

## Supplementary material

S. No.	Species	Number of Protein Sequences
1	<i>H. sapiens</i>	27970
2	<i>P. triglodytes</i>	51931
3	<i>C. familiaris</i>	33654
4	<i>B. taurus</i>	24853
5	<i>M. musculus</i>	26657
6	<i>G. gallus</i>	18532
7	<i>D. melanogaster</i>	19765
8	<i>A. mellifera</i>	9257
9	<i>A. gambiae</i>	12659
10	<i>D. rerio</i>	34438
11	<i>C. elegans</i>	19568
12	<i>A. thaliana</i>	31711
13	<i>S. cerevisiae</i>	5880
14	<i>Influenza virus</i>	8916
15	<i>P. falciparum</i>	5267

**Table 1:** Total number of protein sequences taken for study in each species.