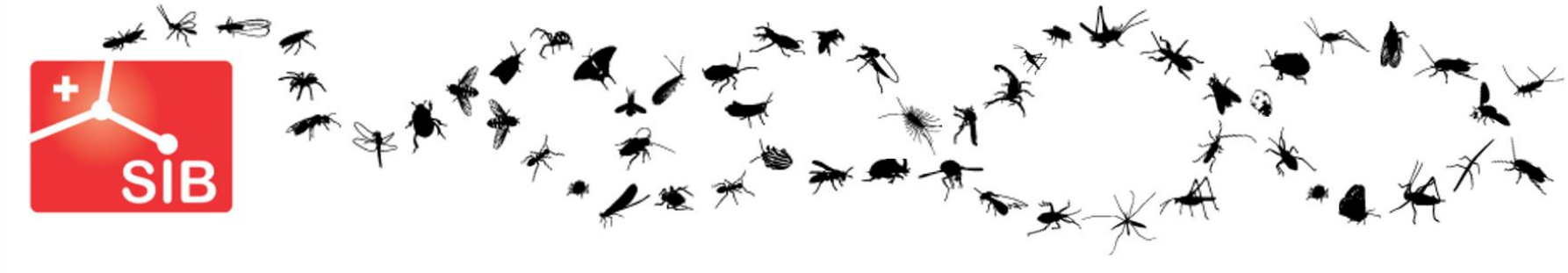
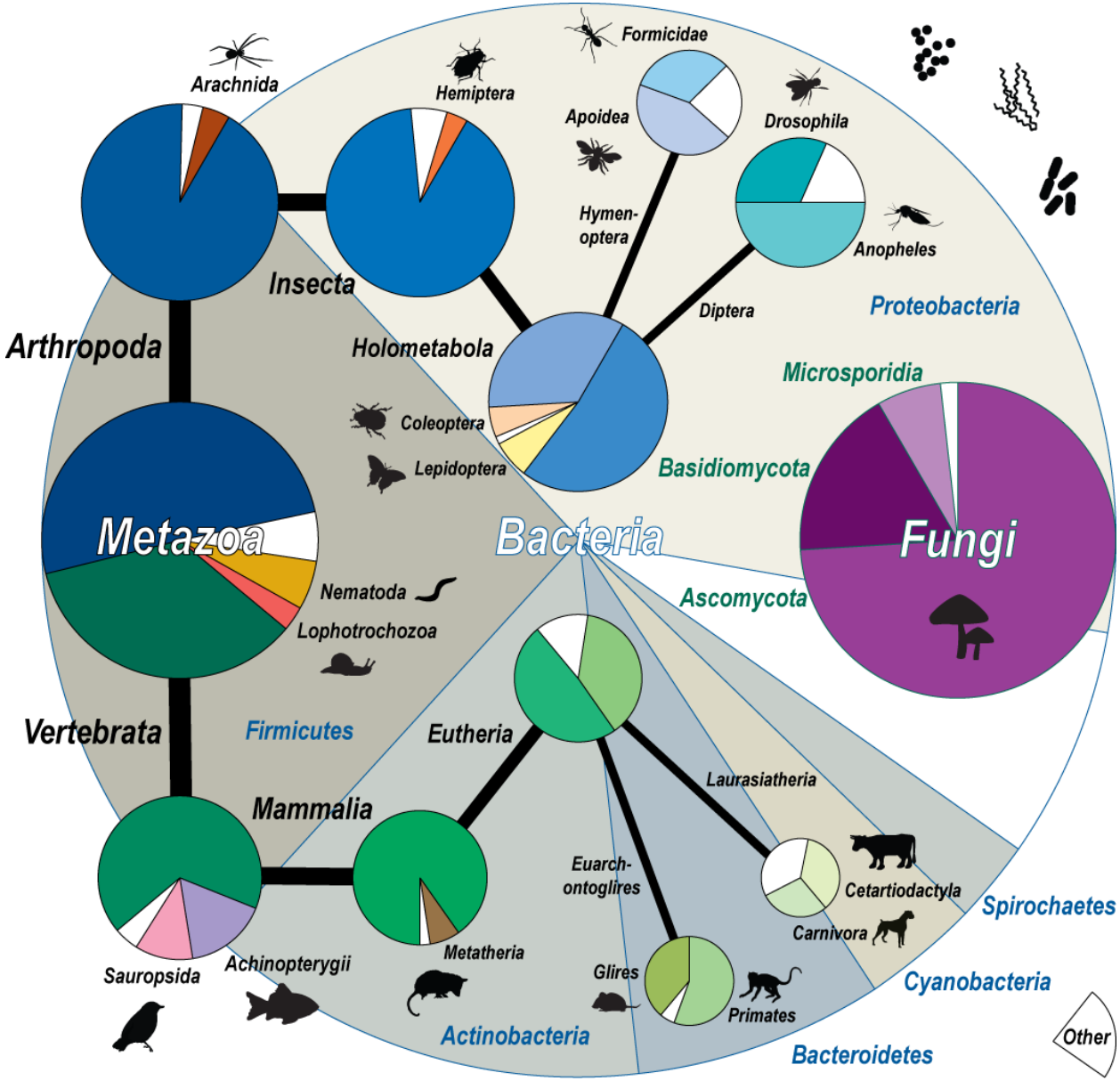


