

Tips and tricks of the trade

The BaCellO Predictor

The lab tips and protocols on the *Nature Protocols* website are usually lengthy and bone-dry, only comprehensible by the experts. Recently, however, the *Lab Times* editor came across a notable exception to the rule.

Lab Hint

While surfing through the *Nature Protocols* site, my attention was caught by one protocol entitled “BaCellO: a Balanced Subcellular Localization predictor”, published by Andrea Pierleoni *et al.* (DOI:10.1038/nprot.2007.165), from Rita Casadios Biocomputing Group at the University of Bologna. Bologna is, obviously, not only home to great tortellini manufacturers but also to excellent bioinformaticians. As the title of the paper implies, the group presents an *in-silico* method for the prediction of the subcellular localization of a given protein. There is, however, more to it than meets the eye. They also provide free online access to the BaCellO prediction server, located on the bicomputing group’s website.

Pierleoni *et al.* keep their protocol as simple and straightforward as possible. My favourites are the “Materials”, “Time Taken” and “Procedure” sections. They read as follows:

“Materials:

- ▶ Sequences of the proteins to be predicted are required in FASTA format.
- ▶ A personal computer with a web browser program (Internet Explorer 6 and upper, Firefox and Opera 8 and upper were tested and support the prediction server).
- ▶ An internet connection

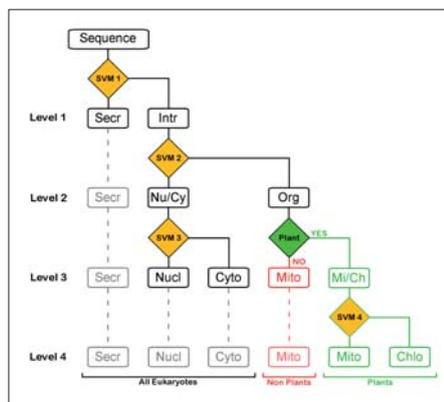


Figure 1: Decision tree of the BaCellO predictor (taken from *Bioinformatics*, 2006: 22, 408-416).



Developed the BaCellO predictor:

Andrea Pierleoni, Pier Luigi Martelli, Piero Fariselli and Rita Casadio

Time Taken:

- ▶ Approximatively 30 seconds per protein sequence.

Procedure:

How to predict the subcellular localization for a protein:

- ▶ Go to <http://gpcr.biocomp.unibo.it/bacello/pred.htm>
- ▶ Select the kingdom of the organism expressing your protein(s) (choosing between Animals, Fungi or Plants).
- ▶ Paste the sequences (up to five sequences per time) in the corresponding field.
- ▶ Submit the request and wait for results.

How to read results:

- ▶ The result page will be available for a maximum of 24h.
- ▶ In the result page you will find, for each protein:
 - a) the prediction of the subcellular localization
 - b) the path along the decision tree (Figure 1)."

Quite simple, isn't it? So easy that even the *Lab Times* editor is able to check, whether the BaCellO predictor really does its job. As a test protein I chose the neutral trehalase (Nth1) of bakers' yeast, which I had worked with during my PhD thesis. I went to the website of the *Saccharomyces* Genome Database (SGD) to get the Nth1 protein sequence in FASTA-format and pasted it into the entry mask of the BaCellO predictor. The result appeared a few seconds later: “Localization: Cytoplasm”, I read on the screen. That was correct. Might be a fluke, I thought to myself, and fed BaCellO with the FASTA-data of bakers' yeast acid trehalase (Ath1). Again, BaCellO stated: “Cyto-

plasm”. That might also be true, however, even yeast experts are not yet sure whether Ath1 is localised on the cell surface or in the cytosol.

According to the numbers Pierleoni *et al.* present under “Anticipated results”, the performance of BaCellO is fairly impressive. It depends on the organism and the number of compartments.

BaCellO performed best when it had to choose between intracellular and extracellular localisation. This equals level 1 of BaCellO's decision tree (see Fig. 1, taken from *Bioinformatics*, 2006: 22, 408-416), whose algorithm and tree architecture is at the heart of the localisation predictions. At this level, BaCellO correctly discriminates 96%, 91% and 96% of the proteins from fungi, animals and plants, respectively. The performance at level 2 is not so good (between 84% and 89% accuracy). Here, BaCellO has to decide between extracellular, nuclear/cytoplasmic and mitochondrial/chloroplast. Finally, at level 4, BaCellO has to differentiate between the five compartments (classes): nucleus, cytoplasm, extracellular, mitochondrion and chloroplast. It's no wonder that the scores then drop to 66%.

According to Pierleoni *et al.*, however, “BaCellO outperforms the other presently available methods for the same task and gives more balanced accuracy and coverage values for each class.”

HARALD ZÄHRINGER

Do you have any useful tips?

Contact us at:

editors@lab-times.org