

Signal peptides of Plasmodiidae

Michał Burdukiewicz

Outline

- 1 Apicomplexa signal peptides
- 2 Data
- 3 Lengths of proteins and signal peptides
- 4 1-gram frequency
- 5 Group frequency
- 6 Length of regions
- 7 Conclusion
- 8 References

Outline

- 1 Apicomplexa signal peptides
- 2 Data
- 3 Lengths of proteins and signal peptides
- 4 1-gram frequency
- 5 Group frequency
- 6 Length of regions
- 7 Conclusion
- 8 References

Comparison of Area Under the Curve, Sensitivity and Specificity for different classifiers.

Software name	AUC	Sensitivity	Specificity
signalP 4.1 (no tm) (Petersen et al., 2011)	0.9416	0.9720	0.9112
signalP 4.1 (tm) (Petersen et al., 2011)	0.9673	0.9579	0.9766
PrediSi (Hiller et al., 2004)	0.8949	0.9065	0.8832
Phobius (Käll et al., 2004)	0.9509	0.9673	0.9346
Philius (Reynolds et al., 2008)	0.9369	0.9533	0.9206
signalHsmm-2010	0.9526	0.9533	0.8832
signalHsmm-1989	0.9562	0.9626	0.8972

Comparison of Area Under the Curve and H-measure for different classifiers considering only proteins belonging to Plasmodiidae.

Software name	AUC	Sensitivity	Specificity
signalP 4.1 (no tm) (Petersen et al., 2011)	0.8356	0.7745	0.8966
signalP 4.1 (tm) (Petersen et al., 2011)	0.7928	0.6471	0.9385
PrediSi (Hiller et al., 2004)	0.6597	0.3725	0.9469
Phobius (Käll et al., 2004)	0.7963	0.6765	0.9162
Philius (Reynolds et al., 2008)	0.7753	0.6176	0.9330
signalHsmm-2010	0.9340	1.0000	0.8436
signalHsmm-1989	0.9326	0.9510	0.8631
signalHsmm-2010 with k-mers	0.9334	0.9902	0.7989

Outline

- 1 Apicomplexa signal peptides
- 2 **Data**
- 3 Lengths of proteins and signal peptides
- 4 1-gram frequency
- 5 Group frequency
- 6 Length of regions
- 7 Conclusion
- 8 References

Query

```
taxonomy:" Eukaryota [2759]" NOT taxonomy:" Plasmodium [5820]"  
annotation:(type:signal evidence:ECO_0000269) AND reviewed:yes
```

2543 proteins.

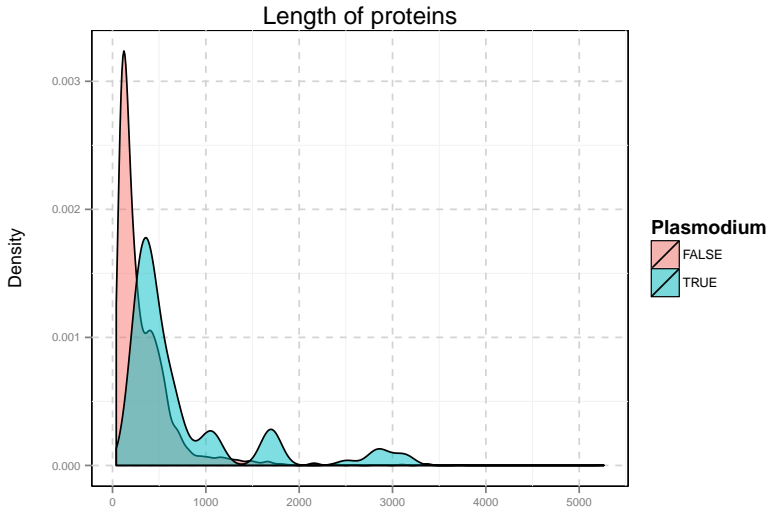
Query

taxonomy:"Plasmodium [5820]" annotation:(type:signal
evidence:manual) AND reviewed:yes

102 proteins.