UniL SIB 3-5 September 2018

Orthology using OrthoDB

robert.waterhouse@gmail.com @rmwaterhouse www.rmwaterhouse.org

Orthology @ OrthoDB



Species Coverage:

- **3663** Bacteria
- **435** Archaea
- **3139** Viruses
- **659** Eukaryota

- **227** Fungi
- **71** Protozoa
- **31** Plants
- **330** Metazoa
  - **133** Arthropoda

Tools:

- ✓ Web browser
- ✓ JSON API
- ✓ SPARQL
- ✓ Data downloads
- ✓ Software package
- ✓ BUSCO

What does OrthoDB offer?

OrthoDB

OrthoDB

Your search for **imd** at Arthropoda level returned 2 groups

Bookmark OrthoDB@Arthropoda | Get All Fasta | Get All as Tab delimited

Download results

Group EOG8GXJ2W at Arthropoda level

Group EOG834ZNB at Arthropoda level View Fasta | View Tab Delimited

Functional descriptions

Functional descriptions

GO Molecular Function	1 gene with GO:0005102: receptor binding 1 gene with GO:0005515: protein binding
GO Biological Process	1 gene with GO:0009609: response to symbiotic bacterium 1 gene with GO:0045824: negative regulation of innate immune res 1 gene with GO:0050777: negative regulation of immune response 1 gene with GO:0061060: negative regulation of peptidoglycan reco
Functional Category	Signal transduction mechanisms

Evolutionary descriptions

Evolutionary descriptions

Phyletic Profile	50 genes in 48 species (out of 87) single copy in 46 species, multi-copy in 2 species
Evolutionary Rate	1.48
Gene Architecture	Median Protein Length 199 (std. 47) Median Exon Count 1 (std. 0.6)

Orthologs by organism

Table of orthologues

Build your query | Search by sequence

Sequence

Text search: imd

Text Search

Phyloprofile: [No filtering]

Phyloprofile filter

Search at: Arthropoda

Clade selector

Species to display: Clear all

- Metazoa *(metazoans)*
- Arthropoda *(arthropods)*
- Insecta *(true insects)*
- Hemiptera
- Acyrthosiphon pisum *(pea aphid)*
- Daphnia pulex *(water flea)*

Species selector

Submit

Select species:

Search species by name:

What does OrthoDB offer?

OrthoDB

1 101823790 TAR (HIV-1) RNA binding protein 2 › 365

2 101831629 protein kinase, interferon-inducible double stranded RNA dependen... › 313

InterPro

Mus musculus

1 Prkra (Q9WTX2) PKR-associated protein X ››› 313 8 IPR014720 32478

2 Tarbp2 (P97473) Protamine-1 RNA-binding protein 365 9 IPR028605 14720

upkws: cytoplasm; double-stranded RNA binding; enzyme binding; micro-ribonucleoprotein complex; miRNA binding; miRNA loading onto RISC involved in gene silencing by miRNA; multicellular organism growth; negative regulation of defense response to virus by host; nucleoplasm; nucleus; perinuclear region of cytoplasm; positive regulation of translation; positive regulation of viral genome replication;

mgi: Tarbp2 TAR (HIV) RNA binding protein 2

Ensembl: ENSMUSG00000023051 TAR (HIV) RNA binding protein 2 [Source:MGI Symbol;Acc:MGI:103027]

phenotype: ABNORMAL SPERMIOGENESIS: anomaly in the process by which a spermatid transforms into a functional spermatozoon; CARDIAC FIBROSIS: formation of fibrous tissue within the heart often resulting from inflammation or injury; COMPLETE LETHALITY AT WEANING: premature death at weaning age of all organisms of a given genotype in a population, often due to the inability to make the

