



**Vilniaus
universitetas**

Doktorantas:
Brendonas Stakauskas
2022-2026

Darbo vadovas:
Dr. Virginijus Marcinkevičius

2025/2026 metai
IV metai, II pusmetis



**Giliais neuroniniais tinklais grįstų
mašininio mokymosi metodų
taikymas viruso mutacijų trajektorijai
prognozuoti**



Studijų plano vykdymas

Studijų planas, vykdymo suvestinė

Studijų metai	Egzaminai		Dalyvavimas konferencijose				Publikacijos				
			Nacionalinėse		Tarptautinėse		Su citavimo rodikliu		Be citavimo rodiklio		Buklė
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	
I (2022/2023)	2	2	1	0	0	1	0	0	1	1 (konf. darbų medžiagoje)	
II (2023/2024)	2	2	1	0	0	1	0	0	1	0	
III (2024/2025)	0	0	0	1	1	0	1	0*	0	1 (konf. darbų medžiagoje)	*Įteikta žurnalui Evolutionary Bioinformatics. Atmesta Įteikta kitam žurnalui Journal of Computational Biology. Gauta recenzija, įteikti pataisymai.
IV (2025/2026)	0	0	0	0	1	0	1	0*	0	0	Įteikta žurnalui Scientific Reports
Iš viso:	4	4	2	1	2	2	2	0	2	3	

Ataskaitinio pusmečio darbo planas ir jo įvykdymas

Publikacijos 2025/2026 (1 pusmetis)			
Planas	Įvykdyta	Būklė	Publikacijos tipas
Journal of Computational Biology	B. Stakauskas, V. Marcinkevičius. Evaluating Protein Language Model Embeddings for Viral Clade Assignment	Įteikta : 2025-10-10 Įteikta (gautos pirmos recenzijos): 2026-02-25	CA WoS su Impact Factor (1.6)
Scientific Reports	B. Stakauskas, P. Górecki. Phylogenetic Signal in Protein Language Model Embeddings: A Zero-Shot Analysis of Distance-Based Tree Inference	Įteikta: 2025-12-31	CA WoS su Impact Factor (3.9)

Visų mokslinių tyrimų ir disertacijos rengimo etapai

1. Mokslinių tyrimų disertacijos tema apžvalga ir analizė
2. Mokslinio tyrimo vykdymas
 - 1.Tyrimo metodikos sudarymas
 - 2.Teorinis tyrimas
 - 3.Empirinis tyrimas
 - 4.Gautų duomenų analizė, apibendrinimas, išvadų parengimas**
3. Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas
4. Daktaro disertacijos parengimas ir svarstymas padalinyje
5. Daktaro disertacijos gynimas

