Computational phylogenetic analysis of Bothrops snakes from venom proteomes

Victor Wichmann Raposo[1,2], Carolina Brás[2,3], Solange M. T. Serrano [2,3], Felipe Grazziotin[2,4], Inácio L. M. Junqueira de Azevedo[2,3], and Marcelo S. Reis[2,5]

[1] Instituto de Matemática e Estatística, Universidade de São Paulo (USP), Brazil;
[2] Center of Toxins, Immune-Response and Cell Signaling (CeTICS), Instituto Butantan, Brazil;
[3] Laboratório Especial de Toxinologia Aplicada,
[4] Laboratório de Coleções Zoológicas,
[5] Laboratório Especial de Ciclo Celular, Instituto Butantan, Brazil

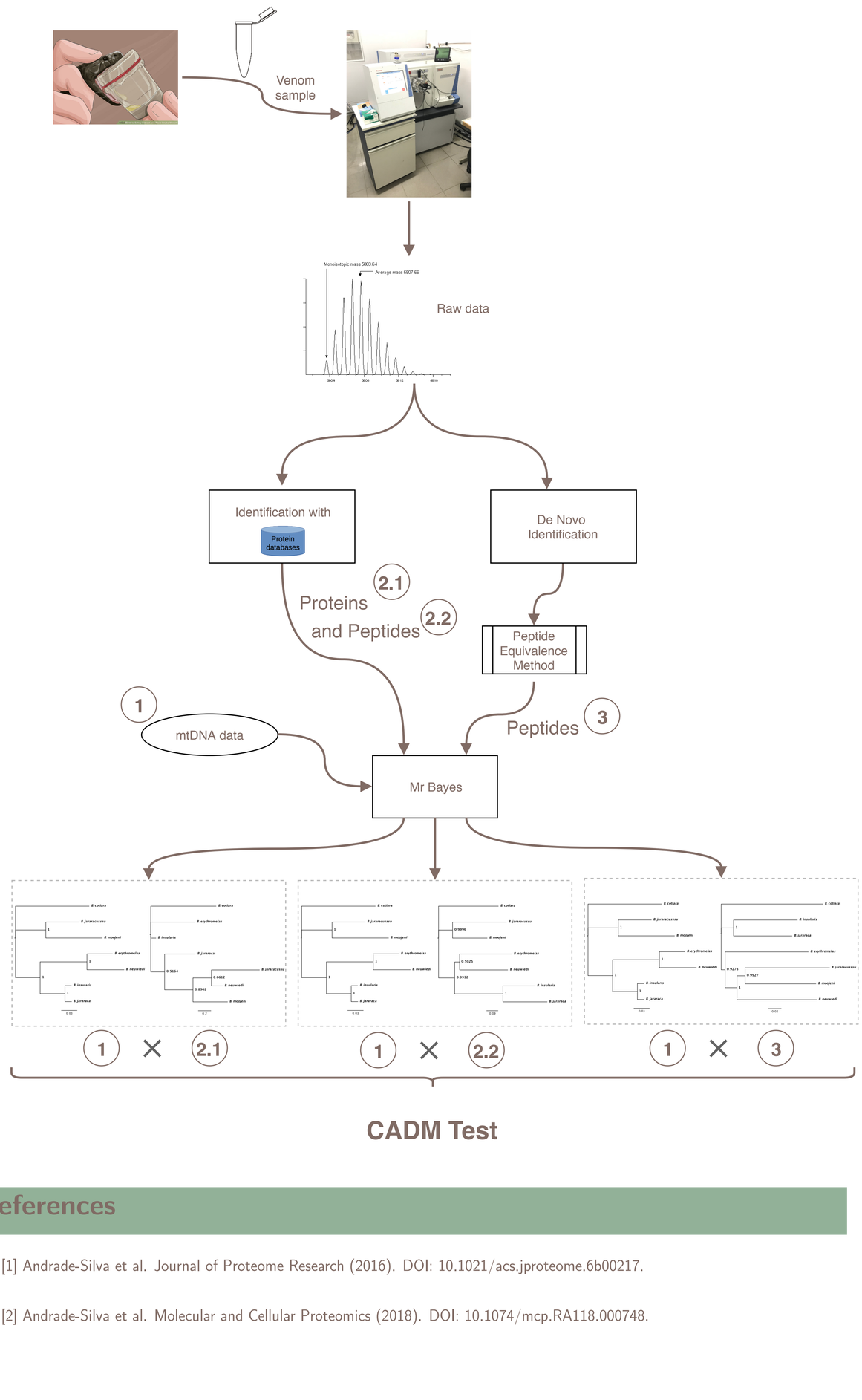
Background and objective

Snake venoms are complex protein-based mixtures, whose proteins can undergo variable levels of glycosylation. There is interspecies variation both in the mixture composition (proteome) and in the types of glycan structures that bind to its proteins. Recently, it was reported that, among *Bothrops* snakes, cladograms obtained using either proteome [1] or N-glycan structures [2] correlate with the phylogenetic tree produced through genomic sequences. However, in these studies, it was not applied quantitative metrics for comparison among different cladograms. Moreover, it was not totally exhausted the usage of information contained in the peptides detected during the mass spectrometry (MS)-based proteomics assays.

Objectives

The general goal of this project is to design a novel methodology for generating, comparing and visualizing phylogenetic trees yielded by non-traditional data in evolutionary analysis (e.g., peptide information of MS-based proteomic assays). Particularly, our goal is to use the developed methodology to test the hypothesis that the proteomic profile of venoms from *Bothrops* snakes is highly correlated to their phylogeny.

