

Análise filogenética computacional de serpentes
do gênero *Bothrops* a partir de proteomas de
venenos

Victor Wichmann Raposo
Supervisor: Marcelo Reis

IME - USP

Novembro 2018

DNA, RNA, proteína e Dogma Central

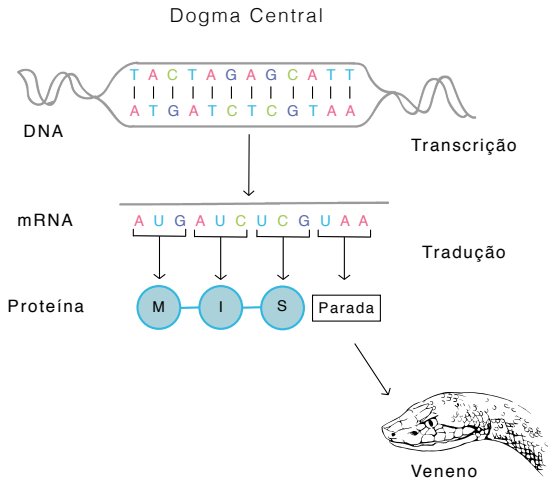


Figure 1: Ilustração da produção de proteínas de venenos em células de serpentes.

Cladogramas

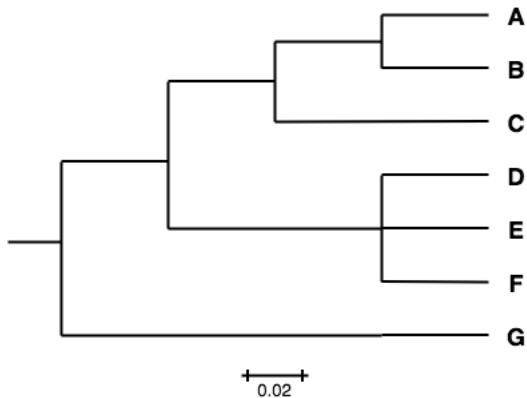


Figure 2: Exemplo de árvore filogenética de sete organismos.

Introdução

- ▶ Venenos de serpentes são misturas proteicas
- ▶ Proteoma e N-Glicanos são correlacionados com a filogenia
- ▶ Métrica de comparação de cladogramas
- ▶ Explorar informação de peptídeos

Objetivos

- ▶ **Geral:** metodologia de geração, comparação e visualização de cladogramas com dados não tradicionais
- ▶ **Específico:** testar a hipótese que o perfil proteômico de espécies do gênero *Bothrops* é correlacionado com sua filogenia

Inferência Bayesiana

$B(s)$:= número de árvores possíveis para s espécies

τ_i := i -ésima árvore

\mathbf{X} := um conjunto de informações biológicas

$$f(\tau_i|\mathbf{X}) = \frac{f(\mathbf{X}|\tau_i) f(\tau_i)}{\sum_{j=1}^{B(s)} f(\mathbf{X}|\tau_j) f(\tau_j)},$$

Inferência Bayesiana

$$f(\tau_i) \sim \text{Uniforme} \left(\frac{1}{B(s)} \right)$$

$$f(\mathbf{X}|\tau_i) = \int_v \int_{\theta} f(X|\tau_i, v, \theta) f(v, \theta) dv d\theta$$

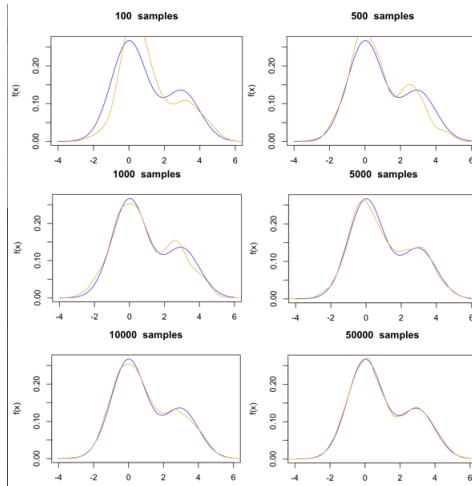


Figure 3: Convergência do algoritmo Metropolis-Hastings. O método MCMC, representado pelo linha laranja, tenta aproximar a distribuição azul. Fonte.

