

Uma abordagem baseada em agrupamento hierárquico para análise filoproteômica de venenos de serpentes medidos por espectrometria de massas

Aluno: Guilherme Costa Vieira
Orientador: Marcelo da Silva Reis

01

Introdução

Venenos de serpentes

Venenos de serpentes são uma complexa mistura de **proteínas e peptídeos**

A composição desses venenos é estudada pela proteômica baseada em experimentos de **cromatografia líquida - espectrometria de massas (LC-MS)**

Estudos recentes indicam que os **perfis proteômicos** das serpentes do gênero *Bothrops* se **correlacionam** com **árvores filogenéticas** desses animais

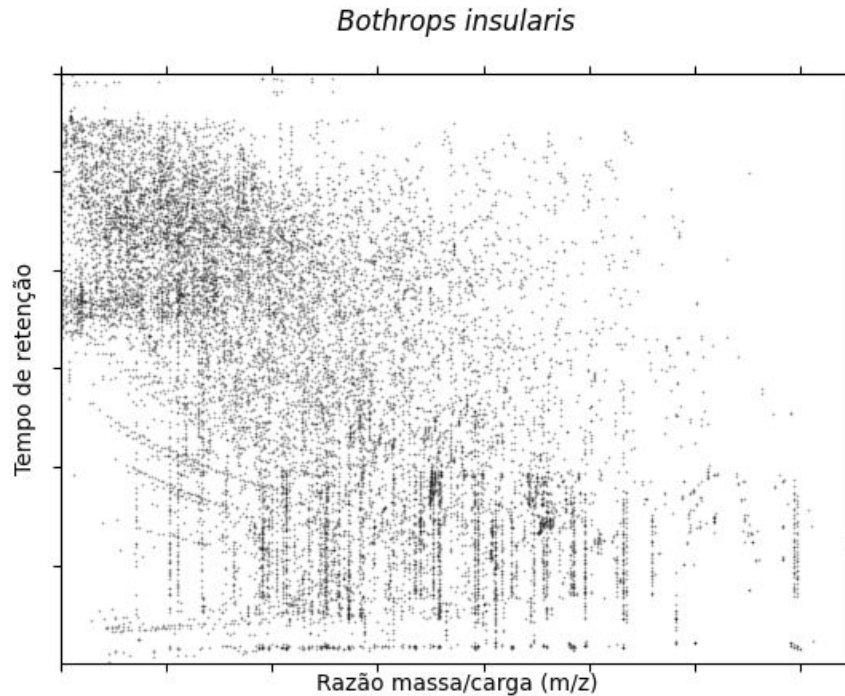
Venenos de serpentes

Problema: recuperar árvores filoproteômicas a partir dados gerados por LC-MS e verificar a existência dessa correlação

Métodos usados no trabalho que mostrou esses indícios apresentavam limitações como **incompletude** de bancos de dados de proteínas

Grupo do Dr. Marcelo Silva Reis desenvolveu métodos que dispensam o uso do banco de dados e apenas utilizam dados brutos de LC-MS

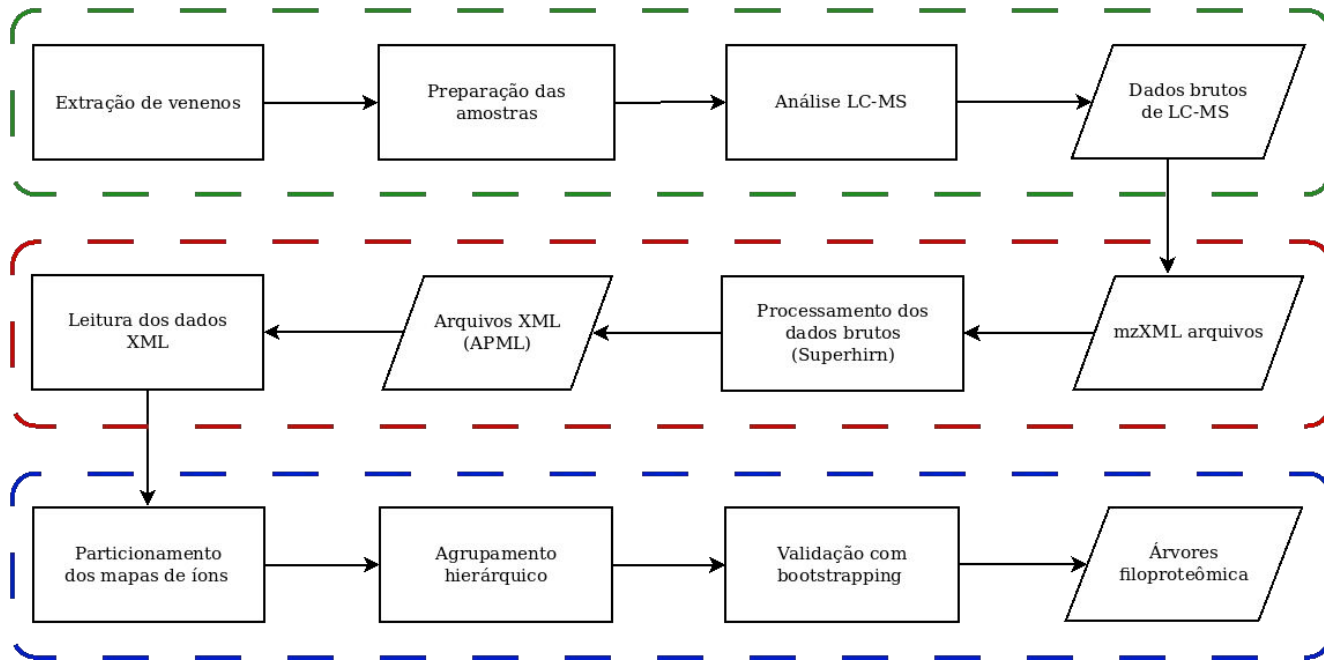
Mapas de íons (MITEs)



