

Table 1 Evaluation of prediction accuracies of plant protein subcellular locations

Subcellular location	Dataset A (total 15028)					Dataset B (total 6908)				
	Total positives	Total negatives	Sn (%)	Sp (%)	MCC (%)	Total positives	Total negatives	Sn (%)	Sp (%)	MCC (%)
Secreted	1485	13543	91.1	98.7	88.5	263	6645	76.8	98.9	74.5
Mitochondrial	919	14109	65.2	82.6	28.4	402	6506	61.4	77.5	21.1
Chloroplast	8124	6904	27.5	90.9	23.5	4918	1990	28.2	90.7	20.4
ER	256	14772	22.3	100.0	46.0	87	6821	18.4	100.0	42.7
Cytosol	77	14951	61.0	78.9	7.0	23	6885	52.2	75.3	3.7
Golgi Apparatus	260	14768	1.5	99.9	6.3	54	6854	0.0	100.0	-0.2
Peroxisome	136	14892	24.3	99.7	31.6	52	6856	13.5	99.5	15.0
Nucleus	3099	11929	62.2	89.2	50.7	788	6120	68.8	85.5	42.7
Plasma Membrane	91	14937	35.2	95.1	10.7	14	6894	21.4	98.9	8.5
Vacuole	273	14755	5.1	99.0	5.5	121	6787	2.5	99.8	6.8
Cytoskeleton	305	14723	13.8	99.7	24.3	186	6722	21.0	99.7	36.0

Note: Sn: sensitivity; Sp: specificity; MCC: Mathews' correlation coefficient