- ▶ Lapointe e Legendre 2004
- ▶ Estimar congruência entre cladogramas
- ▶ Nível de congruência de 0 (incongruência) a 1 (congruência total)

Banco de Dados

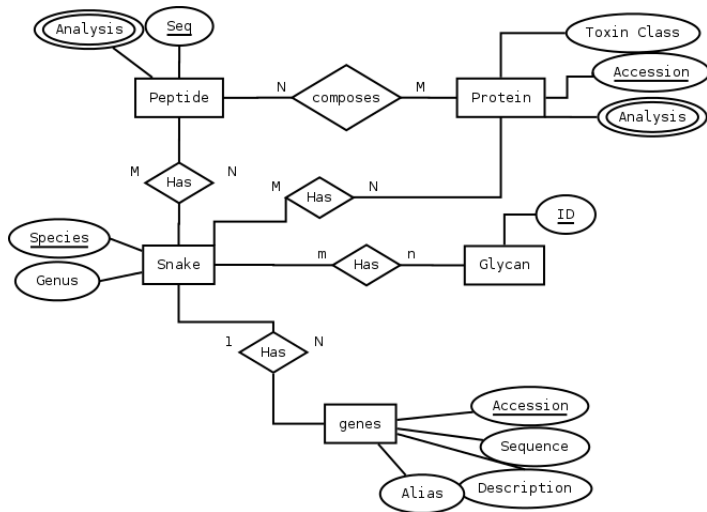


Figure 4: Modelo Entidade Relacional do banco de dados

Programas



Figure 5: Programas e tecnologias utilizadas no projeto

Metodologia

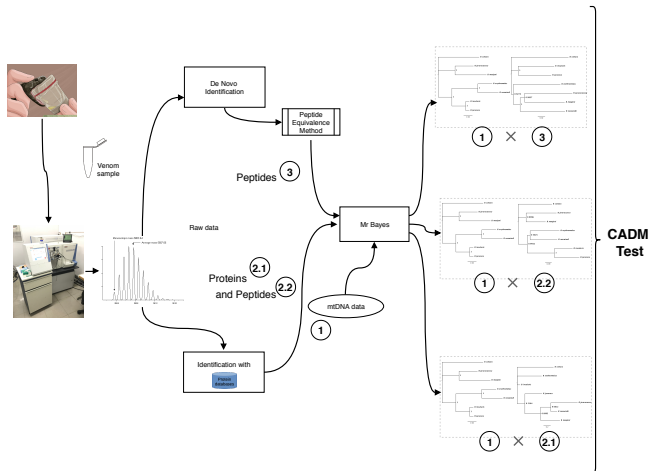


Figure 6: Metodologia aplicada no projeto.

Equivalência de Peptídeos

- ▶ BLAST
 - ▶ Score
 - ▶ E-value
- ▶ Relação de equivalência
- ▶ Union Find

Árvore genômica

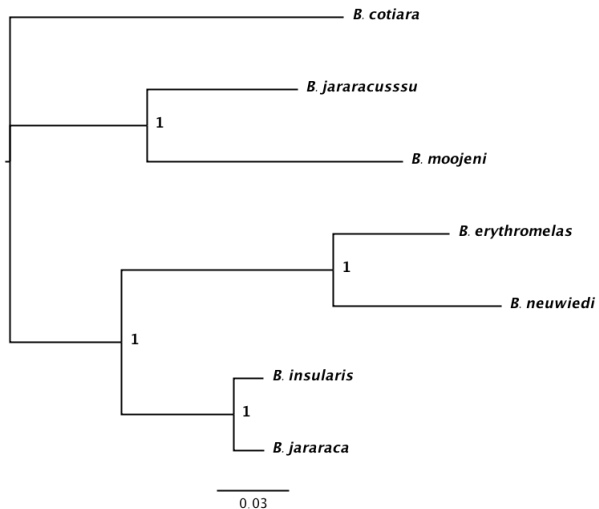


Figure 7

Árvore de Proteína

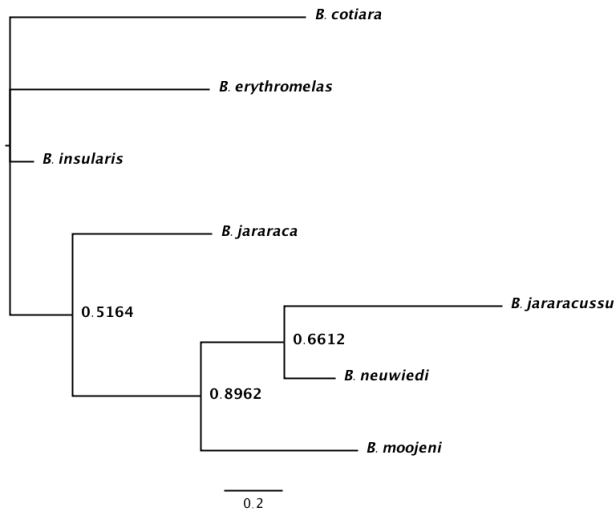


Figure 8: Cladograma obtido por dados proteicos (CADM : 0.4942).

