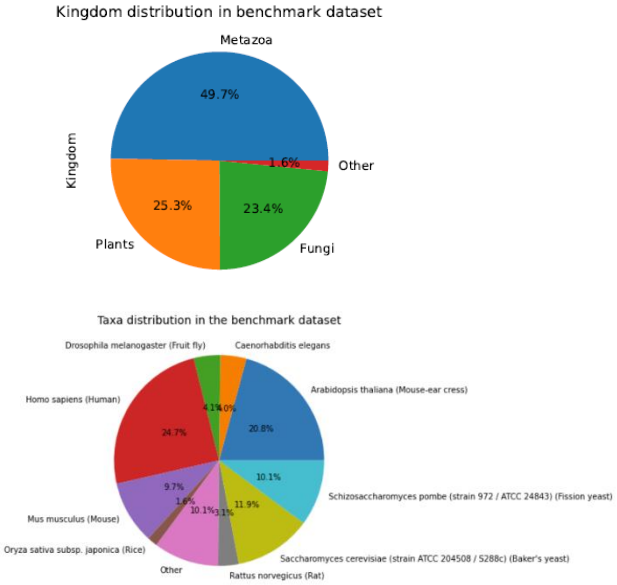
Signal Peptide prediction (Alberti Marta) – supplementary material

Suppl. Fig. 1. Kingdom and taxa distributions of benchmark SP



Prevalence of Metazoa in the kingdom distribution, similarly to the training kingdom distribution. Wider variety of taxa represented with respect to the training kingdom distribution.

Suppl. Table 1. vonHeijne 5-fold CV thresholds.

	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4	Average threshold
threshold	7.929	7.873	8.613	8.036	8.579	8.206
MCC	0.812	0.751	0.797	0.764	0.796	

The optimal threshold for each fold was estimated via a precision-recall curve at varying prediction threshold. The average threshold was used to predict the benchmark sequences.

Suppl. Table 2. CV scores for the all the combinations of C {1, 2, 4}, k {20, 22, 24}, γ {0.5, 1, 'scale'} for SVM. The best combination of parameters, according to the MCC is: C=2, k=20, γ='scale'.

Combo number	C	k	γ	MCC	F1-score	accuracy	recall	precision
1	1	20	0.5	0.779	0.803	0.947	0.733	0.895
2	1	20	1	0.803	0.827	0.952	0.776	0.891
3	1	20	scale	0.831	0.854	0.958	0.830	0.884
4	2	20	0.5	0.804	0.828	0.952	0.779	0.880
5	2	20	1	0.816	0.840	0.954	0.806	0.880
6	2	20	scale	0.840	0.863	0.959	0.853	0.876

7	4	20	0.5	0.811	0.836	0.952	0.803	0.876
8	4	20	1	0.831	0.854	0.957	0.841	0.872
9	4	20	scale	0.829	0.854	0.956	0.845	0.866
10	1	22	0.5	0.709	0.730	0.932	0.616	0.899
11	1	22	1	0.766	0.794	0.943	0.736	0.866
12	1	22	scale	0.791	0.821	0.947	0.806	0.840
13	2	22	0.5	0.763	0.791	0.942	0.729	0.869
14	2	22	1	0.770	0.803	0.943	0.779	0.830
15	2	22	scale	0.797	0.826	0.948	0.818	0.837
16	4	22	0.5	0.780	0.810	0.945	0.779	0.847
17	4	22	1	0.793	0.822	0.948	0.806	0.841
18	4	22	scale	0.788	0.818	0.946	0.810	0.831
19	1	24	0.5	0.646	0.650	0.919	0.508	0.919
20	1	24	1	0.712	0.743	0.932	0.667	0.844
21	1	24	scale	0.766	0.800	0.941	0.779	0.823
22	2	24	0.5	0.704	0.734	0.930	0.655	0.842
23	2	24	1	0.744	0.779	0.936	0.745	0.820
24	2	24	scale	0.773	0.805	0.943	0.795	0.820
25	4	24	0.5	0.742	0.775	0.937	0.733	0.828
26	4	24	1	0.752	0.786	0.938	0.760	0.818
27	4	24	scale	0.772	0.805	0.943	0.795	0.818