

## Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report

John Meinken<sup>1</sup>, Xiang Jia Min<sup>2</sup>

1. Department of Computer Science and Information Systems, Youngstown State University, Youngstown, OH 44555, USA

2. Center for Applied Chemical Biology, Department of Biological Sciences, Youngstown State University, Youngstown, OH 44555, USA

✉ Corresponding Author email: [xmin@ysu.edu](mailto:xmin@ysu.edu); ✉ Author

Computational Molecular Biology, 2012, Vol.2, No.1 doi: 10.5376/cmb.2012.02.0001

**Copyright** © 2012 Meinken and Min. This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract** Computational prediction of protein subcellular locations in eukaryotes facilitates experimental design and proteome analysis. We provide a short review on recent development of computational tools and our experience in evaluating some of these tools. Classical secretomes can be relatively accurately predicted using computational tools to predict existence of a secretory signal peptide and to remove transmembrane proteins and endoplasmic reticulum (ER) proteins. The protocols of differentially combining SignalP, Phobius, WoLFPSORT, and TargetP for identifying a secretory signal peptide in different kingdom of eukaryotes, with TMHMM for removing transmembrane proteins and PS-Scan for removing ER proteins significantly improve the secretome prediction accuracies. Our evaluation showed that current computational tools for predicting other subcellular locations, including mitochondrial or chloroplast localization, still need to be improved.

**Keywords** Eukaryotes; Protein subcellular location; Secretome; Computational prediction

### Introduction

Eukaryotic cells have a complex endomembrane system, in addition to independent organelle structures such as mitochondria and chloroplasts. These subcellular structures include nucleus, endoplasmic reticulum (ER), Golgi apparatus, lysosome, peroxisome, vacuole, cytoskeleton, cytosol, mitochondrion, chloroplast, and plasma membrane. For a subcellular membrane enclosed structure, these structures consist of membrane and internal space such as ER lumen. Outside the plasma membrane, the cell wall and extracellular matrix and space are also important sites for cellular activities.

Eukaryotic cells synthesize thousands of different proteins. For example, *Saccharomyces cerevisiae*, commonly known as baker's yeast, with a relative small genome size of 12 Mb, encodes approximately 5000~6000 different proteins. The proteins encoded by a nuclear genome are synthesized on ribosomes in

cytosol or ribosomes attached to rough ER. However, these proteins need to be translocated to one or more than one specific subcellular location(s) in order to play their biological roles, a process called protein targeting or sorting. Experimental approaches for identifying protein subcellular locations are widely exploited including isolation of organelles, green fluorescence tagging proteins, etc (Heazlewood et al., 2005). Some signal targeting peptides determining protein subcellular locations have been identified experimentally (Blobel and Dobberstein, 1975). Thus protein targeting is believed to be determined by the protein's physical and chemical properties of targeting domains that could be identified from its amino acid sequence. A number of computational tools have been recently developed to predict the subcellular locations of eukaryotic proteins. Nakai and Horton (2007) comprehensively reviewed computation methods and tools for subcellular location

#### Preferred citation for this article:

Meinken and Min, 2012, Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report, Computational Molecular Biology, Vol.2, No.1 1-7 (doi: 10.5376/cmb.2012.02.0001)

Received: 02 Mar., 2012 | Accepted: 10Apr., 2012 | Published: 24 Apr., 2012

prediction. We provide a short review about more recent progress in this area and discuss the challenges remaining for future development based on our research experiences.

### Secretory Signal Peptide and Secretome Prediction

The term secretome is used to refer to a whole set of proteins that are secreted outside a cell-including cell wall, extracellular matrix and extracellular space-in an organism. Recently many efforts have been made to identify secretomes as these proteins have both potential applications in environmental industry and biomedicine (Lum and Min, 2011; Makridakis and Vlahou, 2010). For example, fungal secretomes often contain secreted extracellular enzymes to break down biopolymers that have potential applications in biofuel production (Lum and Min, 2011), and the human secretome plays important biological roles, such as insulin, and provides useful information for the discovery of novel biomarkers such as for cancer diagnosis (Makridakis and Vlahou, 2010).