Extensive X-refs

UniProt: P97473 Protamine-1 RNA-binding protein; TAR RNA-binding protein 2; RISC-loading complex subunit TARBP2; Function:Required for formation of the RNA induced silencing complex (RISC). Component of the RISC loading complex (RLC), also known as the micro-RNA (miRNA) loading complex (miRLC), which is composed of DICER1, AGO2 and TARBP2. Within the RLC/miRLC, DICER1 and TARBP2 are

ExpressionAtlas: ENSMUSG00000023051

GO **Molecular Function:** RNA binding; double-stranded RNA binding; enzyme binding; siRNA binding; miRNA binding; protein homodimerization activity; protein N-terminus binding; pre-miRNA binding

GO **Cellular Component:** intracellular; cell; nucleus; nucleoplasm; cytoplasm; RISC complex; micro-ribonucleoprotein complex; organelle; protein complex; perinuclear region of cytoplasm; RISC-loading complex

GO **Biological Process:** spermatid development; single fertilization; ribonucleoprotein complex assembly; cell differentiation; production of siRNA involved in RNA interference; targeting of mRNA for destruction involved in RNA interference; pre-miRNA processing; cellular nitrogen compound metabolic process; siRNA loading onto RISC involved in RNA interference; multicellular organism growth; miRNA loading onto RISC involved in gene silencing by miRNA; positive regulation of viral genome replication; regulation of viral transcription; anatomical structure development; negative regulation of defense response to virus by host; macromolecular complex assembly

InternationalKnockoutMouseConsortiumKnockoutsEScellsavailable: Tarbp2

Havanagene: OTTMUSG00000033455

Vegagene: OTTMUSG00000033455; Tarbp2

UniProtKBGeneName: Tarbp2; Tarbp2

Entrez: Tarbp2

WikiGene: Tarbp2

E.g. UniProt, Ensembl, Entrez, GO, model organism DBs

www.orthodb.org

OrthoDB

UNIVERSITÉ DE GENÈVE
FACULTÉ DE MÉDECINE

Zdobnov's Computational Evolutionary Genomics group

SIB

OrthoDB start page | Comparative Charts | SPARQL | Own data & mapping | Help

OrthoDB

The Hierarchical Catalog of Orthologs v9.1

OrthoDB is a comprehensive catalog of orthologs, i.e. genes inherited by extant species from their last common ancestor. Arising from a single ancestral gene, orthologs form the cornerstone for comparative studies and allow for the generation of hypotheses about the inheritance of gene functions. Each phylogenetic clade or subclade of species has a distinct common ancestor, making the concept of orthology inherently hierarchical. From its conception, OrthoDB explicitly addressed this hierarchy by delineating orthologs at each major species radiation of the species phylogeny. The more closely related the species, the more finely-resolved the gene orthologies.

Read more or cite
"OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs."
Zdobnov EM et al, NAR, Nov 2016, PMID:27899580

Examples of how you can query OrthoDB
Cytochrome P450, protease | peptidase, kinase -serine, FBgn0036816, GO:0006950, immune response, stress response, breast cancer, diabetes.

Help, **Video Presentation** and **Email**: support[at]orthodb.org

Data downloads Protein sequences and orthologous group annotations for major clades.
OrthoDB software Can be used to compute orthologs on custom data.
BUSCO.v3 Assessing completeness of genome assembly and annotation with single-copy genes.

OrthoDB-News Join the mailing list to keep abreast of the latest developments.