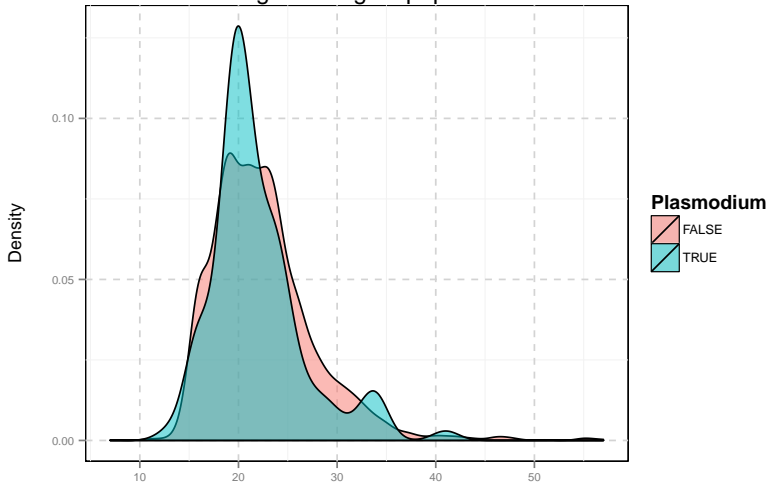
Outline

- 1 Apicomplexa signal peptides
- 2 Data
- 3 Lengths of proteins and signal peptides**
- 4 1-gram frequency
- 5 Group frequency
- 6 Length of regions
- 7 Conclusion
- 8 References



plasmodium	mean	sd	median	IQR
FALSE	327.5368	357.2294	216	291.00
TRUE	734.4804	711.3815	440	380.75

Lengths of signal peptides



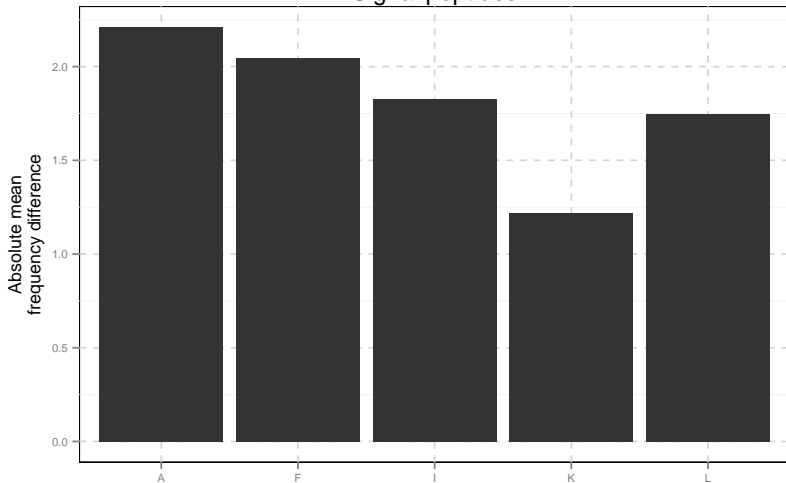
plasmodium	mean	sd	median	IQR
FALSE	22.48722	5.277110	22	6
TRUE	21.91176	4.741022	21	5