02

Metodologia

Encadeamento de processos



Agrupamento hierárquico

Técnica de agrupamento de objetos baseados em similaridade

Distâncias

Euclidiana

Manhattan

Jaccard (binária)

Canberra

Cosseno

Métodos de ligação

Simples: menor distância de qualquer ponto do 1º grupo para qualquer ponto do 2º grupo.

Completa: maior distância de qualquer ponto do 1º grupo para qualquer ponto do 2º grupo.

Média das distâncias: média das distâncias de todos os pontos do 1º grupo em relação aos pontos do 2º grupo.

Bootstrapping

Como verificar se os agrupamentos formados são significativos?

Ideia: Gerar novas amostras aleatoriamente a partir da amostra original

Dois p-valores são fornecidos pela biblioteca **Pvclust**:

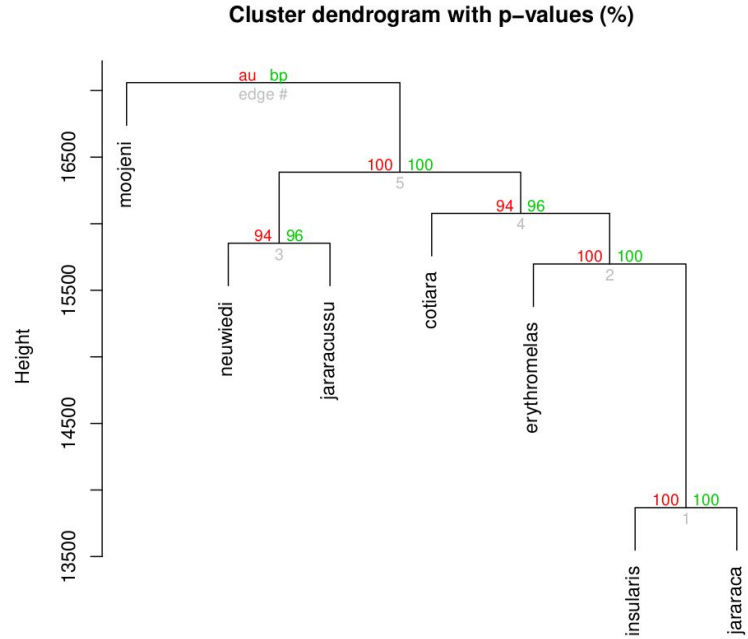
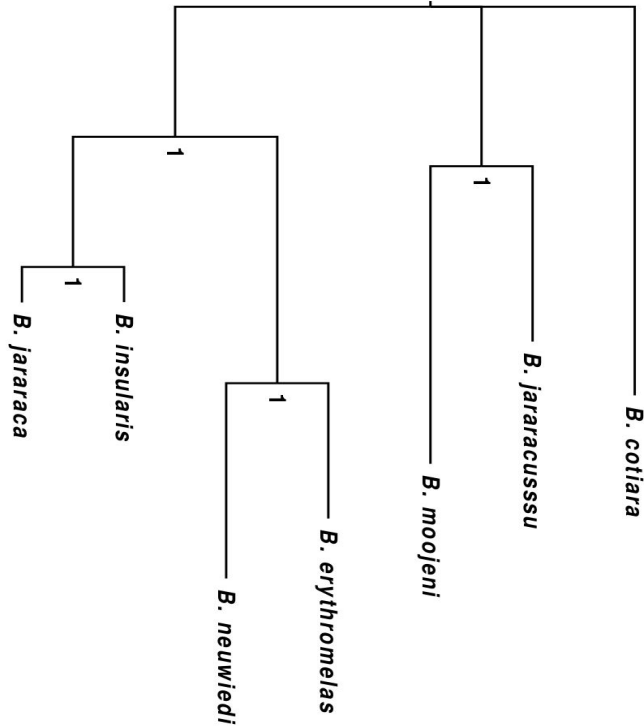
Bootstrap probability (**BP**): frequência que um cluster aparece nas réplicas geradas na reamostragem.