Previous OrthoDB Releases

- OrthoDB9 2015: 172 vertebrates, 133 arthropods, 227 fungi, 25 basal metazoans, 3663 bacteria and 31 plants
- OrthoDB8 2014: 61 vertebrates, 87 arthropods, 227 fungi, 12 basal metazoans, and 2627 bacteria
- OrthoDB7 2013: 64 vertebrates, 57 arthropods, 175 fungi, 14 basal metazoans, and 1115 bacteria
- OrthoDB6 2012: 52 vertebrates, 45 arthropods, 142 fungi, 13 basal metazoans, and 1115 bacteria
- OrthoDB5 2011: 48 vertebrates, 33 arthropods, 73 fungi, and 12 basal metazoans

www.orthodb.org

OrthoDB

Build your query | Search by sequence

Text search: ?

Phyloprofile: ?

[No filtering]

[No filtering]

Search at: ?

Species to display: Clear all

Submit

Build your query | Search by sequence

Copy a protein sequence (<1000 a.a.): ?

Search at: ?

Species to display: Clear all

Submit

www.orthodb.org

OrthoDB

Select species: ?

Search species by name:

- Eukaryota 588 (eucaryotes) e.g. S.cerevisiae, C.elegans, M.oryzae, coelacanth, black-legged tick, water flea,
  - Metazoa 330 (metazoans) e.g. C.elegans, coelacanth, black-legged tick, water flea, platypus, X.tropicalis,
  - Fungi 227 (fungi) e.g. S.cerevisiae, M.oryzae
  - Embryophyta 31 (plants) e.g. A. thaliana, potato, bread wheat
- Bacteria 3663 (eubacteria) e.g. S.pneumoniae, E.coli, E.faecalis, S.agalactiae, H.pylori, A.baumannii
- Archaea 345 e.g. Haloferax volcanii
- Viruses 3139

Start by selecting a few species of interest …

Drosophila melanogaster
Anopheles gambiae
Apis mellifera
Tribolium castaneum
Bombyx mori

Select species:

Search species by name:

apis

- Apis
- Apis cerana (Asiatic honeybee)
- Apis dorsata (giant honeybee)
- Apis florea (little honeybee)
- Apis mellifera (honey bee)
- Spiroplasma apis B31

Selection tree expands and selected species marked

- [ ] Anopheles farauti
- [ ] Anopheles funestus *(African malaria mosquito)*
- [x] Anopheles gambiae *(African malaria mosquito)*
- [ ] Anopheles maculatus (M)
- [ ] Anopheles melas

- [ ] Drosophila erecta
- [ ] Drosophila grimshawi
- [x] Drosophila melanogaster *(fruit fly)*
- [ ] Drosophila mojavensis
- [ ] Drosophila persimilis

- [x] **Lepidoptera** 7 *(butterflies and moths) e.g. silkworm*
  - [ ] **Papilionoidea** 4 *(butterflies)*
  - [x] Bombyx mori *(domestic silkworm)*
  - [ ] Manduca sexta *(tobacco hornworm)*
  - [ ] Plutella xylostella *(diamondback moth)*

- [ ] Apis dorsata *(giant honeybee)*
- [ ] Apis florea *(little honeybee)*
- [x] Apis mellifera *(honey bee)*
- [ ] **Bombinae** 3
- [ ] Dufourea novaeangliae

Species to display: Clear all

- [ ] **Eukaryota** *(eucaryotes)*
  - [ ] **Metazoa** *(metazoans)*
    - [ ] **Arthropoda** *(arthropods)*
      - [ ] **Insecta** *(true insects)*
        - **Endopterygota**
          - * **Diptera** *(flies)*
            - * **Nematocera**
              - * **Anopheles**
                - ✓ Anopheles gambiae (A
            - * **Brachycera**
              - * **Drosophila** *(fruit flies)*
                - ✓ Drosophila melanogas
          - * **Hymenoptera** *(hymenopteran*
            - * **Aculeata**
              - * **Apoidea** *(bees)*
              - * Apis

'Species to display' panel now shows only selected species

www.orthodb.org

OrthoDB

- [ ] Anopheles farauti
- [ ] Anopheles funestus *(African malaria mosquito)*
- [x] Anopheles gambiae *(African malaria mosquito)*
- [ ] Anopheles maculatus M
- [ ] Anopheles melas

- [ ] Apis dorsata *(giant honeybee)*
- [ ] Apis florea *(little honeybee)*
- [x] Apis mellifera *(honey bee)*
- [ ] Bombinae 3
- [ ] Dufourea novaeangliae

