

Prediction of malarial signal peptides using signalHsmm

Michał Burdukiewicz

September 30, 2016

1 Introduction

Medically significant *Plasmodiidae* have unique composition of signal peptides

Heavy adenine-thymine bias of malarial genomes alters amino acid composition of malarial signal peptides making them hard to predict using software trained on other eukaryotes.

2 Reduced alphabet

Group	Amino acids
I	D, E, H, K, N, Q, R
II	G, P, S, T, Y
III	F, I, L, M, V, W
IV	A, C

3 Benchmark

References

- Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: Signalp 3.0. *Journal of Molecular Biology*, 340(4):783 – 795.
- Hiller, K., Grote, A., Scheer, M., Münch, R., and Jahn, D. (2004). PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Research*, 32(suppl 2):W375–W379.
- Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338(5):1027–1036.

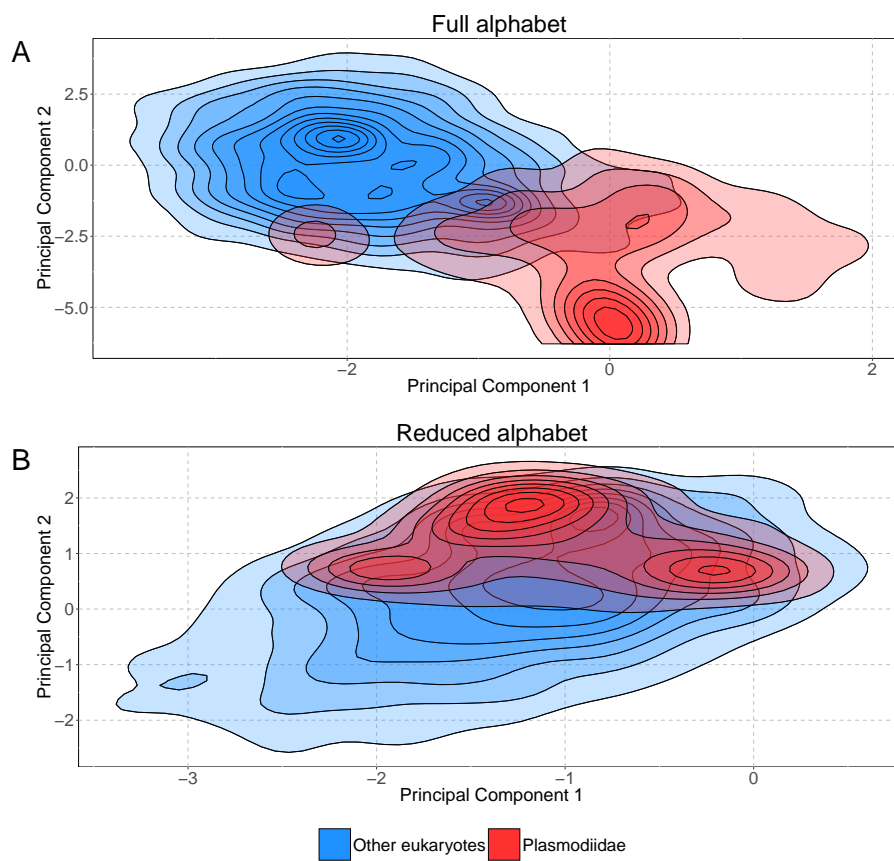


Figure 1: Principal component analysis of amino acid frequency in signal peptides belonging to *Plasmodiidae* and other eukaryotes. A) Frequency of amino acids. B) Frequency of amino acids encoded using the reduced alphabet.

	Sensitivity	Specificity	MCC	AUC
signalP 4.1 (no tm) (Petersen et al., 2011)	0.8235	0.9100	0.6872	0.8667
signalP 4.1 (tm) (Petersen et al., 2011)	0.6471	0.9431	0.6196	0.7951
signalP 3.0 (NN) (Bendtsen et al., 2004)	0.8824	0.9052	0.7220	0.8938
signalP 3.0 (HMM) (Bendtsen et al., 2004)	0.6275	0.9194	0.5553	0.7734
PrediSi (Hiller et al., 2004)	0.3333	0.9573	0.3849	0.6453
Philius (Reynolds et al., 2008)	0.6078	0.9336	0.5684	0.7707
Phobius (Käll et al., 2004)	0.6471	0.9289	0.5895	0.7880
signalHsmm-2010	0.9804	0.8720	0.7409	0.9262
signalHsmm-2010 (hom. 50%)	1.0000	0.8768	0.7621	0.9384
signalHsmm-2010 (raw aa)	0.8431	0.9005	0.6853	0.8718

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10):785–786.

Reynolds, S. M., Käll, L., Riffle, M. E., Bilmes, J. A., and Noble, W. S. (2008). Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *PLoS Computational Biology*, 4(11):e1000213.