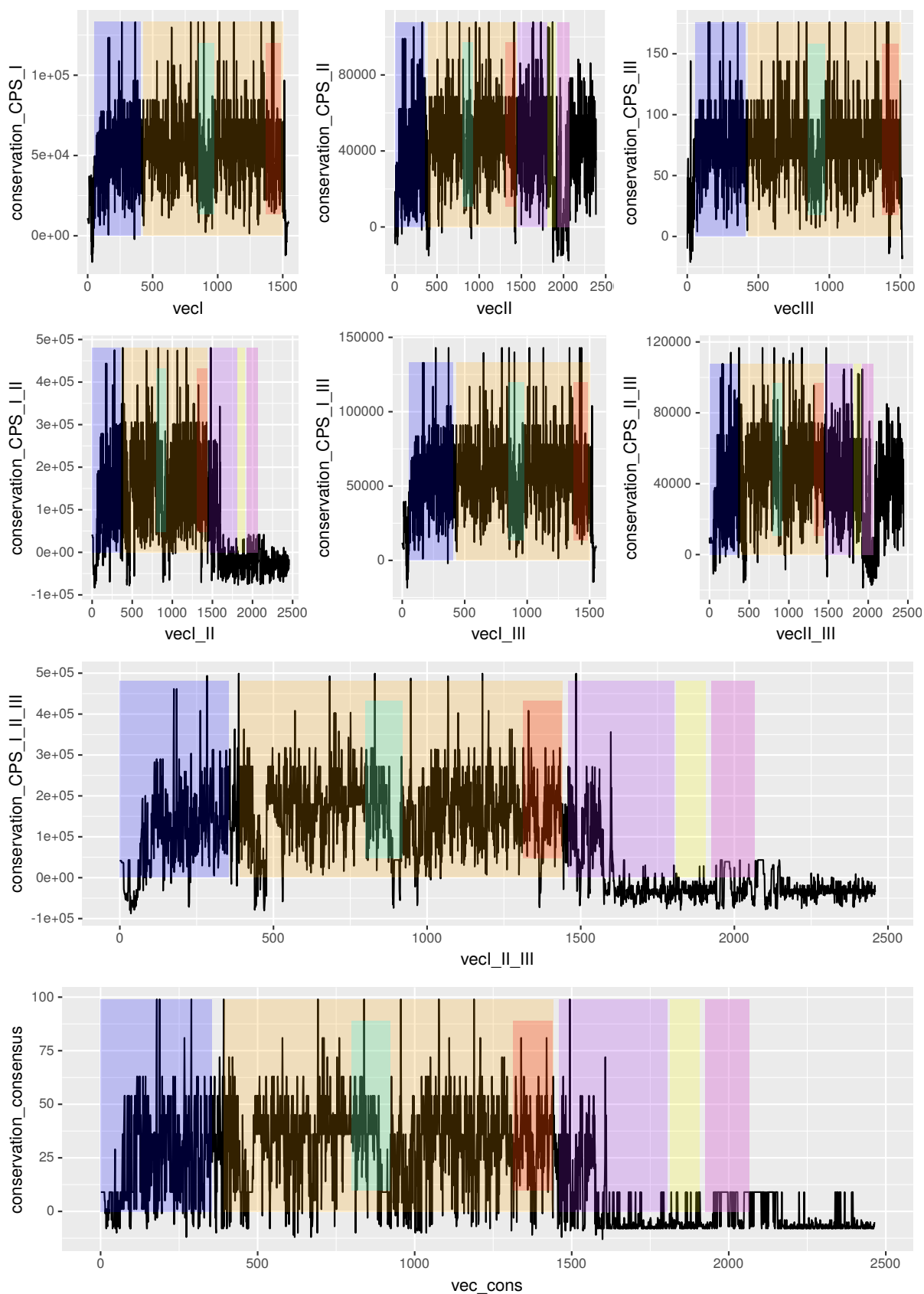


Figure 5: Conserved Residues Among **CPSI-II-III**

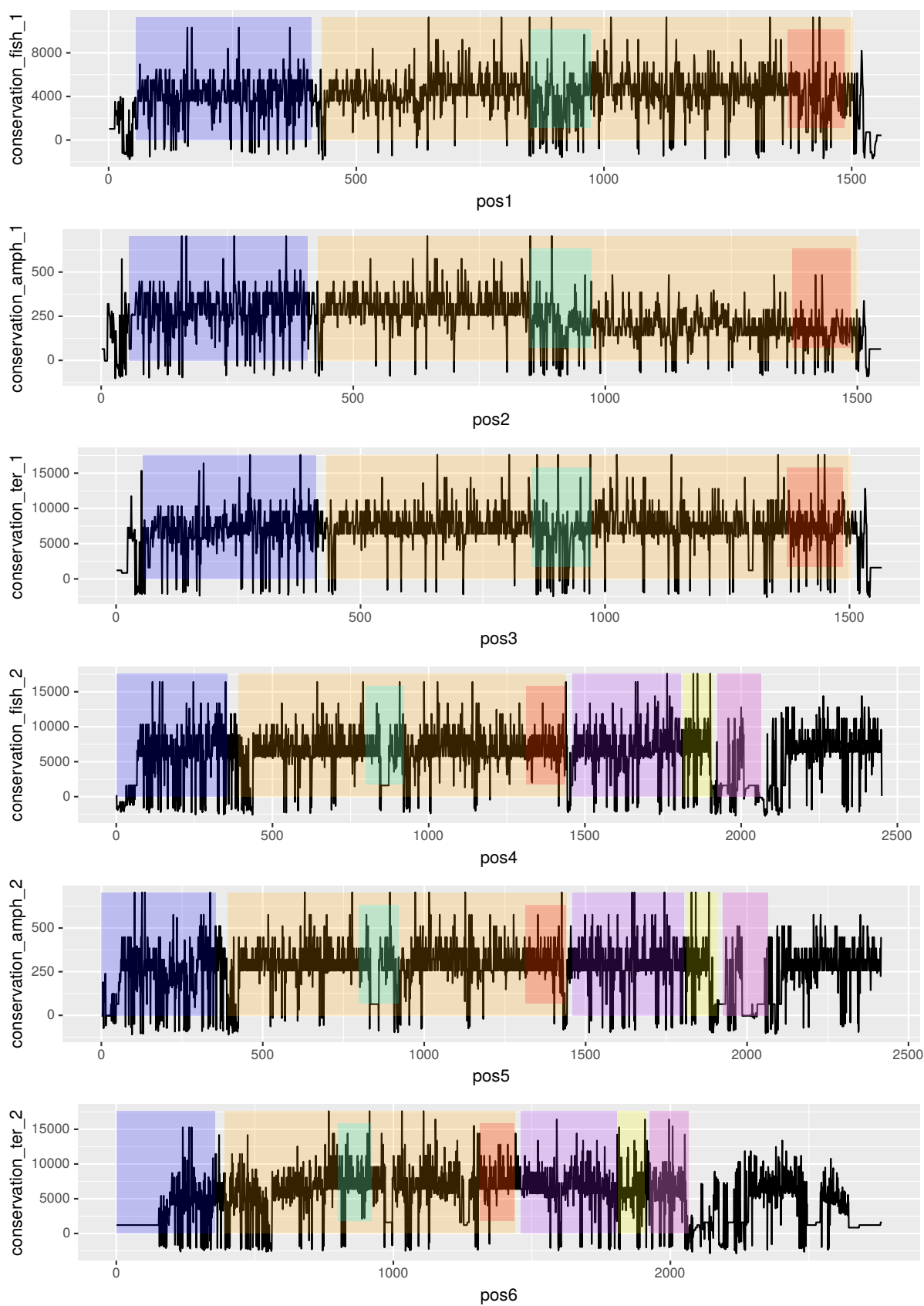


Figure 6: Conserved Residues **CPSI** belongs to (1)Fishes, (2)Amphibians, (3)Terrestrial Animals and **CPSII** belongs to (4)Fishes, (5)Amphibians, (6)Terrestrial Animal. Consensus among all **CPSI** and **CPSII** given equal weight for this comparison.

### **3 Results Discussion**

Due to time limitations, we could not make all our calculations ready, but finally, we compared consensus for CPSI, CPSII and CPSIII to sequences acquired from different species (fish, amphibian and terrestrial). Results are given in Fig 6. We observed some domains are related to big variations among species while some domains are strictly conserved.