A secretome consists of two types of proteins: classical secreted proteins and nonclassical secreted proteins. A typical classical secreted protein contains a secretory signal peptide located on its N-terminus and does not contain other targeting signals in its sequence (Emanuelsson et al., 2007). The secretory signal peptide directs the ribosome to the rough ER to complete the synthesis of a signal peptide containing protein. The secretory signal peptide, typically 15~30 amino acids long, is cleaved off during translocation across the membrane (von Heijne, 1990). One basic concept that should be made clear is that not all secretory signal peptide containing proteins are secreted. A number of papers reported predicted secretomes solely based on existence of secretory signal peptide prediction, resulting in an overestimation of the number of secreted proteins in a proteome. Secreted proteins make up only a fraction of the proteins that enter the ER secretory pathway, as proteins that contain a signal peptide and enter the ER also include residents of the rough ER, smooth ER, Golgi complex, lysosomes, endosomes, and plasma membrane.

Nowadays the commonly used tools for secretory signal peptide prediction include SignalP 3.0 (Bendtsen et al., 2004b), SignalP 4.0 (Petersen et al., 2011), Phobius (Käll et al., 2004 and 2007), TargetP (Emanuelsson et al., 2000), and PrediSi (<http://www.predisi.de/>) (Hiller et al., 2004). In addition, WoLFPSORT and MultiLoc2 can also be used for secreted protein prediction (Horton et al., 2007; Blum et al., 2009). The accuracy of SignalP 4.0 was improved over SignalP 3.0 with a higher specificity by intergrading transmembrane prediction (Petersen et al., 2011). However, SignalP 3.0 is more accurate than SignalP 4.0 in prediction of the cleavage site of the signal peptide. The default length of N-terminal peptide is 70 residues in both SignalP 3.0/4.0 and PrediSi and thus proteins sequences having a long signal peptide (>70 amino acids) cannot be predicted when using the default truncation parameter. Phobius was also a relatively accurate signal peptide predictor that combined transmembrane topology and signal peptide prediction.

Overall the accuracy for signal peptide prediction of the tools mentioned above was acceptable for general use. However, our recent evaluation of these tools showed that the accuracy for classical secretome prediction could be significantly improved by combining multiple tools, mainly due to an increase in prediction specificity (Min, 2010). In addition, adding TMHMM for removing transmembrane proteins and PS-Scan (a standalone version downloaded from Scan-Prosit) for removing ER resident proteins (Prosit: PS00014, Endoplasmic reticulum targeting sequence) significantly improved the accuracy for secretome prediction (Min, 2010). Our evaluation also showed different tools have different strength in processing protein data generated from different kingdoms of eukaryotic organisms. We proposed the following protocols for secretome prediction in different kingdoms of eukaryote: SignalP/WoLFPSORT/Phobius in fungi, Phobius/WoLFPSORT/TargetP in animals, SignalP/Phobius/TargetP in plants, and SignalP/Phobius/TargetP/WoLFPSORT in protists. The specificity for signal peptide prediction is significantly increased when two or more tools are used. In addition, TMHMM and

PS-Scan should be used for all eukaryotic secretome predictions (Min, 2010).

There is only one tool, SecretomeP, available for nonclassical secretome prediction (<http://www.cbs.dtu.dk/services/SecretomeP/>) (Bendtsen et al., 2004a) of mammalian and bacterial organisms. As about 50% of secreted proteins in plants were estimated to be non-classical, i.e. leaderless secreted proteins (LSPs) (Agrawal et al., 2010), certainly a plant specific trained tool or method is needed for prediction of plant specific non-classical secretomes.

### Predictors for Multiple Subcellular Locations

TargetP was implemented to predict subcellular locations of eukaryotic proteins by discriminating between chloroplast transit peptide (cTP, in plants), mitochondrial targeting peptide (mTP) and secretory pathway signal peptide (Emanuelsson et al., 2007). Except for fungal protein data sets, combining TargetP with SignalP, TMHMM and PS-Scan increased secretome prediction accuracy in other eukaryotic protein data sets (Min, 2010). Other widely used tools for predicting multiple subcellular locations are WoLFPSORT and MultiLoc2. WoLFPSORT predicts 12 subcellular locations including chloroplast, cytosol, cytoskeleton, ER, extracellular, Golgi apparatus, lysosome, mitochondria, nuclear, peroxisome, plasma membrane, and vacuolar membrane (Horton et al., 2007). MultiLoc2 predicts 9 subcellular locations for animals and fungi and 10 subcellular locations for plants (Blum et al., 2009). Chou and Shen (2008) developed a package of web servers, called Cell-PLoc, which includes 6 different servers for predicting up to 22 subcellular locations of proteins in various organisms including viruses, bacteria, plants, humans, or general eukaryotes (<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>). However, the servers in the Cell-PLoc package can only process a single sequence per submission, and no stand-alone tools are available, which prevented us for further evaluating the accuracies of these tools.

