README of predicted host-VF protein-protein interactions (PPIs):

"Host-VF-PPI-interactions-predicted.zip"

This zip file includes two folds inside, which correspond to two inference methods of prediction. For each txt file, the name begins with an index number. You can ignore the number. It is then followed with another number, which is the species taxonomy. If a file has no content, it means that there is no interaction inferred for the species.

An example of the data from 4_235_infer_sim:
A6NC05   Q2YJZ1
A8MT40   Q2YKY1
B1ANE3   Q2YP51

235 is a NCBI taxonomy ID: Brucella abortus
https://www.ncbi.nlm.nih.gov/taxonomy/?term=235

Each row: predicted interaction
The first column: "human protein" uniprot ID
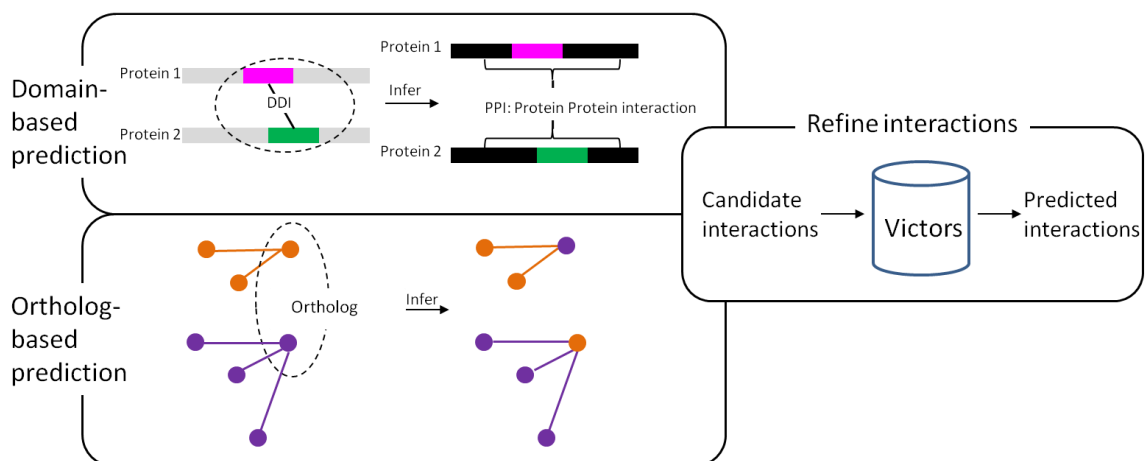   e.g., A6NC05 - human Glutaredoxin-like protein C5orf63
        https://www.uniprot.org/uniprot/A6NC05
The second column: "pathogen protein" uniprot ID
        e.g., Q2YJZ1 – Glutaredoxin nrdH from Brucella abortus (strain 2308)
        https://www.uniprot.org/uniprot/Q2YJZ1

--
Method description:

Project workflow for predicting pathogen-human protein-protein interactions. For any two proteins, one from human and the other from a pathogen, it is inferred that these two proteins interact with each other when they contain interacted domains (domain-domain interaction or DDI), or if one protein is an ortholog of a third protein that interacts with the other protein. All proteins from 30 pathogens (Supplemental Table 2) were used to predict possible human-pathogen protein-protein interactions. The list of Victors virulence factors was used to filter out other pathogen proteins to refine the interaction results.

As domain is the conserved part of a protein, the "domain inferred" prediction is to infer inter-species PPIs from "domain-domain interactions" (DDI) (1). The computational algorithm works as follows:

(i) $A_{pathogen}$ protein has Domain Union U1$\{$ $pu_1,pu_2,…pu_n$ $\}$ and B_{human} protein has Domain Union U2$\{$ $hu_1,hu_2,…hu_n$ $\}$;

(ii) If there are more than one DDI within U1 and U2: DDI(U1,U2) $\neq$0 & >=1;

(iii) Then PPI($A_{pathogen,}$ B_{human}) is true.

The "ortholog inferred" prediction is to infer the inter-species PPIs based on their orthologs interaction information. The orthologous pairs of proteins were defined by employing Inparanoid program (http://inparanoid.cgb.ki.se/) with default parameters (2). The PPI database HPRD (http://www.hprd.org/) was used to find the protein-protein interactions (3). The computational algorithm works as follows:

(i) OrthoU3$\{ou_1,ou_2,…ou_n\}$ is orthologous human proteins of $A_{pathogen}$ protein;

(ii) If any member in OrthoU3 interacts with B_{human} protein: PPI(ou_x, B_{human});

(iii) Then PPI($A_{pathogen}$, B_{human}) is true.

Furthermore, both interactions generated by "domain inferred" and "ortholog inferred" methods were combined into a unique human-pathogen interactome.

References:

1. Zhao, X.M., Zhang, X.W., Tang, W.H. and Chen, L. (2009) FPPI: Fusarium graminearum protein-protein interaction database. *Journal of proteome research*, **8**, 4714-4721.
2. Ostlund, G., Schmitt, T., Forslund, K., Kostler, T., Messina, D.N., Roopra, S., Frings, O. and Sonnhammer, E.L. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic acids research*, **38**, D196-203.
3. Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A. *et al.* (2009) Human Protein Reference Database--2009 update. *Nucleic acids research*, **37**, D767-772.