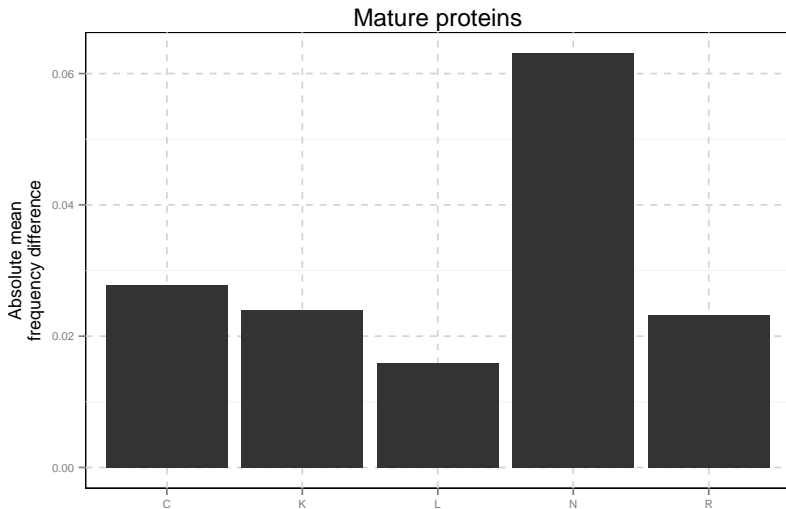
Outline

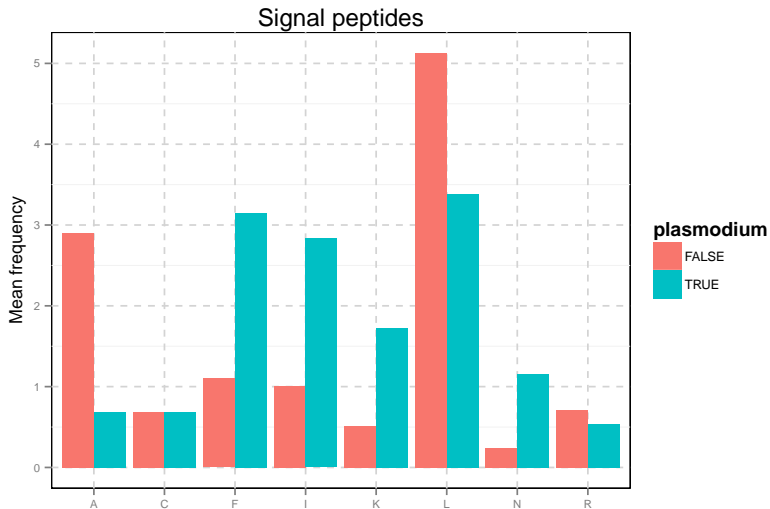
- 1 Apicomplexa signal peptides
- 2 Data
- 3 Lengths of proteins and signal peptides
- 4 1-gram frequency**
- 5 Group frequency
- 6 Length of regions
- 7 Conclusion
- 8 References

Since lengths of signal peptides seem to be comparable between proteins belonging to Plasmodium and other Eukaryotes, we chose not to normalize frequency of amino acids in signal peptides.

Signal peptides

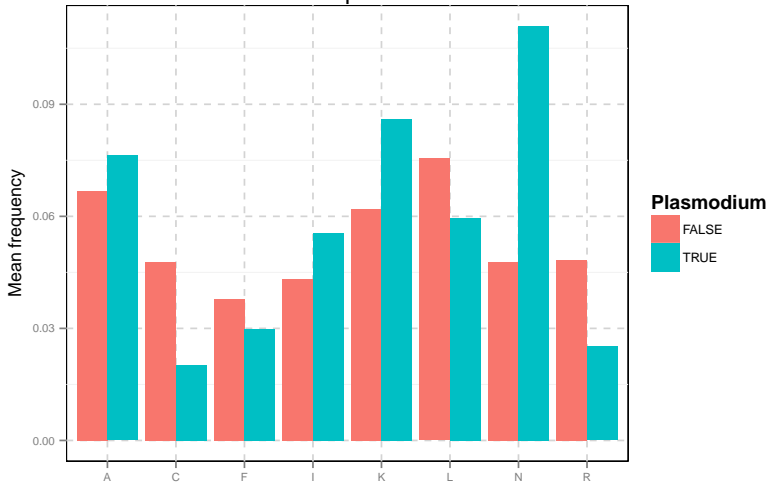






Plasmodium transit peptides are enriched in lysine and asparagine (Zuegge, Ralph, Schmuker, Mcfadden, & Schneider, 2001).

