# CITRUS ADVANCED TECHNOLOGY PROGRAM 

QUARTERLY \& FINAL REPORTS: Control of Citrus Greening, Canker \& Emerging Diseases of Citrus

## INSTRUCTIONS

$\bigcirc$Annual Report $\square$ Final

Required: What is the "headline" for this report (e.g. a one-sentence "newspaper headline" describing what you accomplished)

# Prediction of protein subcellular localization is completed 

Proposal Title
Insight into the causative agent of citrus greening disease (HLB) using computational structure/function analysis of genome encoded proteins.

| Today's Date | Sponsoring Organization (drop-down) | Category (drop down) |
| :---: | :---: | :---: |
| 9/30/10 | Citrus Research and Development Foundation | Unclassified |


#### Abstract

REPORT UPDATE ( 500 words; summarize your accomplishments) Exported proteins are expected to be important for virulence and their comprehensive analysis should aid our understanding of the disease. We combined the results of computer programs and manual curation to identify potential transmembrane and extracytoplasmic proteins. We applied 6 TMH predictors (TMHMM, HMMTOP, TOPPRED, MEMSAT, MEMSAT_SVM and Phobius) and two of them (MEMSAT_SVM and Phobius) detect SPs that are likely to be processed by the Sec complex. In addition, we used the well established SignalP3.0, which contains both Hidden Markov Model (SignaIP_HMM) and Artificial Neural Network (SignaIP_NN) modes for SP prediction. These automatic methods are generally based on the local properties of protein sequences or sequence profiles, resulting in a considerable rate of false predictions. Consequently, we manually inspected all the proteins that are predicted to have TMHs or SPs by any automatic predictors we applied. This broad inclusion can help lower the false negative rate. At the same time, to control the false positive rate, we integrated several lines of evidence, including consensus between predictors, predicted 3D structure (to rule out buried hydrophobic segments in known cytoplasmic proteins) and function (to identify proteins and protein domains known to function outside the cytoplasm or in the membrane), features of a protein's orthologs (to validate if the SPs and TMHs can be constantly predicted in a orthologous group) and specific information about secretion machineries of Ca . L. asiaticus.

After careful analysis of 218 proteins that were predicted to have SPs by any automatic method to identify extracytoplasmic proteins, we hypothesize that 86 proteins with predicted SPs are likely secreted from cytoplasm to periplasm via the Sec machinery. Many proteins from the initial list of 218 candidates were excluded due to the following reasons: (1) the SP cannot be consistently predicted (predicted by only 1 out of 4 methods); (2) the protein is predicted to have multiple TMHs, such as the sensory box/GGDEF family protein (locus: CLIBASIA_01765; gi: 254780468); (3) the confidently predicted function of the protein suggests its localization in the inner membrane or cytoplasm, for example, the ribosomal protein L35, which is predicted to have a SP by 3 out of 4 predictors applied; (4) close homologs likely lack SPs. The resulting list available here: http://prodata.swmed.edu/congqian/paper/supplement_table_S1.pdf will be used to mine for potential virulence factors.


We also hypothesize that 21 proteins without SPs are likely targeted to the extracytoplasmic space through Sec-independent mechanisms. In addition, 184 proteins likely locate in the inner membrane of this Gram-negative bacterium. The lists of these proteins are here: http://prodata.swmed.edu/congqian/paper/supplement_table_S2.pdf

## PI First Name Nick

PI Last Name Grishin
Email grishin@chop.swmed.edu
Phone 214-645-5952

Organization UT Southwestern Medical Center
Contract Number 10-018-314
Project Duration (years) $1 \quad$ Year of Project 1
Total Funds (current year) \$99,902.00