While developing the plant secretome knowledge-base (PlantSecKB), which is now publicly available (<http://proteomics.yzu.edu/secretomes/plant.php>), we

compared the prediction accuracies of TargetP, WoLFPSORT, and MultiLoc2 using a set of plant proteins retrieved from the UniProtKB Swiss-Prot data set. Proteins having multiple subcellular locations or labeled as “fragment”, or having a term of “by similarity” or “probable” or “predicted” in subcellular location annotation were removed. A total of 6908 proteins having annotated subcellular locations were selected. The results are shown in Table 1. If we ignore the subcellular location predictions with less than 100 positive entries, our evaluation showed prediction of secreted proteins was relatively more accurate than prediction of other subcellular locations by all three tools. TargetP was significantly more accurate than the other two tools in predicting secreted proteins. The Matthews' correlation coefficient (MCC) (Matthews, 1975) values for prediction of all other subcellular locations by all three tools were lower than 50%. Thus, an improvement in prediction accuracies for these subcellular locations of plant proteins is really needed. Overall prediction accuracies between WoLFPSORT and MultiLoc2 using its sequence-based prediction method did not show significant differences. MultiLoc2 incorporated phylogenetic profiles and Gene Ontology terms and was reported to perform considerably better than other methods for animal and plant proteins (Blum et al., 2009). However, its accuracy cannot be fairly tested as our data all had Gene Ontology annotation. In addition, we also found that MultiLoc2 was about 500 times slower in data processing than WoLFPSORT, which prevented us from using MultiLoc2 in data processing for our database development.

### Other Computational Tools

Table 2 lists a collection of subcellular localization prediction tools and their related publications. The weblinks for all these tools can be found at our webserver (<http://proteomics.yzu.edu/tools/subcell.html>). This is not an exhaustive list, but focuses on the tools discussed in this paper as well as more recent tools published since 2008. Our knowledgebases currently collect predictions from SignalP 3.0, SignalP 4.0, TMHMM, Phobius, TargetP, WoLFPSORT, PS-Scan and FragAnchor as discussed above.

Table 2 A collection of published protein subcellular localization prediction tools

Tool name	Locations or protein features predicted	Organism categories	Publication citation
TargetP	Extracellular, mitochondrial, chloroplast	Non-plant, plant	Emanuelsson et al., 2000
TMHMM	Transmembrane helices	Any	Krogh et al., 2001
ScanProsite (PS-Scan)	ER retention signal	Any	de Castro et al., 2006
SecretomeP 2.0	Non-classical i.e. not signal peptide triggered protein secretion	Gram-positive bacteria, gram-negative bacteria, mammal	Bendtsen et al., 2004a
Phobius	Signal peptide and transmembrane topology	Any	Käll et al., 2007
WoLF PSORT	> 5 locations	Fungi, plant, animal	Horton et al., 2007
PRED-LIPO	Lipoprotein signal peptides	Gram-positive bacteria	Bagos et al., 2008
ProLoc-GO	> 5 locations	Human and eukaryotes	Huang et al., 2008
KnowPredsite	> 5 locations	Prokaryotes and eukaryotes	Lin et al., 2009
MultiLoc2	> 5 locations	Animal, fungal, plant	Blum et al., 2009
PRED-SIGNAL	Signal peptides	Archaea	Bagos et al., 2009
RSLPred	Chloroplast, cytoplasm, mitochondria, nucleus	Oryza sativa only	Kaundal and Raghava, 2009
SherLoc2	> 5 locations	Animal, fungi, plant	Briesemeister et al., 2009
Cell-PLoc 2.0	> 5 locations	Eukaryote, human, plant, virus, gram-positive bacteria, gram-negative bacteria	Chou and Shen, 2010
CoBaltDB	> 5 locations	Prokaryotes	Goudenège et al., 2010
PSORTb	> 5 locations	Gram-positive and gram-negative bacteria	Yu et al., 2010
SCLPred	Cytoplasm, mitochondrion, nucleus, secretory, chloroplast	Animals, plants, fungi	Mooney et al., 2011
SignalP 4.0	Signal peptide	Eukaryotic, gram-positive bacteria, gram-negative bacteria	Petersen et al., 2011
SlocX	> 5 locations	Aribdopsis thaliana only	Ryngajllo et al., 2011

