Supplementary Material

Figure S2: A) Proteins sequence in fasta format from public databases String, Intact and Psibase. B) Proteins sequence in fasta format from DIP database. C) Blast+ reciprocal hits from alignment of A versus B, and B versus A. Represents the hits generated from Blast+ running with the parameter set to 1 and 20. D) Pairs of PPI (protein-protein interactions) from public databases String, Intact and Psibase. E) Our predicted dataset, generated from the orthologous pairs of PPI mapped from D. The metrics avg(Ab) and min(ab) were used to generated the ROC curves. (C) is used to map each pDB identifier (D) to the ortholog DIP identifier. F) Our gold standard dataset, contain experimental and curated true PPI pairs (T), used to validade the predicted PPI pairs in E. The false PPI set were randomly created, having about five times more pairs than the true PPI set.