Hierarchical Classification of Transposable Elements

Felipe Kenji Nakano[1,2] and Ricardo Cerri[3]

[1] KU Leuven, Campus KULAK - Department of Public Health and Primary Care,
Etienne Sabbelaan 53, 8500 Kortrijk, Belgium
[2] ITEC - imec, Etienne Sabbelaan 51, 8500 Kortrijk, Belgium
[3] Federal University of Sao Carlos - Washington Luis, 13565-905 Sao Carlos, Brazil

Abstract. Transposable Elements (TEs) are DNA sequences capable moving within a genome. Such movement is associated with genetic variability and functionality in genes. TEs classification is usually performed using homology tools which can present drawbacks such as ignoring the hierarchical relationships among TEs. In this work, we have investigated hierarchical classification (HC) applied to TEs classification. We have also proposed two new methods, and experiments show that hierarchical classification can be superior than homology.

Keywords: Hierarchical Classification · Transposable Elements

1 Introduction

Transposable Elements (TEs) are DNA sequences capable of changing their location within a genome. Such movement promotes genetic variability and modifications in the functionality of genes. Moreover, TEs can also make species more resistant, or even be related to cancer [1]. Thus, the correct identification of TEs is essential to understand their influence in the evolution of species.

TEs classification is usually performed using homology methods. Homology consists of finding parts of sequences that are similar to pre-annotated sequences. They can present drawbacks, however, such as overlooking biochemical properties, being limited to sub-groups of TEs, and ignoring hierarchical relationships.

As a countermeasure, recent works have studied hierarchical classification (HC) methods [2, 3]. They build predictive models which can distinguish between any classes while incorporating hierarchical relationships. In this work, we provide a comparison between HC methods and the homology method BLASTn.

We also propose two new local methods: **n**on-**L**eaf **L**ocal **C**lassifier per **P**arent **N**ode (nLLCPN) and **L**ocal **C**lassifier per **P**arent **N**ode and **B**ranch (LCPNB). In a simple manner, LCPNB considers the mean of prediction probabilities, whereas nlLCPN follows the standard top-down approach [2, 3].

2   Experiments

We have collected data from public TE repositories (PGSB and REPBASE), and mapped them to Wicker's Taxonomy, using K-mers as features [3].

In Table 2, we present a comparison between our methods, local baselines from the literature (SWV, SimplePrune and LCPN), and the homology method BLASTn. Parameters detail were presented in [3]

Our method is capable of providing superior results in many cases. More specifically, nlLCPN and LCPNB can provide the highest hF in all datasets, while being competitive in the other evaluation measures. Additionally, all methods from machine learning were substantially superior to BLASTn in all cases.

		SWV	SimplePrune	LCPN	BLASTn	nlLCPN	LCPNB
	hP	0.84	0.84	0.84	0.80	**0.90**	**0.90**
PGSB	hR	**0.93**	**0.93**	**0.93**	0.83	0.91	0.91
	hF	0.88	0.88	0.88	0.81	0.90	**0.91**
	hP	0.82	0.82	0.82	0.55	**0.85**	0.85
REPBASE	hR	**0.89**	**0.89**	**0.89**	0.53	0.87	0.88
	hF	0.85	0.85	0.85	0.54	**0.86**	**0.86**
	hP	0.81	0.81	0.81	0.63	**0.85**	0.85
PGSB+REPBASE	hR	**0.89**	**0.89**	**0.89**	0.63	0.87	0.87
	hF	0.85	0.85	0.85	0.63	**0.86**	**0.86**

Table 1. mean hierarchical precision (hP), recall (hR) and f-measure (hF) measured in 10 runs

3   Conclusion

Since traditional methods were significantly outperformed, our results favour the applicability of machine learning in TEs classification. We have also proposed two methods which can be applied to other hierarchical problems. Future works might consider more datasets, and methods from the deep learning paradigm where features are learned automatically.

References

1. Biémont, C.: A brief history of the status of transposable elements: From junk dna to major players in evolution. Genetics **186**(4), 1085–1093 (2010)
2. Nakano, F.K., Mastelini, S.M., Barbon, S., Cerri, R.: Improving hierarchical classification of transposable elements using deep neural networks. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (July 2018)
3. Nakano, F.K., Pinto, W.J., Pappa, G.L., Cerri, R.: Top-down strategies for hierarchical classification of transposable elements with neural networks. In: 2017 International Joint Conference on Neural Networks (IJCNN) (May 2017)