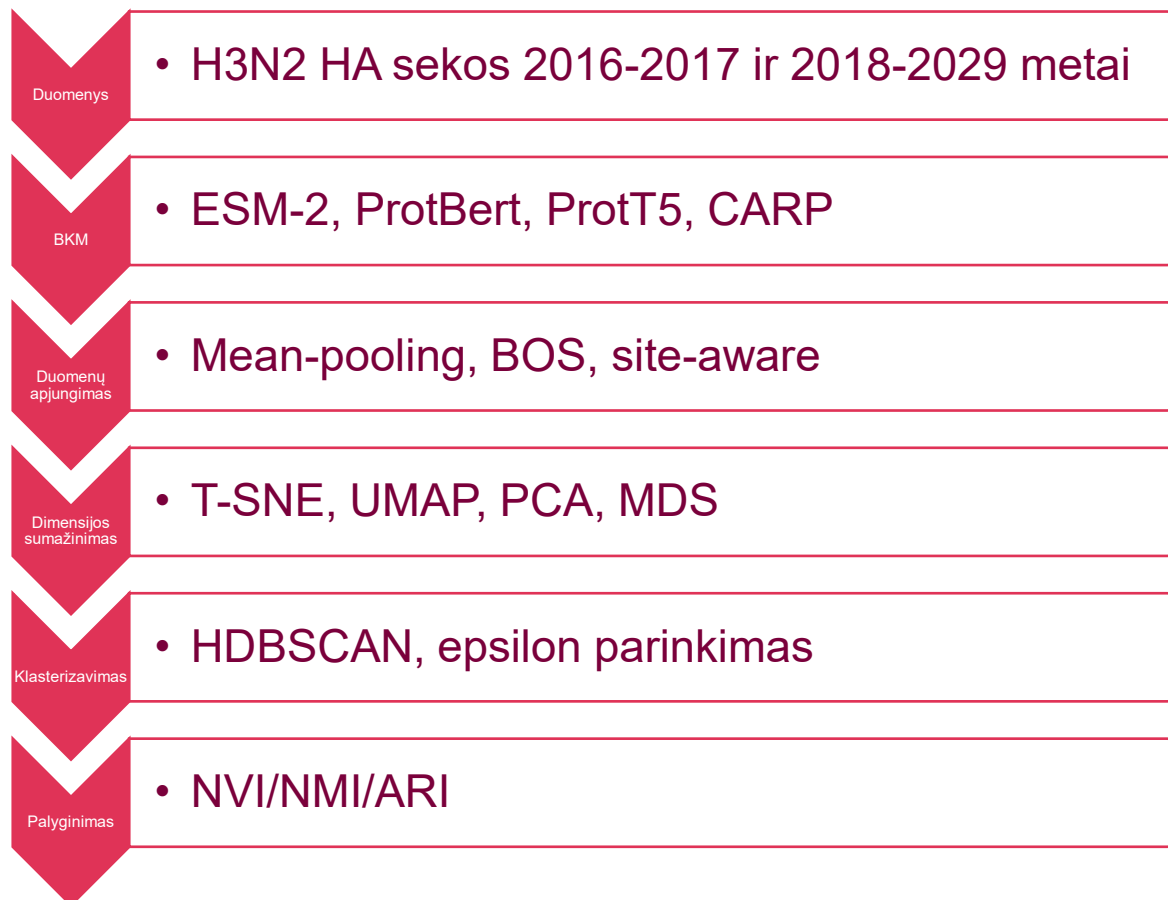
Trumpas per pusmetį gautų mokslinių rezultatų pristatymas

Atlikti tyrimai

B. Stakauskas, V. Marcinkevičius. *Evaluating Protein Language Model Embeddings for Viral Clade Assignment*

B. Stakauskas, P. Górecki. *Phylogenetic Signal in Protein Language Model Embeddings: A Zero-Shot Analysis of Distance-Based Tree Inference*