Five species of interest …

Drosophila melanogaster
Anopheles gambiae
Apis mellifera
Tribolium castaneum
Bombyx mori

- [ ] Drosophila erecta
- [ ] Drosophila grimshawi
- [x] Drosophila melanogaster *(fruit fly)*
- [ ] Drosophila mojavensis
- [ ] Drosophila persimilis

- [x] Lepidoptera 7 *(butterflies and moths) e.g. silkworm*
  - [ ] Papilionoidea 4 *(butterflies)*
  - [x] Bombyx mori *(domestic silkworm)*
  - [ ] Manduca sexta *(tobacco hornworm)*
  - [ ] Plutella xylostella *(diamondback moth)*

www.orthodb.org

OrthoDB

Search for a term, e.g. immunity

Submit

Build your query | Search by sequence

Text search:

immunity

immunity (10)

Phylopromo:

[No filtering]

[No filtering]

Search at:

Endopterygota

Species to display: Clear all

- Eukaryota *(eucaryotes)*
  - Metazoa *(metazoans)*
    - Arthropoda *(arthropods)*
      - Insecta *(true insects)*
        - Endopterygota
          - Diptera *(flies)*
            - Nematocera
              - Anopheles
                - Anopheles gambiae (A

Autocomplete with counts of cached terms shown

NB: 'Search at' is now set automatically to the last common ancestor level of all the species you selected

You can choose older one:

Search at:

Endopterygota

Eukaryota
Metazoa
Arthropoda
Insecta
Endopterygota

Clear all

www.orthodb.org

OrthoDB

10 orthologous groups returned

OrthoDB

Your search for **immunity** at Endopterygota level returned 10 groups

Bookmark OrthoDB@Endopterygota | Get All Fasta | Get All as Tab delimited ?

Group EOG090R04SW at Endopterygota level — 194 genes in 98 species
Similarity:Contains 1 RHD (Rel-like) domain.

Group EOG090R0F9M at Endopterygota level — 90 genes in 85 species
Nuclear cap-binding protein subunit 2

Group EOG090R04J6 at Endopterygota level — 100 genes in 98 species
Arsenite-resistance protein 2

Group EOG090R03S4 at Endopterygota level — 102 genes in 99 species
Nuclear cap-binding protein subunit 1

Group EOG090R0FIQ at Endopterygota level — 348 genes in 97 species
Peptidoglycan recognition protein

Group EOG090R07LX at Endopterygota level — 122 genes in 95 species
Protein kinase domain

Group EOG090R008X at Endopterygota level — 110 genes in 100 species
Similarity:Contains FAD-binding FR-type domain.

www.orthodb.org

OrthoDB

10 orthologous groups returned

OrthoDB

Your search for **immunity** at Endopterygota level returned 10 groups

Bookmark OrthoDB@Endopterygota | Get All Fasta | Get All as Tab delimited ?

1. Drag the Bookmarklet link to your toolbar to be able to quickly and easily conduct future searches at this level

Your browser's toolbar: OrthoDB@Endoptery

RESEARCH ARTICLE

The *Drosophila* Hox gene *Ultrabithorax* acts in both muscles and motoneurons to orchestrate formation of specific neuromuscular connections

Christian Hessinger, Gerhard M. Technau, Ana Rogulja-Ortmann

Development 2017 144: 139-150; doi: 10.1242/dev.143875

Highlight gene name / text
Click the Bookmarklet
Will search OrthoDB at your saved level and with your species selection

Article | Figures & tables | Supp info | Info & metrics | PDF | PDF + SI

Abstract

Hox genes are known to specify motoneuron pools in the developing vertebrate spinal cord and to control motoneuronal targeting in several species. However, the mechanisms controlling axial

www.orthodb.org

OrthoDB

10 orthologous groups returned

OrthoDB

Your search for **immunity** at Endopterygota level returned 10 groups

Bookmark OrthoDB@Endopterygota | Get All Fasta | Get All as Tab delimited ?