Approximately unbiased (**AU**): p-valor aproximadamente não-viesado.

03

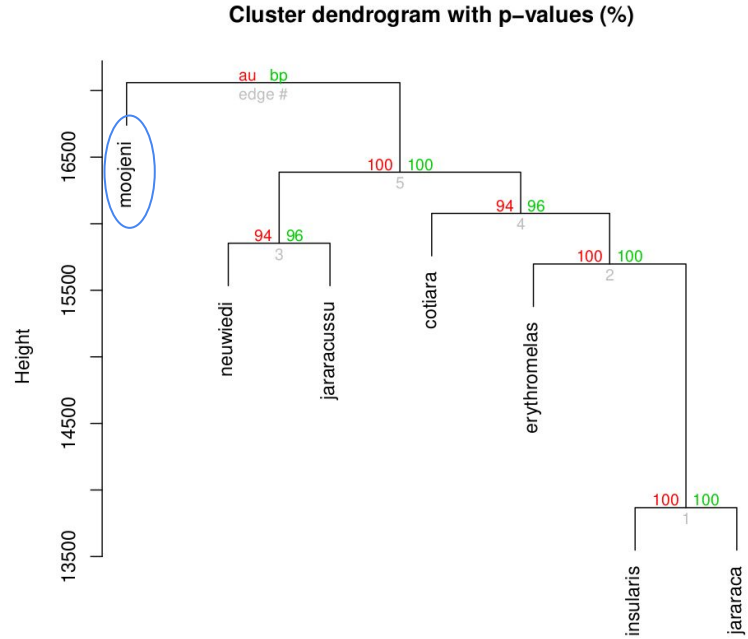
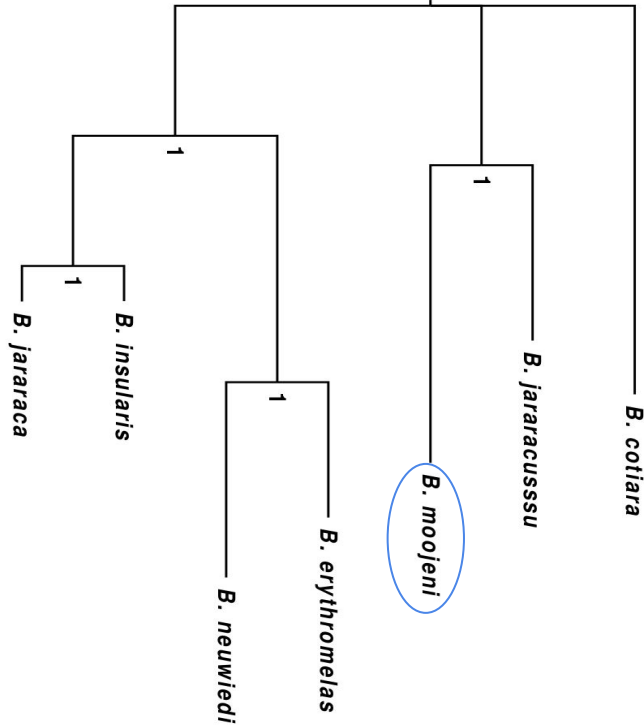
Resultados