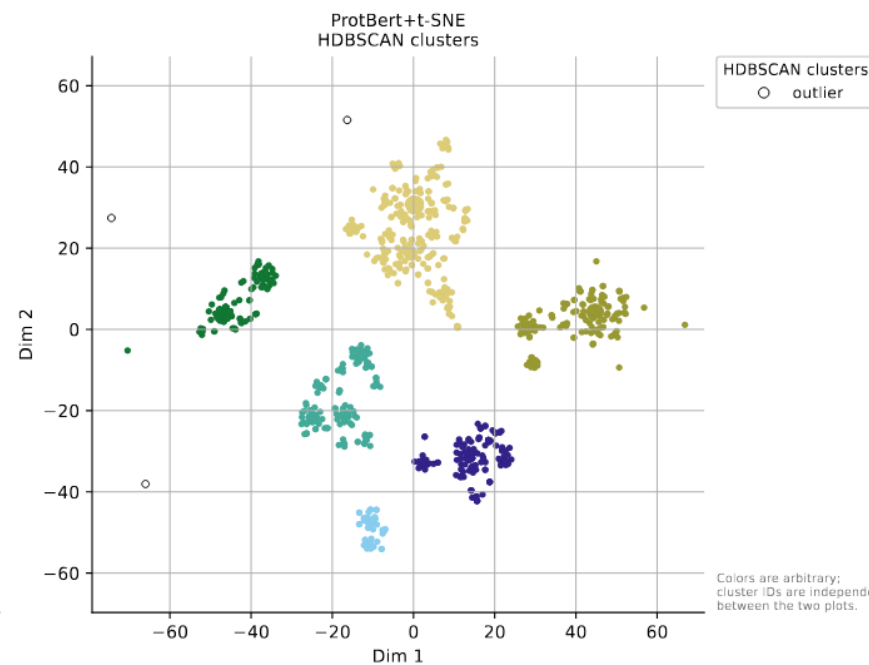
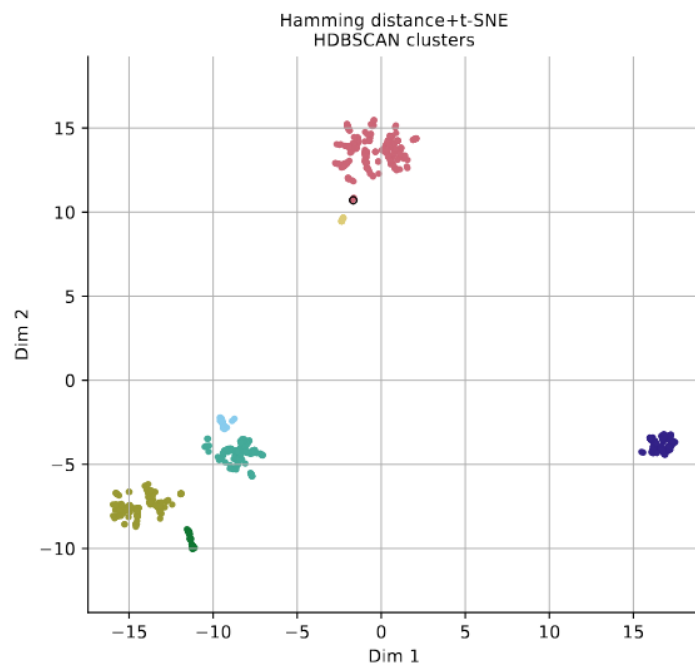
