



## **FERRAMENTA PARA RESTRIÇÃO DO BANCO DE DADOS PARA IDENTIFICAÇÃO DE PROTEÍNAS, BASEADO NO pI E MW PREVISTOS DE SEQUÊNCIAS GENÔMICAS**

Daniel Martins de Souza (Bolsista FAPESP), Marcus Bustamante Smolka, Flavia Vischi Winck, Carlos Eduardo Santoro Pedroso Filho, Prof. Dr. Sérgio Marangoni e Prof. Dr. José Camillo Novello (Orientador), Instituto de Biologia - IB, UNICAMP

A análise do proteoma de um organismo visa revelar quais genes estão sendo expressos em proteínas e como esta expressão muda em diferentes condições. Utiliza-se a técnica de eletroforese 2D para separação das proteínas, que gera dados como ponto isoelétrico (pI) e peso molecular (MW). A identificação é feita por sequenciamento N-terminal (N-term) ou espectrometria de massa (MS). Para organismos com genoma completamente sequenciado, alguns dados da proteína podem identificá-la. Porém alguns genes não podem ser identificados somente com os dados obtidos por N-term ou MS (dados ambíguos). Um procedimento de restrição do banco de dados baseado no cálculo teórico de pI e MW foi testado no proteoma da *Xylella fastidiosa* para identificação exata da proteína. Primeiramente, o pI/MW teórico de todos os genes da *Xylella* foram calculados. Foi então construído um mapa 2D teórico, (um gráfico Cartesiano de pontos onde o pI está no eixo X e MW no eixo Y) que representaria a migração teórica de cada gene, que foi sobreposto ao mapa obtido experimentalmente, demarcando uma região que seleciona candidatos ao gene de interesse. Com este procedimento foi possível a identificação do gene GROEL.

Proteoma - *Xylella fastidiosa* - Banco de Dados