

Creating species databases

These are downloaded from the links below. Use the recent most version.

In the instructions/database folder, these are named as

“**MusMusculus_geneprot_database**”, “**HomoSapiens_gene_prot_database**”, etc...

1. Homo Sapiens

ftp://ftp.ensemblgenomes.org/pub/release-44/pan_ensembl/tsv/ensembl-compara/homologies/homo_sapiens/

2. Dictyostelium Discoideum

ftp://ftp.ensemblgenomes.org/pub/release-44/pan_ensembl/tsv/ensembl-compara/homologies/dictyostelium_discoideum/

3. Mus Musculus

ftp://ftp.ensemblgenomes.org/pub/release-44/pan_ensembl/tsv/ensembl-compara/homologies/mus_musculus/

4. Danio Rerio

ftp://ftp.ensemblgenomes.org/pub/release-44/pan_ensembl/tsv/ensembl-compara/homologies/danio_rerio/

5. Rabbit

<https://www.ensembl.org/biomart/martview/f20b5589735189ea2553e1f3bcc5cb67>

example:

select required columns. Below is an example.

The screenshot shows the Ensembl Biomart interface. The left sidebar displays the 'Dataset' as 'Human genes (GRCh38.p12)' and the 'View' as 'Orthologous Rabbit Genes: Only'. The main table shows 10 rows of results. The columns are: Gene stable ID, Protein stable ID, Rabbit homology type, Rabbit gene stable ID, Rabbit protein or transcript stable ID, Rabbit gene name, %Id. query gene identical to target Rabbit gene, and Rabbit orthology confidence [0 low, 1 high].

Gene stable ID	Protein stable ID	Rabbit homology type	Rabbit gene stable ID	Rabbit protein or transcript stable ID	Rabbit gene name	%Id. query gene identical to target Rabbit gene	Rabbit orthology confidence [0 low, 1 high]
ENSG00000198888	ENSP00000354687	ortholog_one2one	ENSOCUG000000029086	ENSOCUP000000026179	ND1	77.673	1
ENSG00000198763	ENSP00000355046	ortholog_one2one	ENSOCUG000000029090	ENSOCUP000000026180	ND2	62.536	1
ENSG00000198804	ENSP00000354499	ortholog_one2one	ENSOCUG000000029096	ENSOCUP000000026181	COX1	92.0076	1
ENSG00000198712	ENSP00000354876	ortholog_one2one	ENSOCUG000000029099	ENSOCUP000000026182	COX2	72.2467	1
ENSG00000228253	ENSP00000355265	ortholog_one2one	ENSOCUG000000029101	ENSOCUP000000026183	ATP8	44.7761	0
ENSG00000198899	ENSP00000354632	ortholog_one2one	ENSOCUG000000029102	ENSOCUP000000026184	ATP6	76.9911	1
ENSG00000198938	ENSP00000354982	ortholog_one2one	ENSOCUG000000029103	ENSOCUP000000026185	COX3	86.9732	1
ENSG00000198840	ENSP00000355206	ortholog_one2one	ENSOCUG000000029105	ENSOCUP000000026186	ND3	72.1739	1
ENSG00000212907	ENSP00000354728	ortholog_one2many	ENSOCUG000000029107	ENSOCUP000000026187	ND4L	76.5306	1
ENSG00000212907	ENSP00000354728	ortholog_one2many	ENSOCUG000000025982	ENSOCUP000000022571		73.4694	0

At the bottom, a message states: 'In order to maintain service for all users, BioMart browser sessions running for more than 5 minutes are terminated. If you have queries that you think will run longer than this, please choose have the results emailed to you. Note that queries that run for longer than 6 hours will be terminated even when submitted this way. If this happens please reformat your query or contact us for details on how to approach this.'