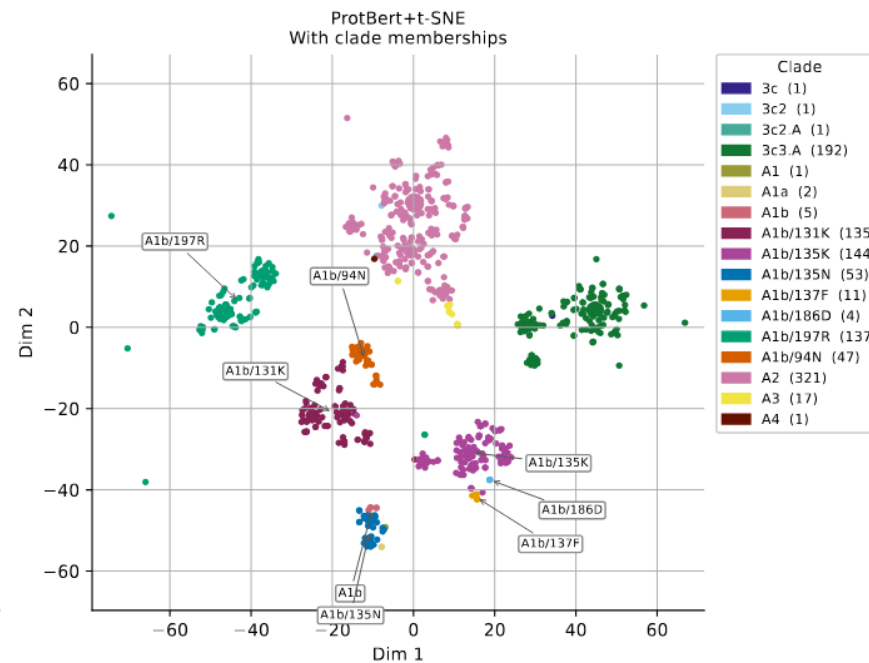
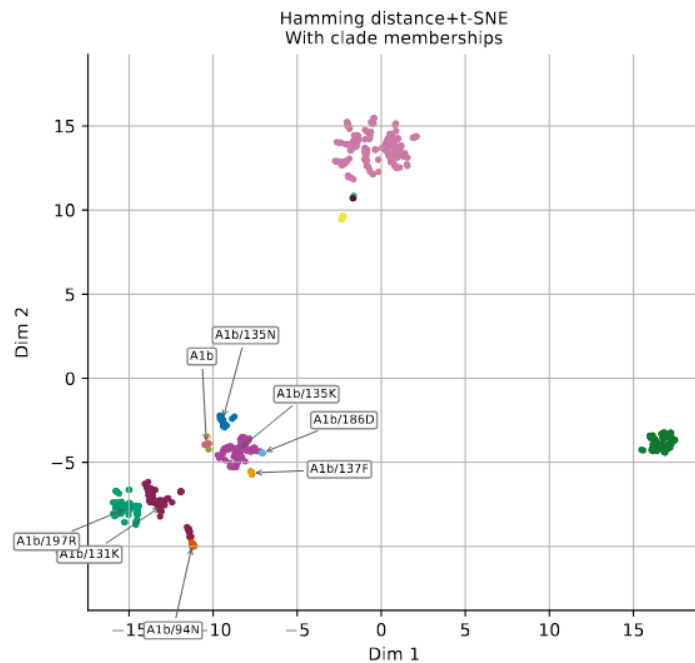
Kladų priskyrimas