Mature proteins



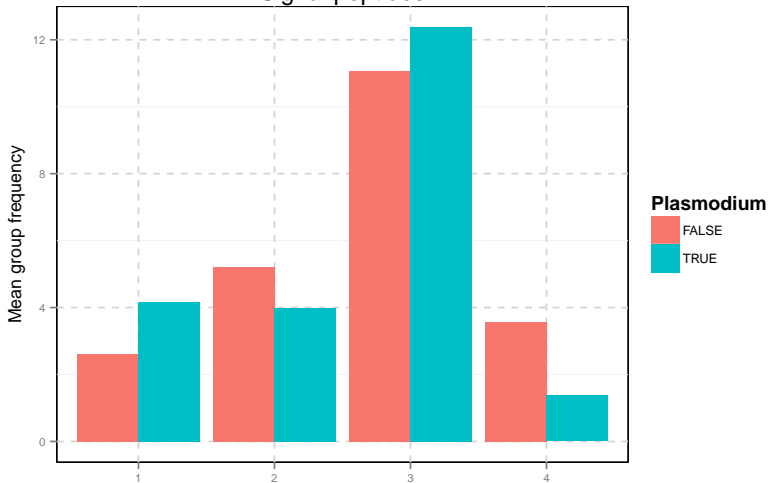
Outline

- 1 Apicomplexa signal peptides
- 2 Data
- 3 Lengths of proteins and signal peptides
- 4 1-gram frequency
- 5 Group frequency**
- 6 Length of regions
- 7 Conclusion
- 8 References

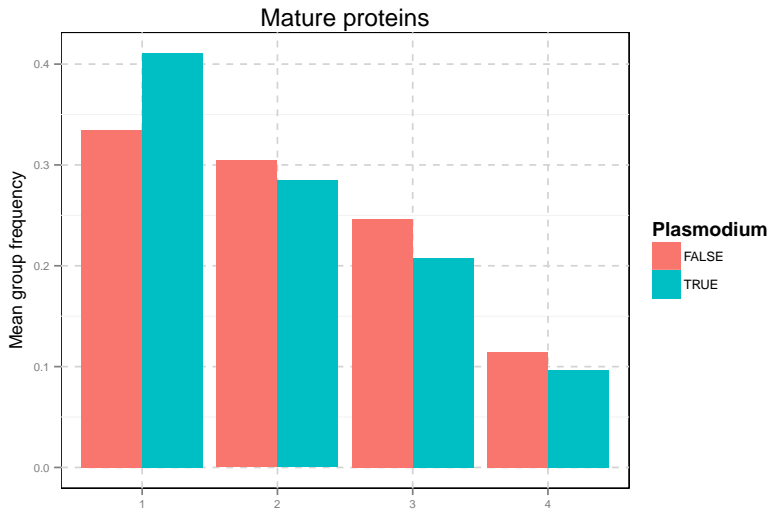
The signalHsmm encoding.

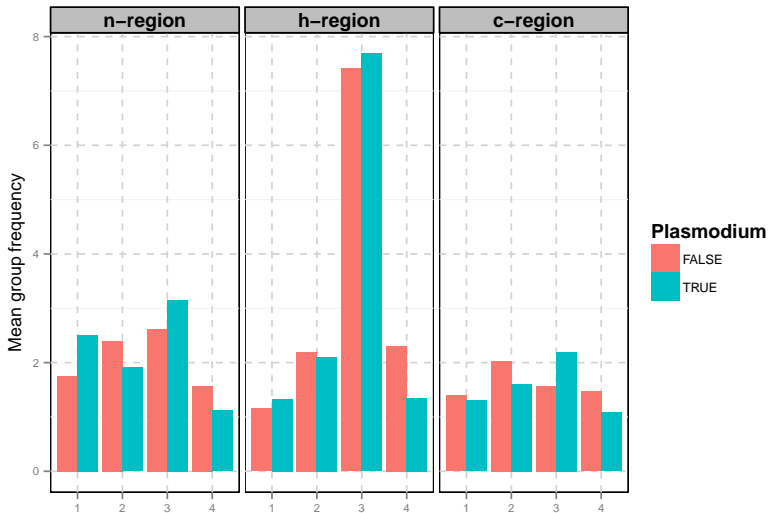
Group ID	Amino acids
1	D, E, H, K, N, Q, R
2	G, P, S, T, Y
3	F, I, L, M, V, W
4	A, C