Some tools make predictions for only a single subcellular location or identify the presence of a single protein feature (such as a signal peptide). Then there are more comprehensive tools that can make predictions for many locations, and may employ a combination of multiple computational methods as well. The trend in recent years seems to be toward more comprehensive tools. Of the tools we collected that were published since 2008, twelve out of fifteen contain predictions for four or more subcellular localizations.

With the emergence of so many tools that can already predict a variety of subcellular locations, one might ask if our approach of combining analysis results from

multiple tools into a database is still relevant. We believe our work can make several valuable contributions in this area. Firstly, a combination of data from multiple predictions often produces more accurate results than the individual predictions. This principle has been demonstrated in our specific work with secretomes (Min, 2010) and is also a widely recognized statistical concept. Also, a database can be used in ways that a prediction tool cannot. For most of the prediction tools, analysis is performed at the time of request. The user must know which protein(s) they are interested in before they can get analysis results. With our database, the user can work in the other direction as well. They can start with a subcellular location and species they are interested in and get a list of proteins that meet those criteria.

In addition, the development of so many tools that can perform the same task creates a dilemma for researchers, who must choose which tool(s) they will use. There is a need for testing that compares different tools and identifies their relative strengths and weaknesses. Perhaps some tools perform better for plants while others perform better for bacteria. Some tools may have better specificity for a certain subcellular location while others may have better sensitivity. Our knowledgebases can serve as a rich dataset for performing such comparisons. In this work, we compare the prediction accuracies for plant proteins using TargetP, WoLFPSORT and MultiLoc2. Much more work is needed to continue these types of comparative studies for improving the prediction accuracy of proteome-wide protein subcellular location in the future.

### Acknowledgments

The work was supported by the Ohio Plant Biotechnology Consortium and Youngstown State University (YSU) Research Council to XJM.

### References

- Agrawal G.K., Jwa N.S., Lebrun M.H., Job D., and Rakwal R., 2010, Plant secretome: unlocking secrets of the secreted proteins, *Proteomics*, 10: 799-827  
<http://dx.doi.org/10.1002/pmic.200900514>  
 PMID:19953550
- Bagos P.G., Tsirigos K.D., Plessas S. K., Liakopoulos T. D., and Hamodrakas S. J., 2009, Prediction of signal peptides in archaea, *Protein engineering, design & selection: PEDS*, 22(1): 27-35  
 PMID:18988691
- Bagos P.G., Tsirigos K.D., Liakopoulos T.D., and Hamodrakas S.J., 2008, Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model, *J. proteome res.*, 7(12): 5082-5093  
<http://dx.doi.org/10.1021/pr800162c>  
 PMID:19367716
- Bendtsen J.D., Jensen L.J., Blom N., Von Heijne G., and Brunak S., 2004a, Feature based prediction of non-classical and leaderless protein secretion, *Protein Eng. Des. Sel.*, 17(4): 349-356  
<http://dx.doi.org/10.1093/protein/gzh037>  
 PMID:15115854
- Bendtsen J.D., Nielsen H., Von Heijne G., and Brunak S., 2004b, Improved prediction of signal peptides: SignalP 3.0, *J. Mol. Biol.*, 340: 783-795  
<http://dx.doi.org/10.1016/j.jmb.2004.05.028>  
 PMID:15223320
- Blobel G., and Dobberstein B., 1975, Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma, *J. Cell Biol.*, 67: 835-851  
<http://dx.doi.org/10.1083/jcb.67.3.835>  
 PMID:811671
- Blum T., Briesemeister S., and Kohlbacher O., 2009, MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction, *BMC Bioinformatics*, 10: 274  
<http://dx.doi.org/10.1186/1471-2105-10-274>  
 PMID:19723330 PMCID:PMC2745392
- Briesemeister S., Blum T., Brady S., Lam Y., Kohlbacher, O., and Shatkay H., 2009, SherLoc2: a high-accuracy hybrid method for predicting subcellular localization of proteins, *J. proteome res.*, 8(11): 5363-5366  
<http://dx.doi.org/10.1021/pr900665y>  
 PMID:19764776
- Chou K.C., and Shen H.B., 2008, Cell-PLoc: a package of web servers for predicting subcellular localization of proteins in various organisms, *Nat. protoc.*, 3(2): 153-162  
<http://dx.doi.org/10.1038/nprot.2007.494>  
 PMID:18274516
- Chou K., and Shen H., 2010, Cell-PLoc 2.0: an improved package of web-servers for predicting subcellular localization of proteins in various organisms, *Natural Science*, 2: 1090-1103, doi: 10.4236/ns.2010.210136  
<http://dx.doi.org/10.4236/ns.2010.210136>
- de Castro E., Sigrist C.J., Gattiker A., Bulliard V., Langendijk-Genevaux P.S., Gasteiger E., Bairoch A., and Hulo N., 2006, ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins, *Nucleic Acids Res.*, 34(Web Server issue): W362-365
- Emanuelsson O., Brunak S., Von Heijne G., and Nielsen H., 2007, Locating proteins in the cell using TargetP, SignalP and related tools, *Nat. Protoc.*, 2: 953-971  
<http://dx.doi.org/10.1038/nprot.2007.131>  
 PMID:17446895
- Emanuelsson O., Nielsen H., Brunak S., and Von Heijne G., 2000, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, *J. mol. Biol.*, 300(4): 1005-1016  
<http://dx.doi.org/10.1006/jmbi.2000.3903>  
 PMID:10891285
- Gouden è ge D., Avner S., Lucchetti-Miganeh C., and Barloy-Hubler F., 2010, CoBaltDB: Complete bacterial and archaeal orfeomes subcellular localization database and associated resources, *BMC microbiol.*, 10: 88



