1 Introduction

Pathogenesis studies of infectious diseases as well as the development of their effective medical treatments will depend greatly on the ability of researchers to decipher the function and network of proteins. With the fruitful achievement of several model organism genome projects started from the last decade of the 20th century, approaches of ‘functional genomics’ are emerging to comprehensively annotate gene functions, bringing the biological research society into the era of proteomics (as of protein functions in a genome-wide sense). We propose to take advantage of such advances to further understand the pathology of Candida albicans related diseases and to explore drug targets for effective medical treatments.

Candida albicans is both a commensal and pathogen of humans that can infect a broad range of body sites. Endogenous C. albicans infections are established by cells that normally colonise mucosal surfaces or skin as harmless commensals, and that are triggered to cause infection by changes in the host immune system or microflora. Life-threatening Candida infections are frequently seen in transplant recipients and have been reported in cancer patients at autopsy. In these severe cases, C. albicans penetrates into deeper tissue and may enter the bloodstream. From the bloodstream, it has the potential to invade almost all body sites and organs, causing life-threatening systemic infection that requires adaptation to a variety of different environmental stresses. The limited number of effective and safe antifungal antibiotics exacerbates these problems. Given the serious nature of diseases caused by the organism and the role of proteomics in our understanding of bacterial pathogenesis, development of the proteomics of C. albicans is an important endeavor.

2 Methods and Results

Database of integrated protein sequence information is established with so-called LAMP system (Linux, Apache, MySQL, and PHP). In this project, we integrate CDSPD and biological annotations, including GenBank, Gene Ontology (http://www.geneontology.org/), KEGG (http://www.genome.jp/kegg/), interpro (http://www.ebi.ac.uk/interpro/), OMIM, PDB, and dbEST with spatiotemporal information. Information describing locations of gene products, gene classification, and protein structural orientations will be included to increase the accuracy of prediction. Due to the limited interactions found in various organisms, we will collect available datasets as the fundamental material from two main sources but focused on the physical interactions not covering indirect relationships.

We use Perl and XML parser to perform them into database format for effective data management. To uncover the domain composition, we will use InterProScan to dissect all available proteins into their components in term of domains defined by InterPro member-database.

A mixed statistical model of Association and Maximum Likelihood Estimation (MLE), “Hybrid Model”, is proposed to estimate the probability (DDI index, Domain-domain Interaction Index) for each domain pair to interact in S. cerevisiae from its protein interaction data derived from two-hybrid assays. Here we can
perform the Associate method to get the initial $\lambda_{mn}$ for MLE. $\lambda_{mn}$ means the probability of a given pair of domains ($m$ and $n$) having interaction, two proteins ($P_i, P_j$) have interaction ($P_{ij}$=1) if at least one pair of their domains is identified to have interaction (equation 1). Therefore, the probability of each pair of domains having interaction, $\lambda_{mn}$, is proposed to be estimated. The event that having a domain interaction is presumed to be independent from others. This proposed Hybrid Model takes the advantages of biologically legitimate initials, association measures, in performing heuristic computation for MLE. The DDI index generated from *S. cerevisiae* is then used to assign putative interacting partners of proteins of *C. albicans* (37,734 putative protein interactions, 2,357 out of them have probability equal to 1).

$$\text{Prob}(P_{ij} = 1) = 1 - \prod_{D_{mn} \in P_{ij}} (1 - \text{Prob}(D_{mn} = 1))$$  

(equation 1)

### 3 Visualization and Application of *C. albicans* Protein Network

This putative network for *C. albicans* rewired the pathogen protein interactions by incorporating the association measures and MLE to infer putative interacting relationships among proteins from domain interactions in yeast by the strategy of comparative proteomics. It performs as a convenient user-friendly interface integrated with rapid graphical networking maps plus instant gene annotation (Figure 1). With the power of combining experimental and inferred interactions, this value-added system is capable of depicting the global view of *C. albicans* protein interactions network, annotating ORFs without available functional expositions, and offering proper candidates to narrow down the scale of further high-throughput screening and validating experiments for drug targets. Specific protein-protein and protein-ligand interactions are central to most biological processes and are the focus of many avenues of research to develop small molecule-based therapies that will disrupt these essential interactions. The critical target proteins in those networks will be unrevealed by the topological analysis of protein network. Using the *in silico* Gene Knockout to silence a particular target in the network, we can observe the skew or destruction of protein network and challenge the robustness of whole network. The simulation results will be validated by the gene silence experimental data done in lab. We hope our work can provide further understand the pathology of *Candida albicans* related diseases and explore drug targets for effective medical treatments.

![Figure 1](image-url)

**Figure 1:** A diagram of the experimental and inferring visualized interaction networks of *C. albicans* interactomes based on the comparative proteomics strategy. The protein of interest can be easy to query by the fulltext search with several filters having biological meanings. After selecting the threshold, the protein network will be depicted with dynamic annotation windows. The solid and black lines indicate the interactions existing in the experimental results, and the red and solid lines depict the putative relations in the highest possibility. The inferring interactions are represented in the red and dot lines in the various breadths. The information of ORFs can be revealed by the mouse-over and all the interactions can saved as CSV format for further applications.