Signal peptides



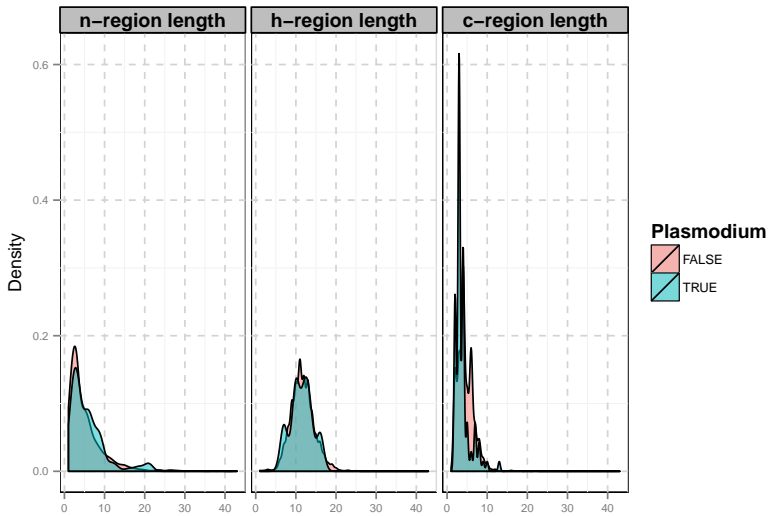
Group	Plasmodium	Mean frequency
1	FALSE	2.624852
1	TRUE	4.176471
2	FALSE	5.203696
2	TRUE	3.980392
3	FALSE	11.079040
3	TRUE	12.382353
4	FALSE	3.579630
4	TRUE	1.372549





Outline

- 1 Apicomplexa signal peptides
- 2 Data
- 3 Lengths of proteins and signal peptides
- 4 1-gram frequency
- 5 Group frequency
- 6 Length of regions**
- 7 Conclusion
- 8 References



plasmodium	variable	mean	sd	median	IQR
FALSE	n-region length	5.233582	4.551361	4.0	5
FALSE	h-region length	11.651986	2.943809	11.0	3
FALSE	c-region length	4.601652	2.037751	4.0	3
TRUE	n-region length	5.725490	4.396571	4.5	4
TRUE	h-region length	11.411765	2.765963	12.0	3
TRUE	c-region length	3.774510	1.882209	3.0	1

Outline

- 1 Apicomplexa signal peptides
- 2 Data
- 3 Lengths of proteins and signal peptides
- 4 1-gram frequency
- 5 Group frequency
- 6 Length of regions
- 7 Conclusion**
- 8 References

The mean frequency of group 3 (large, hydrophobic amino acids) are more common in the Plasmodium signal peptides.

Outline

- 1 Apicomplexa signal peptides
- 2 Data
- 3 Lengths of proteins and signal peptides
- 4 1-gram frequency
- 5 Group frequency
- 6 Length of regions
- 7 Conclusion
- 8 References**

- Hiller, K., Grote, A., Scheer, M., Münch, R., & Jahn, D. (2004). Predisi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Research*, 32, W375-W379.
- Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology*, 338, 1027-1036.
- Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). Signalp 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, 8, 785-786.
- Reynolds, S. M., Kall, L., Riffle, M. E., Bilmes, J. A., & Noble, W. S. (2008). Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *PLoS Comput Biol*, 4(11), e1000213.

Zuegge, J., Ralph, S., Schmuker, M., Mcfadden, G. I., & Schneider, G. (2001). Deciphering apicoplast targeting signals – feature extraction from nuclear- encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene*, 280, 19–26.