- <http://dx.doi.org/10.1186/1471-2180-10-88>  
PMid:20331850 PMCID:PMC2850352
- Heazlewood J.L., Tonti-Filippini J., Verboom R.E., and Millar A.H., 2005, Combining experimental and predicted datasets for determination of the subcellular location of proteins in Arabidopsis, *Plant Physiol.*, 139(2): 598-609  
<http://dx.doi.org/10.1104/pp.105.065532>  
PMid:16219920 PMCID:PMC1255979
- Hiller K., Grote A., Scheer M., Münch R., and Jahn D., 2004, PrediSi: prediction of signal peptides and their cleavage positions, *Nucleic Acids Res.*, 32(Web Server issue): W375-379
- Horton P., Park K.J., Obayashi T., Fujita N., Harada H., Adams-Collier C.J., and Nakai K., 2007, WoLF PSORT: protein localization predictor, *Nucleic acids res.*, 35(Web Server issue): W585-587
- Huang W.L., Tung C.W., Ho S.W., Hwang S.-F., and Ho S.Y., 2008, ProLoc-GO: utilizing informative Gene Ontology terms for sequence-based prediction of protein subcellular localization, *BMC bioinformatics*, 9: 80  
<http://dx.doi.org/10.1186/1471-2105-9-80>  
PMid:18241343 PMCID:PMC2262056
- Käll L., Krogh A., and Sonnhammer E.L., 2004, A combined transmembrane topology and signal peptide prediction method, *J. Mol. Biol.*, 338: 1027-1036  
<http://dx.doi.org/10.1016/j.jmb.2004.03.016>  
PMid:15111065
- Käll L., Krogh A., and Sonnhammer E.L.L., 2007, Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server, *Nucleic acids res.*, 35(Web Server issue): W429-432
- Kaundal R., and Raghava G.P.S., 2009, RSLpred: an integrative system for predicting subcellular localization of rice proteins combining compositional and evolutionary information, *Proteomics*, 9(9): 2324-2342  
<http://dx.doi.org/10.1002/pmic.200700597>  
PMid:19402042
- Krogh A., Larsson B., von Heijne G., and Sonnhammer E.L., 2001, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. mol. Biol.*, 305(3): 567-580  
<http://dx.doi.org/10.1006/jmbi.2000.4315>  
PMid:11152613
- Lin H.N., Chen C.T., Sung T.Y., Ho S.Y., and Hsu W.L., 2009, Protein subcellular localization prediction of eukaryotes using a knowledge-based approach, *BMC bioinformatics*, 10(Suppl 15): S8  
<http://dx.doi.org/10.1186/1471-2105-10-S15-S8>
- Lum G., and Min X.J., 2011, FunSecKB: the fungal secretome knowledgebase, *Database-the Journal of Biological Databases and Curation*, Vol. 2011, doi: 10.1093/database/bar001  
<http://dx.doi.org/10.1093/database/bar001>
- Makridakis M., and Vlahou A., 2010, Secretome proteomics for discovery of cancer biomarkers, *J. proteomics*, 73(12): 2291-2305  
<http://dx.doi.org/10.1016/j.jprot.2010.07.001>  
PMid:20637910
- Matthews B.W., 1975, Comparison of the predicted and observed secondary structure of T4 phage lysozyme, *Biochim. Biophys. Acta*, 405: 442-451  
[http://dx.doi.org/10.1016/0005-2795\(75\)90109-9](http://dx.doi.org/10.1016/0005-2795(75)90109-9)
- Min X.J., 2010, Evaluation of computational methods for secreted protein prediction in different eukaryotes, *J. Proteomics Bioinform*, 3: 143-147
- Mooney C., Wang Y.H., and Pollastri G., 2011, SCLpred: protein subcellular localization prediction by N-to-1 neural networks, *Bioinformatics*, 27(20): 2812-2819  
<http://dx.doi.org/10.1093/bioinformatics/btr494>  
PMid:21873639
- Nakai K., and Horton P., 2007, Computational prediction of subcellular localization, *Methods Mol. Biol.*, 390: 429-466  
[http://dx.doi.org/10.1007/978-1-59745-466-7\\_29](http://dx.doi.org/10.1007/978-1-59745-466-7_29)  
PMid:17951705
- 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  
<http://dx.doi.org/10.1038/nmeth.1701>  
PMid:21959131
- Ryngajillo M., Childs L., Lohse M., Giorgi F.M., Lude A., Selbig J., and Usadel B., 2011, SLocX: predicting subcellular localization of Arabidopsis proteins leveraging gene expression data, *Frontiers plant sci.*, 2: 43
- Von Heijne G., 1990, The signal peptide, *J. Membr. Biol.*, 115: 195-201  
<http://dx.doi.org/10.1007/BF01868635>  
PMid:2197415
- Yu N.Y., Wagner J.R., Laird M.R., Melli G., Rey S., Lo R., Dao P., Sahinalp S.C., Ester M., Foster L.J., and Brinkman F.S.L., 2010, PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, *Bioinformatics*, 26(13): 1608-1615  
<http://dx.doi.org/10.1093/bioinformatics/btq249>  
PMid:20472543 PMCID:PMC2887053