Resultados



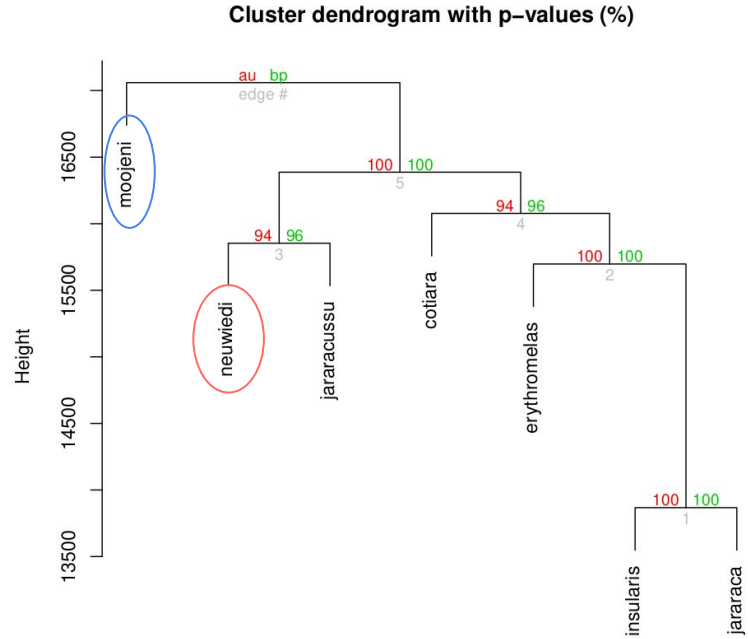
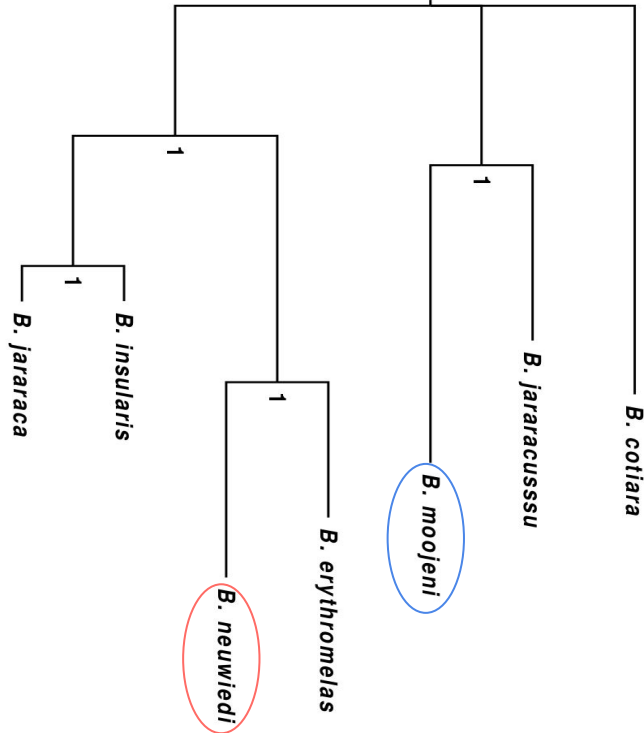
Distance: canberra
Cluster method: average

Resultados



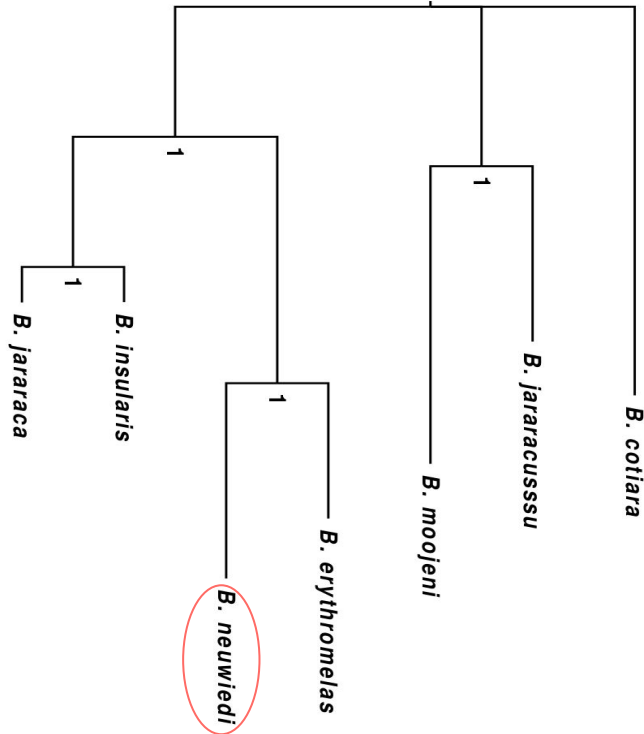
Distance: canberra
Cluster method: average

