Sequence based prediction of Type III secreted proteins

Fred Kleine, Stefan Brandmaier, Patrick Tischler, Roland Arnold and Thomas Rattei

Many Gram-negative symbiotic bacteria possess a complex protein secretion system termed “Type III secretion system (TTSS)” to deliver bacterial effector proteins into host cells in order to modulate their cellular functions. Unlike the highly conserved TTSS, the currently known effector proteins lack sequence homology and are expected to be unique for each bacterial species. The mechanism of protein recognition and translocation by the TTSS is widely unknown and different theories have been proposed.

Genome wide screens for effector proteins in *Chlamydiae* and *Pseudomonas syringae* suggest the existence of a N-terminal signal that is necessary for Type III secretion. Based on this experimental evidence, we have developed a generalized prediction tool for the sequence based identification of Type III secreted proteins. For the N-terminal sequence of a query protein, the frequency of amino acids, dipeptides, tripeptides and the presence of small sequence patterns are classified by a support vector machine, which results in the prediction of Type III secretion for this protein. The positive datasets for training and testing consist of proteins from *Chlamydiae, Yersinia, Salmonella* and *E. Coli*, for which Type III secretion has been previously reported. In the absence of a sufficient number of known non-secreted proteins, randomly chosen proteins from the respective genomes have been used as negative datasets.

The quality of prediction has been tested by 10-fold cross-validation, which resulted in ~70% specificity and ~70% sensitivity. By using two genomes for training and the remaining two genomes for testing, we could show the general applicability of our method even across different phyla. This result could be confirmed by an additional test using proteins from *Pseudomonas syringae*.

The sequence based prediction of Type III secreted proteins is a very fast method to select candidate proteins for experimental secretion screens and provides evidence for the existence of a universal secretion signal in the N-terminus of Type III secreted effector proteins.