

Additional data file 7 - List of sources and versions of predicted protein datasets from 45 eukaryotes used in this work.

Organism	Source	Version	Web reference	Download filename
<i>Apis mellifera</i>	beebase	v2 (prerelease)	racerx00.tamu.edu	Amel_pre_release2_OGS_pep.fa
<i>Arabidopsis thaliana</i>	TAIR	TAIR7	www.arabidopsis.org/	TAIR7_pep_20070425
<i>Aureococcus anophagefferens</i>	JGI	v1.0	www.jgi.doe.gov/	proteins.Auran1_FilteredModels3.fasta.gz
<i>Batrachochytrium dendrobatidis</i>	Broad Institute	v1.0	www.broad.mit.edu/annotation/fgi/	batrachochytrium_dendrobatidis_1_proteins.fasta.gz
<i>Caenorhabditis elegans</i>	WormBase	WS170	www.wormbase.org/	wormpep170.tar.gz
<i>Capitella sp. I</i>	JGI	v1.0	www.jgi.doe.gov/	FilteredModels1.0.aa.fasta.gz
<i>Chlamydomonas reinhardtii</i>	JGI	v3.1	www.jgi.doe.gov/	Chlre3_1.GeneCatalogProteins.6JUL06.fasta.gz
<i>Ciona intestinalis</i>	JGI	v2.0	www.jgi.doe.gov/	FM1.aa.fasta.gz
<i>Cryptosporidium parvum</i>	CryptoDB	v3.4	www.cryptodb.org/	CparvumAnnotatedProtein.fsa.zip
<i>Cyanidioschyzon merolae</i>	C.merolae genome project	?	merolae.biol.s.u-tokyo.ac.jp/	cds.fasta
<i>Danio rerio</i>	VEGA	May 2007	vega.sanger.ac.uk/	Danio_rerio.VEGA.may.pep.tot.fa.gz
<i>Dictyostelium discoideum</i>	dictyBase	05-20-2007	dictybase.org/	dicty_predictions_protein.gz
<i>Drosophila melanogaster</i>	ENSEMBL	BDGP4.3, 44.43b	www.ebi.ac.uk/ensembl/	Drosophila_melanogaster.BDGP4.3.44.pep.all.fa.gz
<i>Encephalitozoon cuniculi</i>	NCBI	?	www.ncbi.nlm.nih.gov	encephalitozoon_cuniculi Field: Organism
<i>Entamoeba histolytica</i>	geneDB	17102005	www.genedb.org/	GeneDB_Ehistolytica_Proteins_17102005
<i>Gallus gallus</i>	ENSEMBL	WASHUC2, 44.2b	www.ebi.ac.uk/ensembl/	Gallus_gallus.WASHUC2.44.pep.all.fa.gz
<i>Giardia lamblia</i>	GiardiaDB	v1.0	www.giardiadb.org/	GlambliaAnnotatedProteins_GiardiaDB-1.0.fasta
<i>Homo sapiens</i>	VEGA+Refseq	'Jun 2007	vega.sanger.ac.uk/ , www.ncbi.nlm.nih.gov	Homo_sapiens.VEGA.jun.pep.tot.fa.gz
<i>Leishmania major</i>	geneDB	v5.2	www.genedb.org/	LmjFwholegenome_20060901_V5.2.pep
<i>Lottia gigantea</i>	JGI	v1.0	www.jgi.doe.gov/	Lotgi1_GeneModels_FilteredModels1_aa.fasta.gz
<i>Monosiga brevicollis</i>	JGI	v1.0	www.jgi.doe.gov/	Monbr1_best_proteins.fasta.gz
<i>Naegleria gruberi</i>	JGI	v1.0	www.jgi.doe.gov/	Naegr1_best_proteins.fasta.gz
<i>Nematostella vectensis</i>	JGI	v1.0	www.jgi.doe.gov/	proteins.Nemve1FilteredModels1.fasta.gz
<i>Neurospora crassa</i>	Broad Institute	v7.0	www.broad.mit.edu/annotation/fgi/	neurospora_crassa_7_proteins.fasta.gz

<i>Oryza sativa</i>	TIGR	v5.0	www.tigr.org/tdb/e2k1/osa1/	all.pep
<i>Ostreococcus tauri</i>	JGI	v2.0	www.jgi.doe.gov/	O.tauri.FM.aa.fasta.gz
<i>Paramecium tetraurelia</i>	ParameciumDB	v1.04	paramecium.cgm.cnrs-gif.fr/	Ptetraurelia_peptides_v1.04.fasta
<i>Phaeodactylum tricornutum</i>	JGI	v2.0	www.jgi.doe.gov/	Phatr2_geneModels_FilteredModels2_aa.fasta.gz
<i>Physcomitrella patens</i>	JGI	v1.1	www.jgi.doe.gov/	proteins.Phypha1_1.FilteredModels.fasta.gz
<i>Phytophthora sojae</i>	JGI	v1.1	www.jgi.doe.gov/	proteins.FM_Physo1_1.fasta.gz
<i>Plasmodium falciparum</i>	PlasmoDB	v5.2	www.plasmodb.org/	PfalciparumAnnotatedProteins_plasmoDB-5.2.fasta
<i>Populus trichocarpa</i>	JGI	v1.1	www.jgi.doe.gov/	proteins.Poptr1_1.JamboreeModels.fasta.gz
<i>Rhizopus oryzae</i>	Broad Institute	v3.0	www.broad.mit.edu/annotation/fgi/	rhizopus_oryzae_3_proteins.fasta.gz
<i>Saccharomyces cerevisiae</i>	Saccharomyces genome database	20070713	www.yeastgenome.org/	orf_trans_all.fasta.gz
<i>Schizosaccharomyces pombe</i>	Sanger	v19	www.sanger.ac.uk/Projects/S_pombe	pompep
<i>Strongylocentrotus purpuratus</i>	Baylor College of Medicine	Spur_2.1, v4	www.hgsc.bcm.tmc.edu/	freeze4.tar.gz
<i>Takifugu rubripes</i>	ENSEMBL	Assembly4, 44.4e	www.ebi.ac.uk/ensembl/	Takifugu_rubripes.FUGU4.44.pep.all.fa.gz
<i>Tetrahymena thermophila</i>	TIGR	10/24/2006	www.tigr.org/tdb/e2k1/tta1/	TTA1.pep
<i>Thalassiosira pseudonana</i>	JGI	v3.0	www.jgi.doe.gov/	Thaps3_geneModels_FilteredModels2_aa.fasta.gz
<i>Theileria annulata</i>	GeneDB	v4	www.genedb.org/	TANN.GeneDB.pep
<i>Toxoplasma gondii</i>	ToxoDB	v4.1	www.toxodb.org/	TgondiiAnnotatedProteins_toxoDb-4.1.fasta
<i>Trichomonas vaginalis</i>	TIGR	20050331	www.tigr.org/tdb/e2k1/tvg/	T.vaginalis_20050331_aa.fasta.gz
<i>Trichoplax adhaerens</i>	JGI	v1.0	www.jgi.doe.gov/	Triad1_best_proteins.fasta.gz
<i>Trypanosoma brucei</i>	geneDB	v4	www.genedb.org/	Tb927_Proteins_May06_v4.fas.gz
<i>Ustilago maydis</i>	Broad Institute	v1.0	www.broad.mit.edu/annotation/fgi/	ustilago_maydis_1_proteins.fasta.gz
