A multiresolution based method for phyloproteomic analysis of snake venoms

Gustavo M. Maciel[1,2,4], **Débora Andrade-Silva**[3,4], **Solange M.T. Serrano**[3,4] **e Marcelo S. Reis**[2,4]

[1] Instituto de Matemática e Estatística, Universidade de São Paulo

[2] Laboratório de Ciclo Celular, Instituto Butantan

[3] Laboratório de Toxinologia Aplicada, Instituto Butantan

[4] Center of Toxins, Immune-response and Cell Signaling (CeTICS), Instituto Butantan

gustavo.maciel@usp.br
marcelo.reis@butantan.gov.br

Introduction

Snake venoms are complex protein-based mixtures, essential for the survival of the venomous species. Liquid chromatography-mass spectrometry (LC-MS) is a popular analytical technique that has been widely used in identifying the composition of these mixtures. Recent studies utilizing this technique have shown a strong connection between the venom proteomic profiles of *Bothrops* snakes and its phylogenetic classification. However, two of them [1, 2] did not apply quantitative metrics to assess the similarities among evolutionary trees obtained from the data. The other one [3] relied on the identification of peptides by *de novo* sequencing to avoid the overrepresentation of some species in protein databases. But this method introduces false positive candidate peptides, which can negatively affect the trees topology and/or branch lengths. To overcome this, a possibly suitable strategy is to use raw LC-MS data, without peptide identification.

Goals

- The main goal of this work is to develop a novel method to infer phyloproteomic trees from raw LC-MS data, without peptide identification, based on a multiresolution approach [4].
- A more specific objective is to apply this methodology to study the evolutionary relations between seven species of *Bothrops* snakes, utilizing proteomic data from its venoms, with the belief that it will lead to better results on the congruence of the inferred phyloproteomic trees compared to a reference phylogenetic one.

Methods

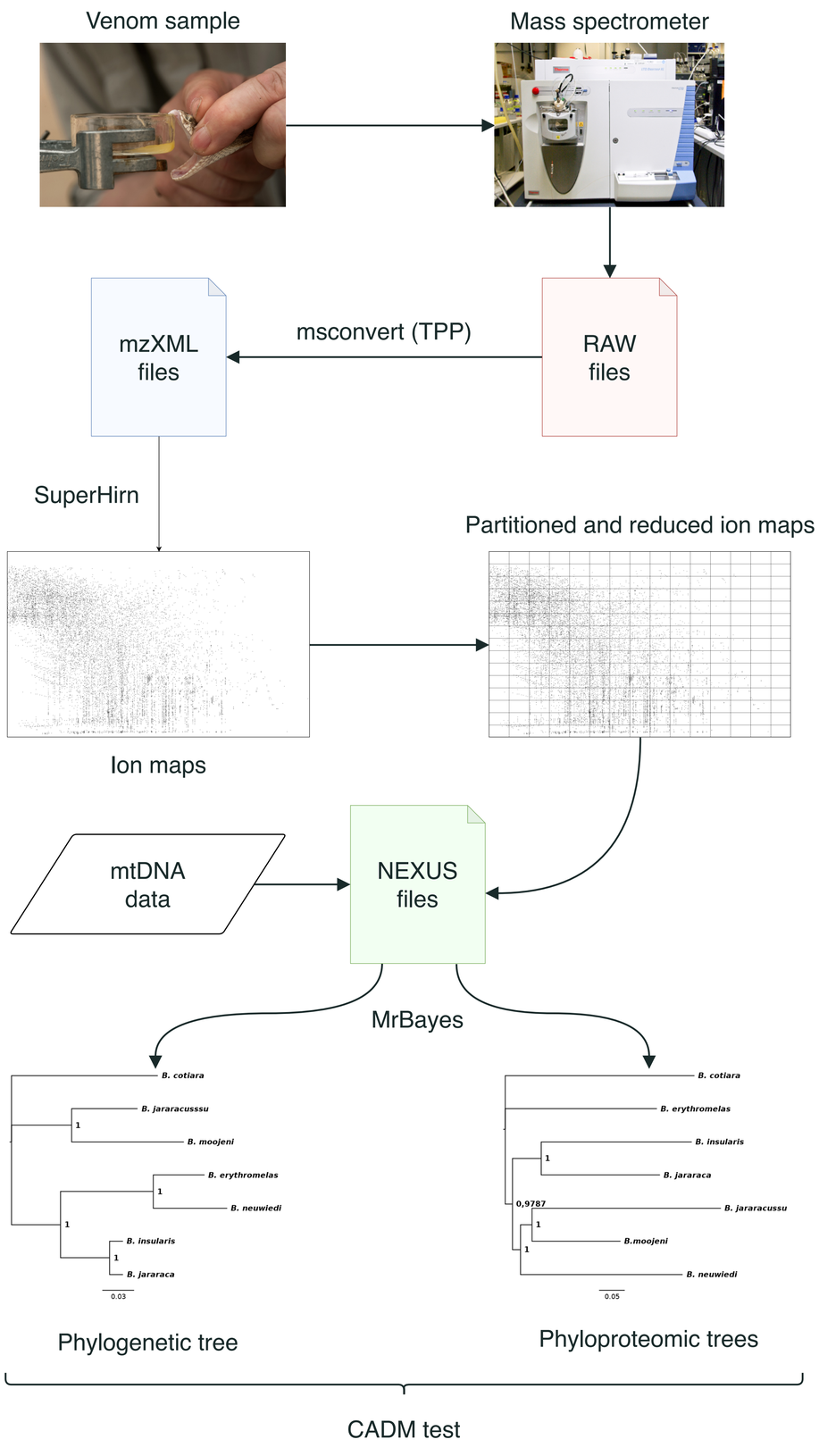

Figure 1: Main flow of the developed pipeline.