2. Get ALL protein sequences (FASTA format) from the selected species for <u>ALL 10</u> of the search result orthologous groups


```
>7070:00073c {"pub_gene_id":"TC002498", "pub_og_id":"EOG090R008X", "og_name":"Similarity:Contains FAD-binding FR-type domain.","level":33392, "description":"Putative uncharacterized protein  "}
MVSLTSLLFHVLTFCVIVLVAISTVRPEVRTEKQRYDGWFNNLAHPDWGSVGSHLIRRAPSAYSDGVYMLAGQNRPSPRKLSRLFMRGLDGLGSMNNRTALLAFFGQMVTSEVMMASESGCPIEMHHIEIEKCDDVYDKECRGDKYIPFHRAGYDRKTGQSPNSPREQLNQVTSWIDGSFIYSTSEPWVNAMRSFRNGTFLTDA
RNSMRVPLFNNPVPHVMKMLSTERLFLLGDPRTNQNPALLTISILFFRWHNVIAERVQKEHPDWSDEDVFQRTRRIVVATVQNIIAYEYIPAFLGQPLPEYSGYKQDVHPGVTHVFQSAAFRYGHSLIPPGLYRRDGECNFEESPMGLRLCATWWDSNDVLISHSLEKLLMGMASQLAEREDSVLCSDVRDKLFGPMEFSRRDL
RGRDSGLPDYNTIRAYYGLPKIKKWADINPKLFEEKPELLRALVSAYANNINNIDVYVGGMLESYGGPGELFTTVIKEQFARLRDSDRFWFENEGNGIFTREEIASLRKVTLWDVIVNSTNIAPNHIQKNVFFWMSGDPCPQPMQLNTSLMEPCKILNGYDYFEGNELVYIYACVFLAFVPILCAGAGYGVVKLQNRRRRRLKI
NCVEGKLPVDTMIVQEWLHANHKRLVKIKFGPEIAIHIMGRKGEKLRTVNFKNTETVYVEISVELTNQKKPMLLIRVPRDHDLVIELDSIGSRKKFLTKLELFLTSNKKSLIITEVVREIMLAKAETSERRQKKLEHFFREAYALTFGLRPGERRRRSDASVDGEVVTVMRTSLTKSEFASALGMKHDAVFVKKMFNIVDKDGD
EFLDTVVLFSRGKTDDKLRIIFDMCDNDGNGVIDKEELSEMLRSLVEIARTTSLSDDHVTELIEGMFQDAGLQHKHFLTYNDFKEMMKEYKGEFVAIGLDCKGAKQNFLDTSTNVARMTSFHIEPTLDSEKNYLELKWNSLTTFLEENRQNIFYLFIFYVITIALFVERFIHYSFMAEHTDLRHIMGVGIAITRGSAASLSFCYS
MSRNLLTKLKEFSIQQYIPLDTHIQFHKIAACTALFFSLLHTVGHIVNFYHVSTQPLENLRCLTNEVRFPSDYKPGITFWLFQTITGLTGVFLFVIMSIIIVFAHPTIRKKAYNFFWATHNLYIILYILCLIHGLARLTGAPRFWLFFIGPGTIYILDKVVSLRTKYIPLDVIETELLPSDVIKIKFYRPPNLKYLSGQWVRLA
HEFHSFTLTSAPHENFLSCHIKAQGPWTWKLRNYFDPSNYNPEDQPKILLEGPFGGGNQDWYKFEVAVMVGGGIGVTPYASILNDLVFGTSTNRYSGVACKKVYFLWICPSHKHFEWFIDVLRDVEKKDVTNVLEIHIFITQFFHKFDLRTTMLYICENHFQRLSKTSIFTGLKAVNHFGRPDMTSFLKFVQKRHSYVSKIGVF
LTKSVMSACDDVNKGRKLPYFIHHFENFG
>7091:001565 {"pub_gene_id":"BGIBMGA005478", "pub_og_id":"EOG090R008X", "og_name":"Similarity:Contains FAD-binding FR-type domain.","level":33392
