Materials & Methods

**Assumption:** within-species PPIs can be used to predict cross-species PPIs

- **Compute the probability that two proteins interact given they each contain a set of functional domains**

\[
Pr\{I(g, h) | D(g, d), D(h, e)\} = \frac{|S_{g,e}|}{|P_d||P_e| - |P_d \cap P_e|}
\]

**Evaluation Methods**

- **Predictions yield a weighted bipartite graph where the weight of an interaction is the computed probability**
- **H-H-P triplets: one pathogen protein predicted to interact with two host proteins (similarly H-P-P triplet)**
- **Enrichment scores are collapsed to most specific and most enriched**

**Datasets Used**

**Prediction:**

- **PPIs:** Reactome (Ishii-Tope 2005), IntAct (Hermjakob 2004), BIND (Gilbert 2005), DIP (Saliwonski 2004), and MIPS (Guldener 2004)
- **Functional Domains:** InterProScan (Quevillon 2005)

**Evaluation:**

- **Protein Function:** Gene Ontology (Ashburner 2001)
- **Gene Expression:** Malaria expression measured during merozoite invasion of human red blood cells (Bozdech 2003, Le roch 2003), Human expression measured in peripheral blood mononuclear cells from infected and healthy individuals (Boldt unpublished, Ockenhouse 2006)

**Results**

**Predictions**

- **Focus predictions on those proteins likely to be involved in pathogenesis**
- **Remove proteins annotated with functions related to the nucleus, ribosome, nucleic acid binding, helicase activity, and proteolysis**
- **Train using 7,876 human PPIs and 214 Plasmodium PPIs**
- **Predict a total of 516 interactions between 158 human proteins and 30 Plasmodium proteins with probability of at least 0.50**
- **Predict no interactions in a hypothetical fly-Plasmodium system**

**Datasets Used**

**Prediction:**

- **PPIs:** Reactome (Ishii-Tope 2005), IntAct (Hermjakob 2004), BIND (Gilbert 2005), DIP (Saliwonski 2004), and MIPS (Guldener 2004)
- **Functional Domains:** InterProScan (Quevillon 2005)

**Evaluation:**

- **Protein Function:** Gene Ontology (Ashburner 2001)
- **Gene Expression:** Malaria expression measured during merozoite invasion of human red blood cells (Bozdech 2003, Le roch 2003), Human expression measured in peripheral blood mononuclear cells from infected and healthy individuals (Boldt unpublished, Ockenhouse 2006)

**Evaluation Methods**

- **Predictions yield a weighted bipartite graph where the weight of an interaction is the computed probability**
- **H-H-P triplet: one pathogen protein predicted to interact with two host proteins (similarly H-P-P triplet)**
- **Enrichment scores are collapsed to most specific and most enriched**

**Datasets Used**

**Prediction:**

- **PPIs:** Reactome (Ishii-Tope 2005), IntAct (Hermjakob 2004), BIND (Gilbert 2005), DIP (Saliwonski 2004), and MIPS (Guldener 2004)
- **Functional Domains:** InterProScan (Quevillon 2005)

**Evaluation:**

- **Protein Function:** Gene Ontology (Ashburner 2001)
- **Gene Expression:** Malaria expression measured during merozoite invasion of human red blood cells (Bozdech 2003, Le roch 2003), Human expression measured in peripheral blood mononuclear cells from infected and healthy individuals (Boldt unpublished, Ockenhouse 2006)

**Evaluation Methods**

- **Predictions yield a weighted bipartite graph where the weight of an interaction is the computed probability**
- **H-H-P triplet: one pathogen protein predicted to interact with two host proteins (similarly H-P-P triplet)**
- **Enrichment scores are collapsed to most specific and most enriched**

**Proximity**

**Coexpression**

**Weighted Functional Enrichment**

- **Predicted Network**
- **Known Network**
- **Expression Datasets**

**Conclusion**

- **Predicting host-pathogen interactions is an important and unsolved problem with implications in biomedicine**
- **Our algorithm is able to identify biologically important interactions**
- **Future plans:** incorporate additional data sources as they become available
- **Experiments:** verify biologically interesting interactions