Duomenų rinkinys

Split	N	Clades	Within-clade Hamming	Between-clade Hamming
Train	1,523	11	0.0059	0.0158
Test	1,073	17	0.0054	0.0247

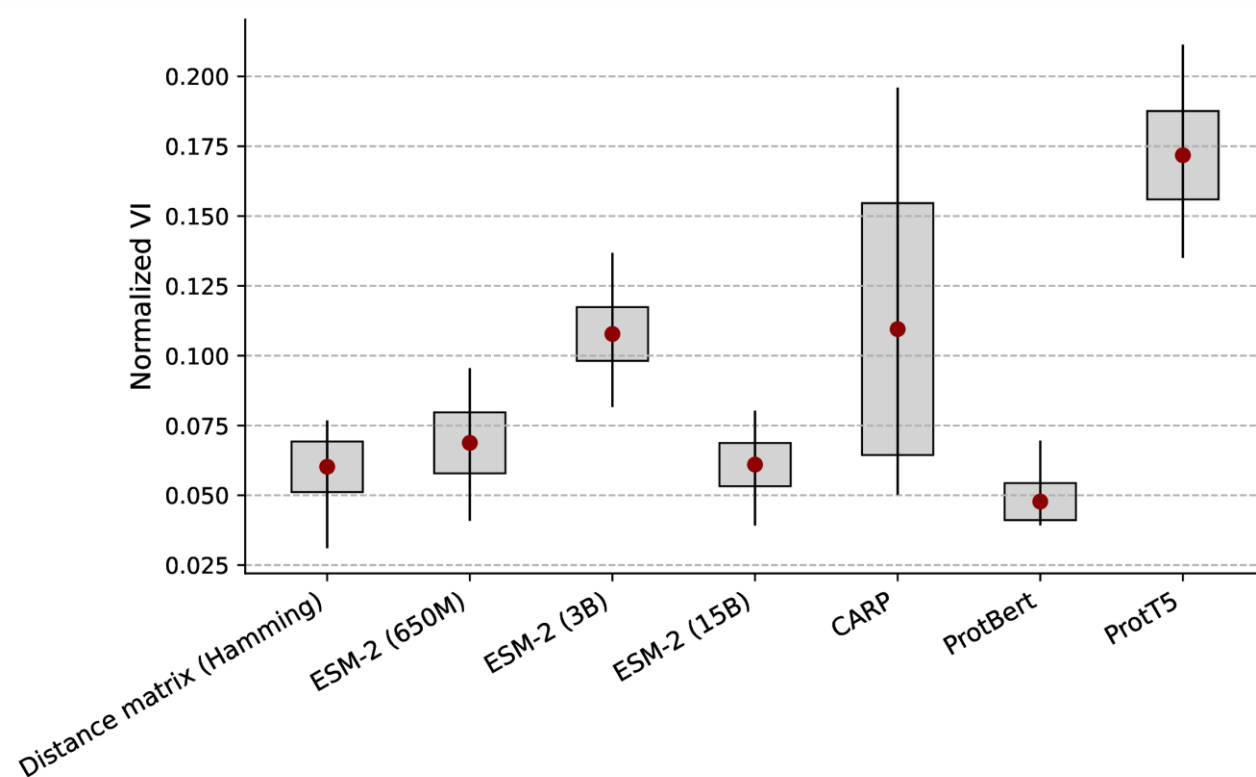
Atlikta 100 pakartojimų su atsitiktiniais duomenų poaibiais



Colors are arbitrary:
cluster IDs are independent
between the two plots.