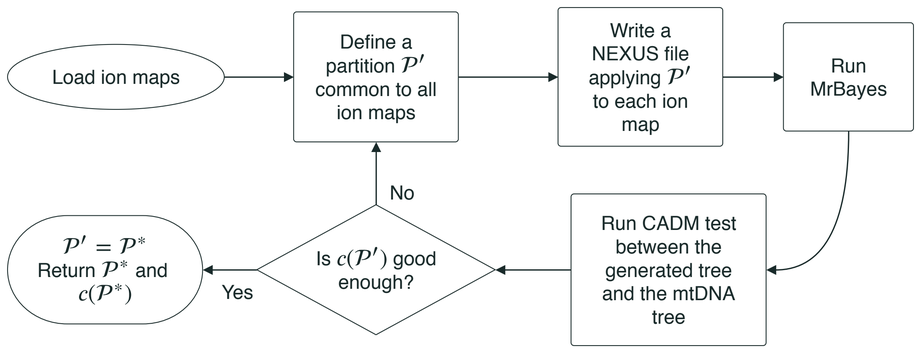

Figure 2: General idea of the optimization procedure used to maximize the CADM test score.

Results

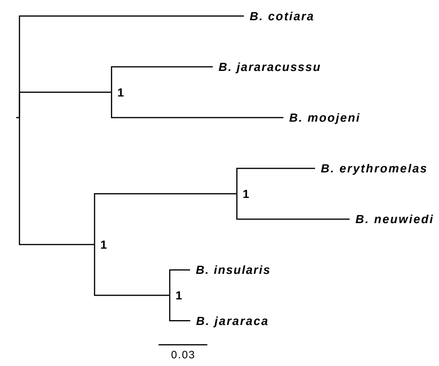

Figure 3: Phylogenetic tree inferred from mtDNA data. This tree is used as reference in all CADM tests performed.

$W = 0.4464048$ $W = 0.6080151$

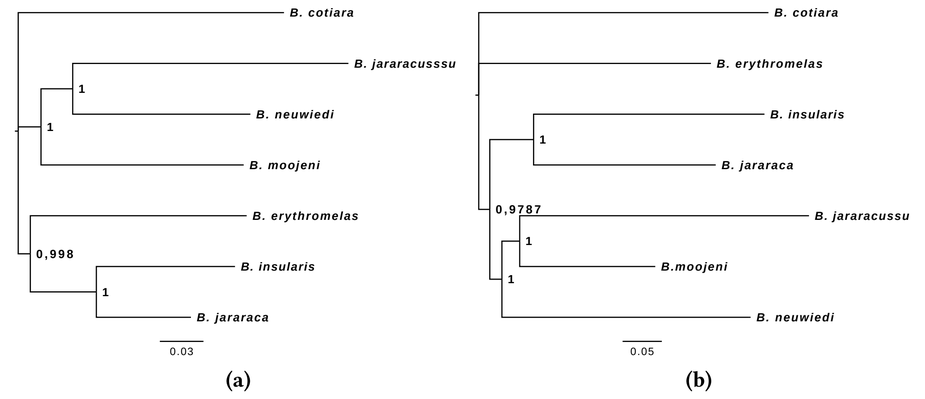

Figure 4: Phyloproteomic trees constructed from raw LC-MS data. The tree in Figure 4a was inferred using a naive method, and the other, shown in Figure 4b, using an optimization procedure based on the idea described in Figure 2.

As can be seen in Figure 4, the tree generated with the partitioning method performed way better than the other in the CADM test. This result was determined by the value of Kendall's coefficient of concordance W, but is also noticeable by looking at their topologies. Moreover, the posterior probabilities of the branches are sufficiently high to ensure that it is a trustworthy result.

Final considerations

- We developed a method for phyloproteomic analysis of raw data obtained from mass spectrometry experiments, without relying on peptide identification techniques, that showed coherent results.
- This work also reinforced that the venom proteomes of the seven *Bothrops* snakes studied are indeed correlated to their phylogeny. But, as well as stated in a previous work [3], the *B. neuwiedi* classification obtained diverges from the reference tree.

Future work

- The exploration of the optimization problem search space could be done utilizing some heuristic, such as genetic algorithms or simulated annealing.
- The method developed in this project could also be applied in other types of snake venom data, for instance, the structure of N-glycans of *Bothrops* venoms studied in a previous work [2].

References

[1] Andrade-Silva et al. Proteomic and glycoproteomic profilings reveal that post-translational modifications of toxins contribute to venom phenotype in snakes. *Journal of proteome research*, 2016.

[2] Andrade-Silva et al. Structures of N-Glycans of *Bothrops* venoms revealed as molecular signatures that contribute to venom phenotype in viperid snakes. *Molecular and Cellular Proteomics*, 2018. In revision.

[3] V. W. 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] D. A. Vaquero et al. A maximum-likelihood approach for multiresolution w-operator design. In *XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'05)*. IEEE, 2005.