# Computational Molecular Biology

2012, Vol.2, No.1, 1-7 <http://cmb.sophiapublisher.com>



Table 1 Comparison of prediction accuracies of plant protein subcellular locations by different tools

	Total positives	Total negatives	TargetP			WoLF PSORT			MultiLoc2		
			Sn (%)	Sp (%)	MCC (%)	Sn (%)	Sp (%)	MCC (%)	Sn (%)	Sp (%)	MCC (%)
Secreted	263	6645	76.8	98.7	72.3	43.7	99.6	58.0	51.3	98.6	53.7
Mitochondrial	402	6506	61.4	77.5	21.1	33.3	96.2	30.4	60.9	84.2	27.3
Chloroplast	4918	1990	28.2	90.7	20.4	28.2	83.9	12.7	19.2	98.6	23.2
ER	87	6821	-	-	-	9.2	99.9	24.5	29.9	98.4	23.0
Cytosol	23	6885	-	-	-	52.2	65.6	2.2	95.7	51.5	5.4
Golgi Apparatus	54	6854	-	-	-	0.0	99.9	-0.2	24.1	98.5	15.5
Peroxisome	52	6856	-	-	-	15.4	99.3	14.3	42.3	98.1	23.9
Nucleus	788	6120	-	-	-	80.3	78.1	41.1	26.8	93.3	22.4
Plasma Membrane	14	6894	-	-	-	21.4	97.9	6.1	7.1	99.4	3.9
Vacuole	121	6787	-	-	-	14.0	99.5	21.2	5.0	99.8	11.6

Note: Sn: Sensitivity; Sp: Specificity; MCC: Mathews' correlation coefficient