Rezultatai

Model	NVI ↓	NMI ↑	ARI ↑
ProtBert	0.0445	0.9145	0.9046
ESM-2 (650M)	0.0544	0.9007	0.9111
ESM-2 (15B)	0.0562	0.9017	0.8710
Baseline (Hamming t-SNE)	0.0538	0.8969	0.8682
CARP	0.0728	0.8697	0.8459
ESM-2 (3B)	0.1058	0.8107	0.7869
ProtT5	0.1155	0.7967	0.6182



Filogenetinio medžio rekonstravimas

Duomenų rinkiniai:

- 500 skirtingų baltymų šeimų iš PANTHER DB
- Generuotos sekos naudojant SimPhy ir AliSim.

Modeliai:

ProtT5, ProtBert, E1, ESMC, *MSA Pairformer*

Metrikos:

- **RF** Robinson-Foulds: topologiniai skirtumai
- **KF** Kuhner-Felsenstein: šakų ilgių neatitikimas
- **Nye** Pomedžių panašumai
- **SPI** Shared Phylogenetic Information

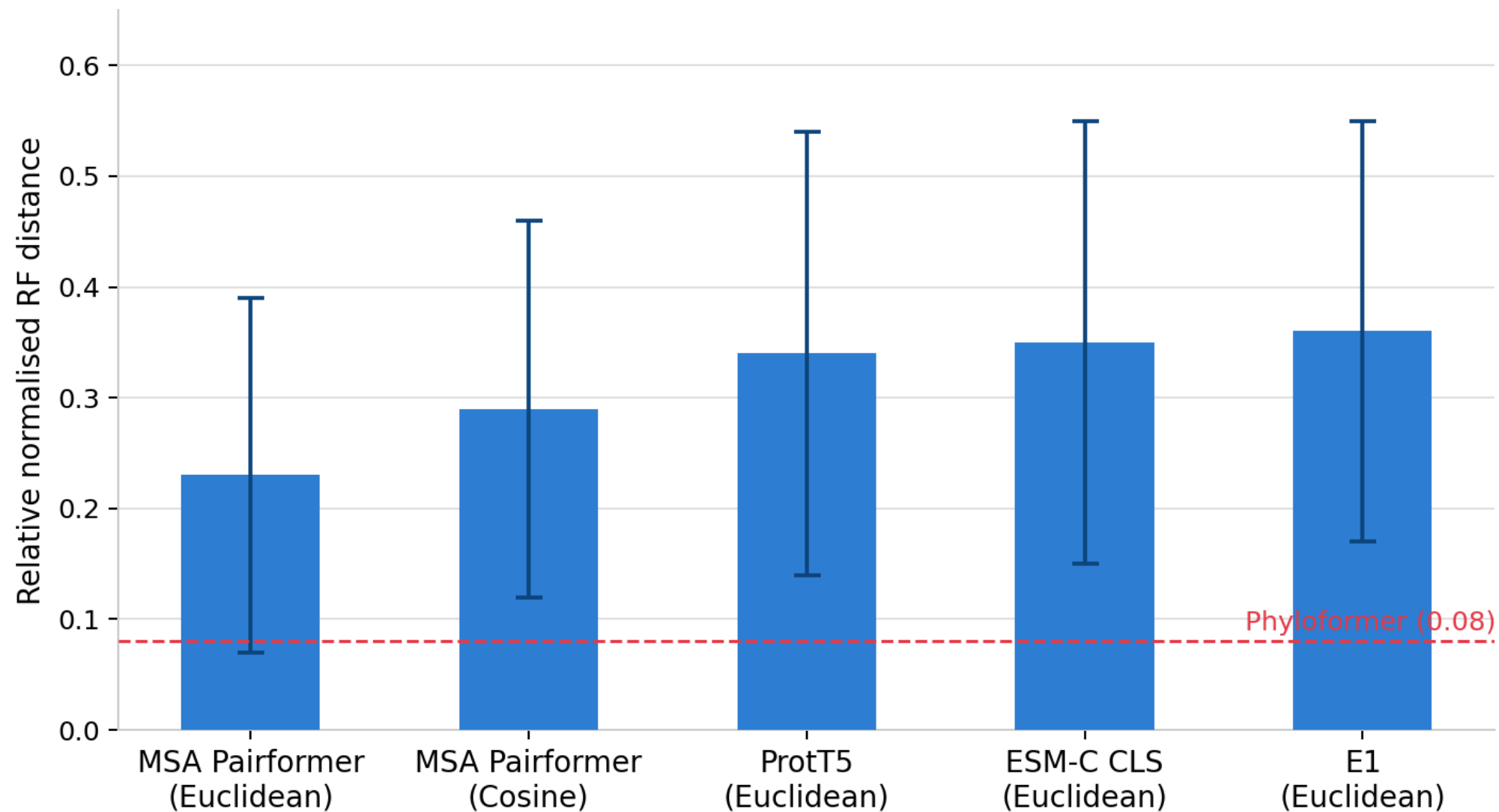
Metrika gaunama lyginant su „žinomu“ filogenetiniu medžiu

