

Substrate Prediction of Membrane Transporter in *Arabidopsis thaliana*

Nadine Schaadt, Volkhard Helms

Chair of Computational Biology, Saarland University, Saarbrücken

Membrane transporters catalyze active one- and bi-directional transport of small chemicals across biological barriers such as lipid bilayer membranes. Based on their genome sequences, it has been postulated that model organisms such as *Escherichia coli* or *Arabidopsis thaliana* contain between 500 and 700 membrane transporters among their 4000 (*E.coli*) and 25.000 (*Arabidopsis*) protein coding genes. Unfortunately, the experimental annotation of which transporters transport which substrates is far from complete and will likely remain so for much longer. Therefore, it is highly desirable to develop statistical methods that may aid in the substrate annotation of putative membrane transporter proteins.

Here, we have present an analysis of the dataset on the *Arabidopsis* transporters compiled in the database Aramemnon (<http://aramemnon.botanik.uni-koeln.de/>). Various measures for the similarity of membrane transporter sequences were tested based on their amino acid composition, higher sequence order informations, amino acid characteristics or sequence conservation. We defined several positive sets including amino acid, oligopeptide, ion and sugar transporters and trained two classification schemes either based on ranked lists or on a SVM for classifying unknown sequences into these positive sets. We found that the amino acid profile allows ranking transporter sequences with an accuracy of 75% or higher according to the substrate type. Integrating additional information further improved the prediction performance. This study shows that substrate annotations of membrane transporters are well feasible based on their bare amino acid compositions. We hope that this work may prove useful for biological applications.