Árvore de N-glicanos

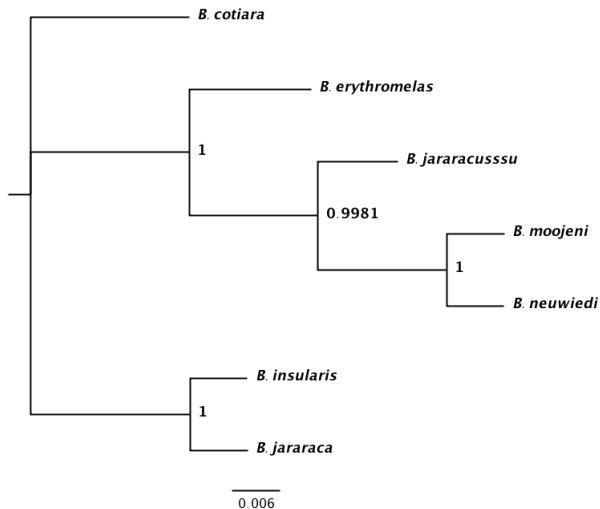


Figure 9: Cladograma obtido por dados de N-glicanos (CADM: 0.4653).

Árvore de Peptídeos

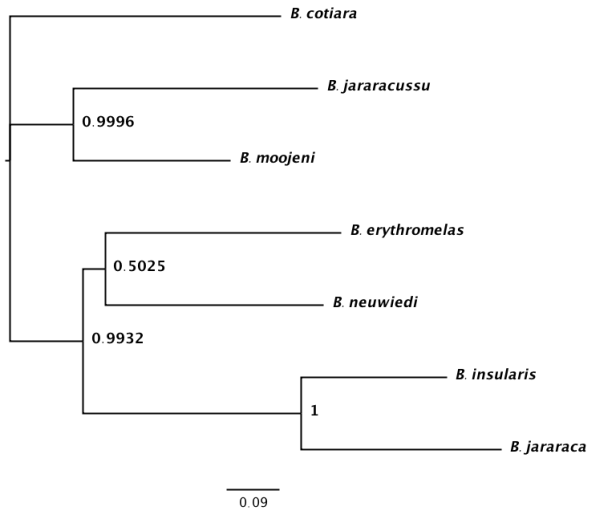


Figure 10: Cladograma obtido por dados peptídicos (CADM: 0.7927).

Árvore de Peptídeos (de novo)

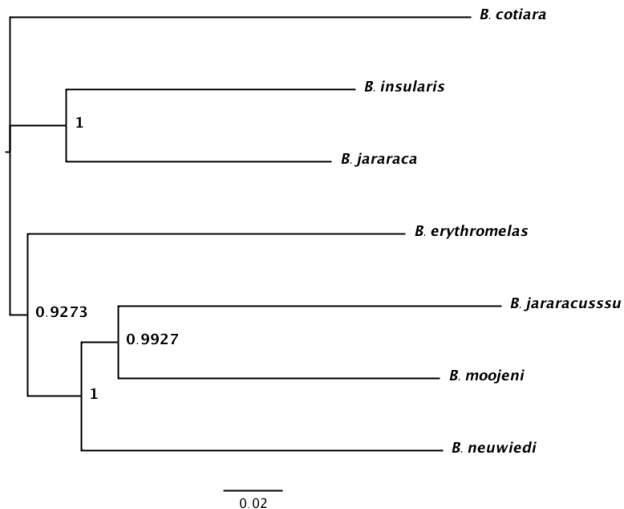


Figure 11: Cladograma obtido por dados de peptídeos identificados pelo protocolo de novo (CADM: 0.7214).

Resultado da Metodologia

Árvore (MIN_VALUE)	CADM	Número de Sequências
(a) (10^{-10})	0.6662	5408
(b) (10^{-5})	0.7207	4901
(c) (10^{-2})	0.7214	3258

Table 1: Resultados do teste CADM comparando as árvores de peptídeos de proteoma total sequenciados do modo de novo após aplicada a metodologia de equivalência de peptídeos variando o MIN_VALUE.

Conclusão

Contribuições Tecnológicas

- ▶ Pipeline desenvolvido
- ▶ Método de Equivalência de Peptídeos

Contribuições Científicas

- ▶ Árvores geradas com dados peptídicos são topologicamente congruentes
- ▶ Divergência da *B. newiedi*

Trabalhos Futuros

- ▶ Estender metodologia para novas serpentes
- ▶ Aplicar em outros contextos (e. g. modificações epigenéticas em células cancerígenas)
- ▶ Identificar o porquê da divergência da *B. neuwiedi*

Agradecimentos

- ▶ Dr. Inácio L. M. Junqueira de Azevedo (LETA)
- ▶ Dra. Solange M. T. Serrano (LETA)
- ▶ Dr. Felipe Grazziotin (Coleções Zoológicas)
- ▶ Carolina Brás (Doutoranda LETA)