Rezultatas lyginamas su Phyloformer rezultatu

Atstumų agregavimas

- Mean-pooling
- Slenkantis langas
- *Padengiamumas*
- *Atsitiktiniai poabiai*

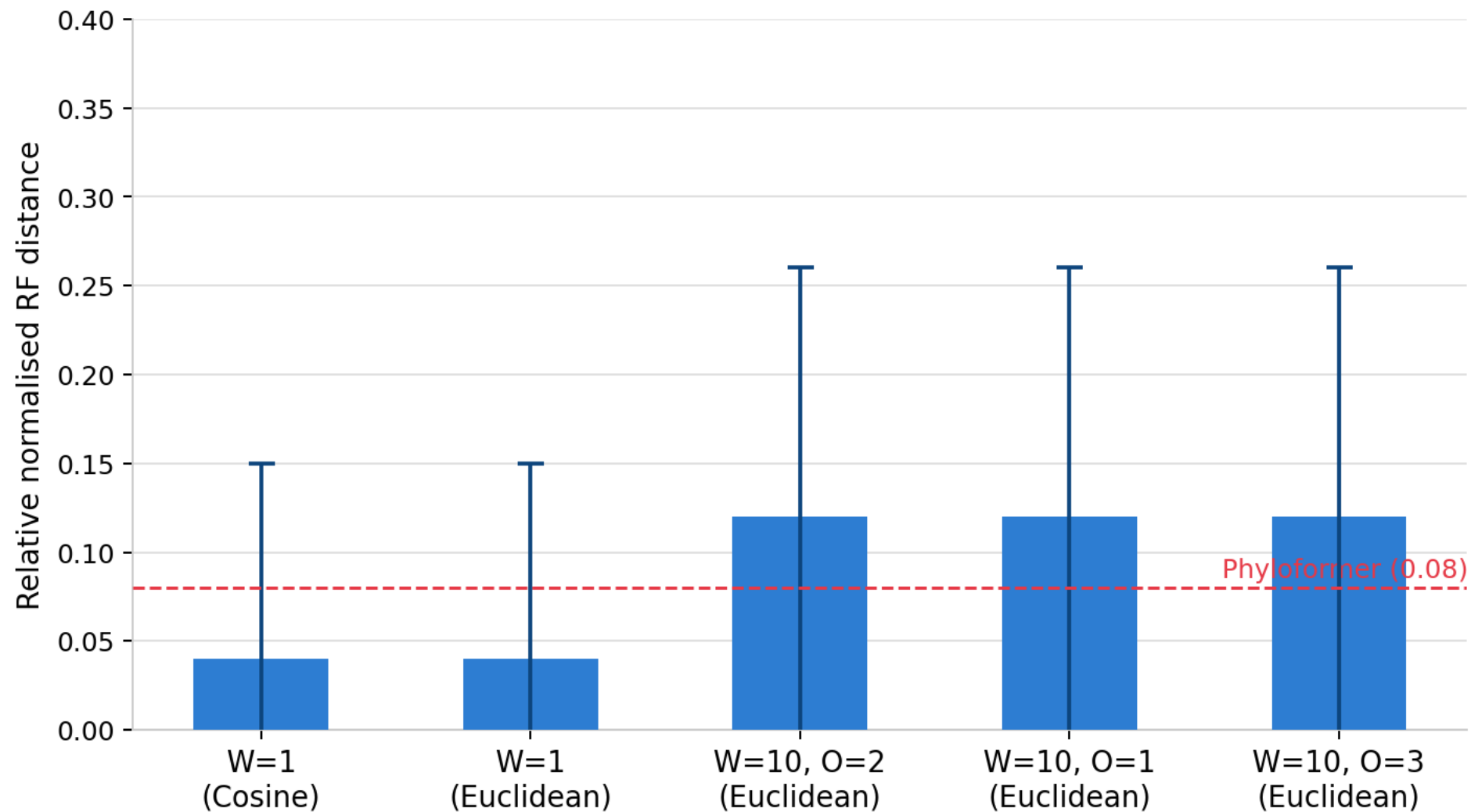
Priskiriami vienetiniai svoriai, arba svoriai sudaromi pagal unikalių simbolių skaičių arba Shannon entropiją