Resultados

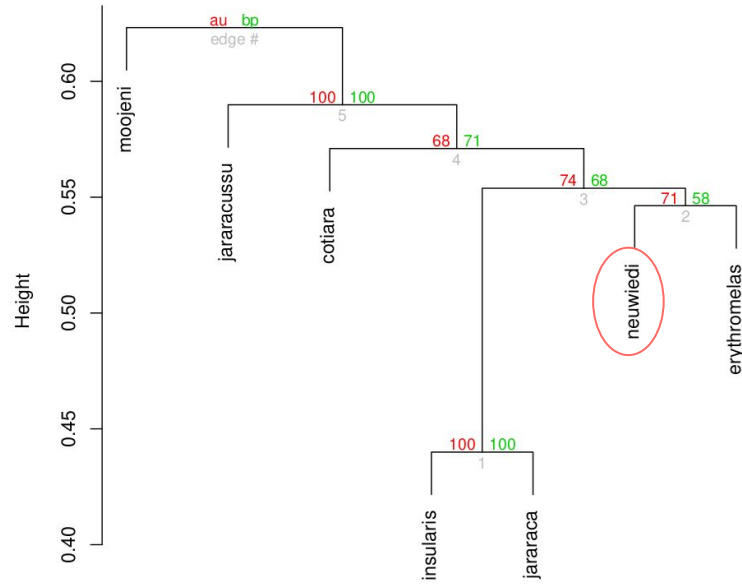


Distance: canberra
Cluster method: average

Resultados



Cluster dendrogram with p-values (%)



Distance: binary
Cluster method: average

04

Conclusões

Conclusões

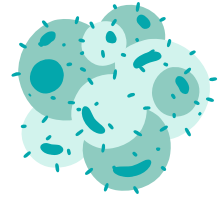
- Os agrupamentos confirmaram uma relação próxima entre as árvores filoproteômicas e filogenéticas
- Análises realizadas com dados brutos são muito similares aos resultados obtidos com identificação de peptídeos e proteínas
- A perspectiva é que essas metodologias sejam aplicadas em outros venenos e outros tipos de dados proteômicos

Referências

- [1] Débora Andrade-Silva, André Zelanis, Eduardo S Kitano, Inácio LM Junqueira-de Azevedo, Marcelo S Reis, Aline S Lopes, and Solange MT Serrano. Proteomic and glyco-proteomic profilings reveal that post-translational modifications of toxins contribute to venom phenotype in snakes. *Journal of proteome research*, 15(8):2658–2675, 2016.
- [2] D. Andrade-Silva, D. Ashline, T. Tran, A.S. Lopes, S.R. Travaglia Cardoso, M.S. Reis, A. Zelanis, S.M.T. Serrano, and V. Reinhold. Structures of N-Glycans of Bothrops venoms revealed as molecular signatures that contribute to venom phenotype in viperid snakes. *Molecular & Cellular Proteomics*, 17(7):1261–1284, 2018.
- [3] Victor Wichmann Raposo. Análise filogenética computacional de serpentes do gênero bothrops a partir de proteomas de venenos. Technical report, Instituto de Matemática e Estatística, Universidade de São Paulo, 2018. Monografia de graduação em Computação.
- [4] Gustavo Mendes Maciel. Um método baseado em multirresolução para análise filo-proteômica de venenos de serpentes. Technical report, Instituto de Matemática e Estatística, Universidade de São Paulo, 2019. Monografia de graduação em Computação.
- [5] Jay W Fox and Solange MT Serrano. Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. *Proteomics*, 8(4):909–920, 2008.
- [6] Gentleman, R., Carey, V., Huber, W., Irizarry, R., & Dudoit, S. (Eds.). (2005). *Bioinformatics and computational biology solutions using R and bioconductor* (2005th ed.). doi:10.1007/0-387-29362-0.
- [7] Suzuki, R. and Shimodaira, H. (2004) "An application of multiscale bootstrap resampling to hierarchical clustering of microarray data: How accurate are these clusters?", The Fifteenth International Conference on Genome Informatics 2004, P034.
- [8] Suzuki, R. and Shimodaira, H. (2006) "Pvclust: an R package for assessing the uncertainty in hierarchical clustering", *Bioinformatics*, 22 (12): 1540-1542.
- [9] Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51:492–508.
- [10] Shimodaira, H. (2004). Technical details of the multistep-multiscale bootstrap resampling. Research Report B-403, Dept. Mathematical and Computing Sciences, Tokyo, Institute of Technology, Tokyo.

Thanks!

Do you have any questions?
gcvieira@protonmail.com



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**

Please keep this slide for attribution