Methods



Results

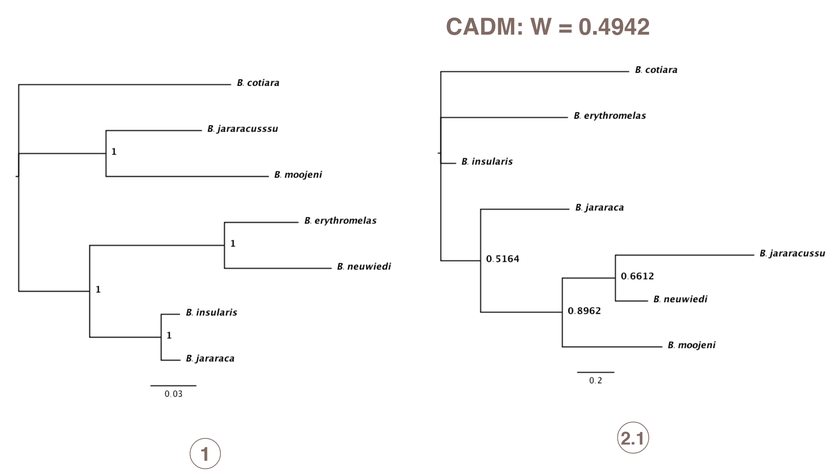


Figure: Comparison of the cladogram obtained with genomic data (1) with proteic data (2.1). With a visual inspection, we can see that the trees are not topologically congruent. This conclusion is corroborated by the CADM test.

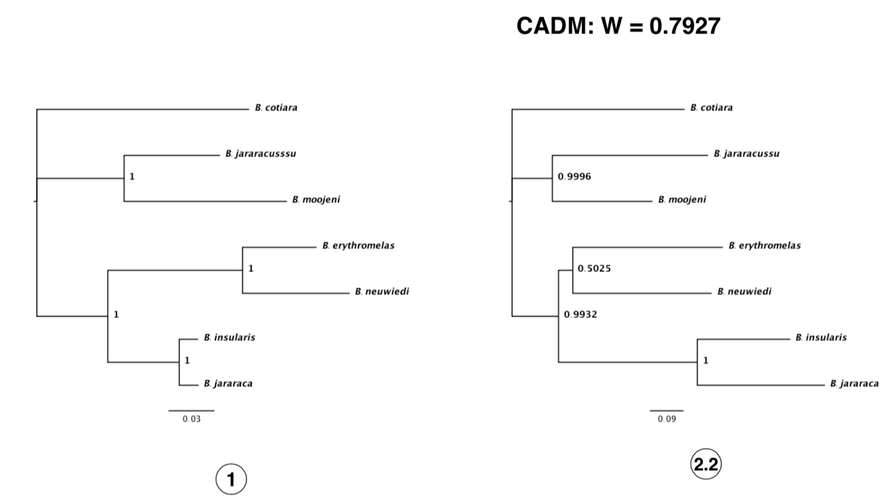


Figure: Comparison of the cladogram obtained with genomic data (1) with data from peptides identified using a protein database (2.2). The result of the CADM test shows that, the trees are very topologically congruent; However the *posteriori* probability for divergence between *B. neuwiedi* and *B. erythromelas* is small, which implies that the postioning of *B. neuwiedi* is not robust.

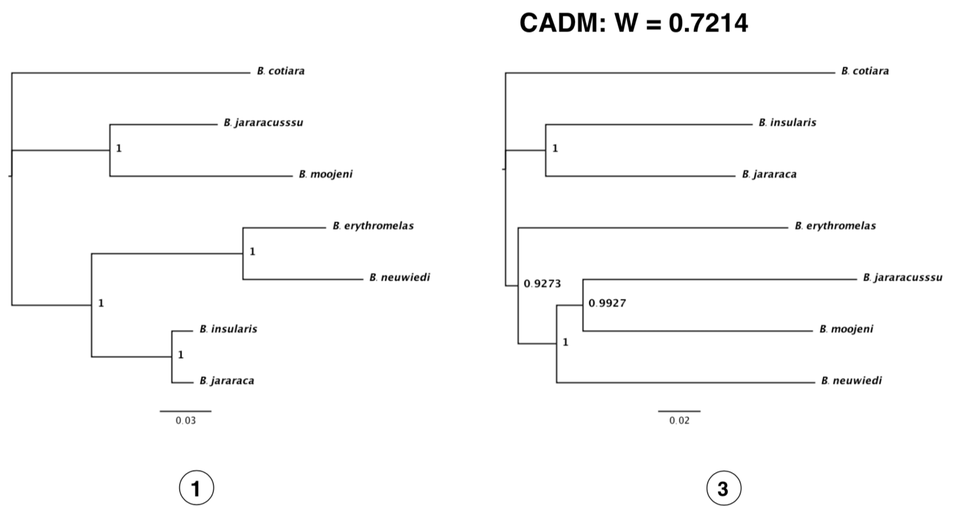


Figure: Comparison of the cladogram obtained with genomic data (1) with data from peptides identified using the de novo protocol (3). The result of the CADM test shows that, the trees are very topologically congruent, hence validating the peptide equivalence method.

Conclusions and future work

- We developed and validated a pipeline for designing of phyloproteomic trees using MS-based data and for cladogram comparison against mtDNA-based phylogenetic trees.
- With this pipeline we showed that a tree generated with peptide data from *Bothrops* snakes venoms is topologically congruent to the phylogenetic one, with the exception of *B. neuwiedi*.
- The next step in this research line is to extend the methodology to other snakes and also to apply it in other biological contexts (e.g. evaluation of epigenetic modifications in cancer cells).

References

[1] Andrade-Silva et al. Journal of Proteome Research (2016). DOI: 10.1021/acs.jproteome.6b00217.

[2] Andrade-Silva et al. Molecular and Cellular Proteomics (2018). DOI: 10.1074/mcp.RA118.000748.

victor.raposo@usp.br