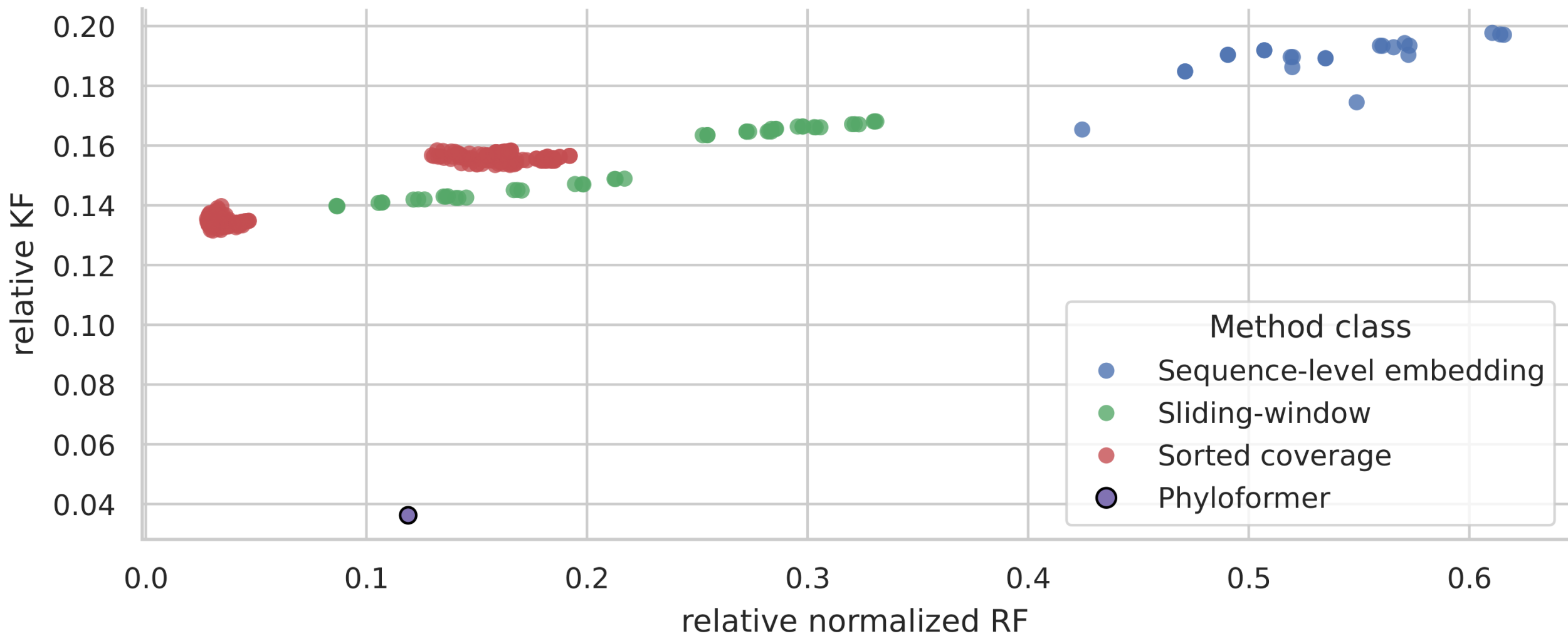


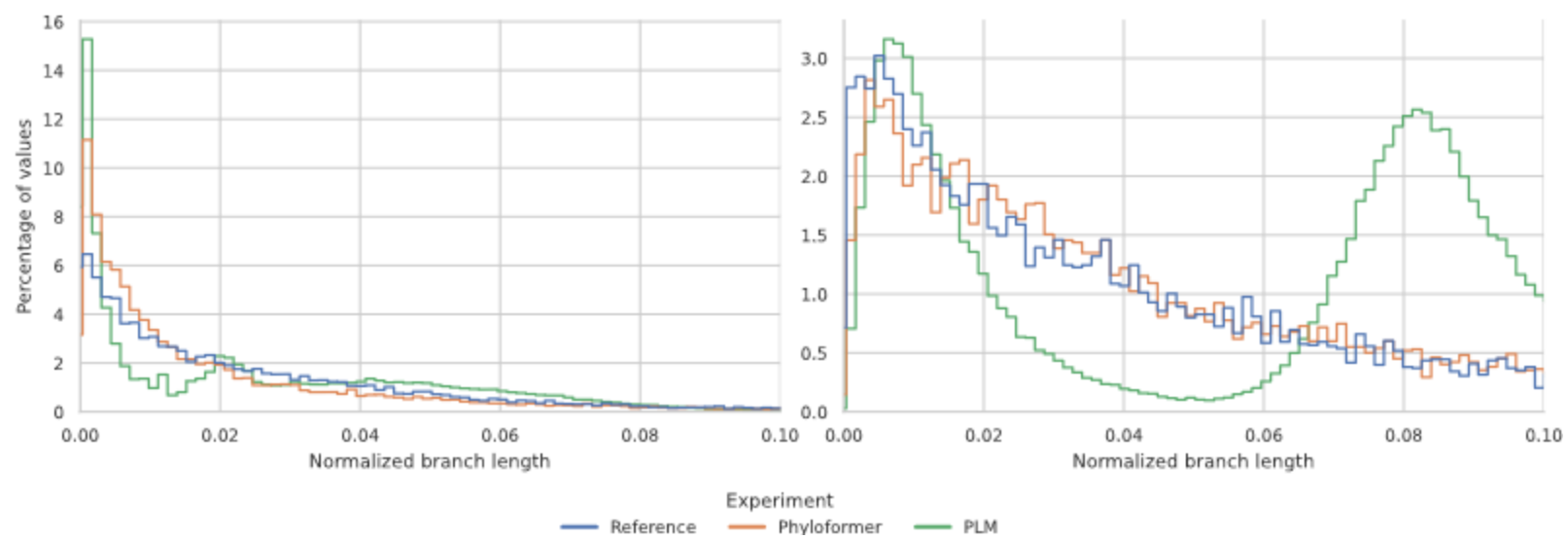
Slenkantis langas

- Sudaromi įterpiniai kiekvienai pozicijai
- Sudaromi persidengiantys langai
- Apskaičiuojamas atstumas kiekvinem langui
- Susumuojami visi gauti atstumai

Ekstremalus atvejis: $N=1$







Kito pusmečio darbo planas

Mokslinė veikla:

- Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:
 - Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas.
 - Analitinės disertacijos dalies parengimas.
 - Teorinės disertacijos dalies parengimas.
 - Eksperimentinės disertacijos dalies parengimas.
 - Bendrųjų išvadų formulavimas.
- Daktaro disertacijos parengimas ir svarstymas padalinyje.
- Daktaro disertacijos gynimas.