MAGPERPGARTLSKIFMRGQDGLPSLTNRTALLAFFGQVVTGEIVMASESGCPIEHHRIPVDKCDHMYDSECQGAKYMPFLRAAYDRNTGQSPNSPREQINQMTSWIDGSFVYSTSEAWVNAMRSFQNGSFASDGGMPLRNTKRVPLFNNPVPHYMRMLSPERLFLLGDPRTNQNPAMVSFGILLFRWHNVVAARVHKQHPDWSI
RARRIVIASLQNIILYEYVPAFLGAALPPYEGYKPEVAPGVTHEFAVAAFRFGHTLVPPAILLRNRKCGYGRAPGGHDAIRLCQTWWDGNDVMSQVPLEEVLMGMASQLSEREDALLCSDVRDNLFGPMEFSRRDLGALNIMRGRDNGVPDYNTVREYFGLRKIKTFNEINPDLFEKNPDLLQKFIQIYNGDVDNIDVYIGGML
GELFRAIIIDQFTRIRDSDRFWFENDKNGIFTPRDIEELRRITLWDIIVNSTTIGAKDIQRDVFHWRAGDPCPQPYQLNASKLPPCKYLKGYDYFEGNEFAYIYICLILVFVPILCAGAGYGVVKLQNSRRRRLKIQQEHLKNAQCKGSVDKMVCREWVHASHKRLVKLRLGPESALHVTDRKGEKLRTLPLDHTDQLTVVESQ
RPLVLVRVPREHDLVLEMDSIASRRKFLVKLDTFLAQHKKALNLTQGPREQILATAETRERRQRKLEHFFREAYAITFGLAPGEKRRRSEDADPESIVMRTSLSKSEFASALGMKGDAVFVKKMFNIVDKDGDGRISFQEFLDTVVLFSRGATEDKLRIIFDMCDNDRNGVIDKGELSEMLRSLVEIARTTSLRDEHVTELIDG
LQHKDHLTYSDFKLMMKEYKGEFVAIGLDCKGAKQNFLDTSTNVARMTSFHIEPSMEQTRHWLLLKWDTLTTFLEENRQNIFYLFIFYVVTIGLFVERFAHYSFMSEHLDLRHIMGVGIAITRGSAASLSFCYSLLLLTMSRNLLTKLKEFSIQQYIPLDSSIQFHKIVACTALFFSLLHTAGHMVNFYHVSTQPVENLRCLTK
DFRPGITYWVFQTVTGVSGVLLFVIMCIIFVFAHPVIRKRAYPWFWRAHSLHVALYALCLVHGLARLTGAPRFWIFFLGPAIIYTLDKVVSLRTEYLALDVLETEMLPSDVIKIKFYRPPNLKYLSGQWVRLSCTAFKKEEFHSFTLTSAPHENFLSCHIKAQGPWTWKLRNYFDPCNYNTDDHPKIRIQGPFGGGNQDWYKFE
GGIGVTPYASILNDLVFGTSTNRYSGVACKKVYFLWICPSHKHFEWFIDVLRDVERKDVTNVLEIHIFITQFFHKFDLRTTMLYICENHFQRLSRTSMFTGLKAVNHFGRPDMSSFLKFVQKKHSYVSKIGVFSCGPRPLTKSVMSACEEVNRTRRLPYFIHHFENFG
>7165:002379 {"pub_gene_id":"DUOX", "pub_og_id":"EOG090R008X", "og_name":"Similarity:Contains FAD-binding FR-type domain.","level":33392, "description":"Similarity:Contains FAD-binding FR-type domain."}
MSHVEKQRYDGWYNNLAHPDWGAVDNHLTRKAPSAYSDGVYVMAGSNRPSPRKLSRLFMRGTDGLPSMENRTALLAFFGQVVTNEIVMASESGCPIEMHRIEIEKCDEMYDRECRGDRYIPFHRAAYDRNTGQSPNAPREQINQMTAWIDGSFIYSTSEAWLNAMRSFQDGALLTDKQGTMPVKNTMRVPLFNNPVPHVMRMLSI
LGDPRTNQNPALLSFAILFLRWHNVVAKRVRRQHRDWSDEEIFQRARRVVIASLQNIVAYEYLPAFLDKEIPPYDGYKADTHPGVSHMFQAAAFRFGHSLIPPGLFRRDGQCNFRRTNMDFPALRLCSTWWNSNDVLDNTPVEEFIMGMASQIAEKEDPLLCSDVRDKLFGPMEFTRRDLGALNIMRGRDNGLPDYNTARAAYR
WRDINPAVFERQPELLDLLIKTYDNQLDNVDVYVGGMLESDGRPGELFSAVIIDQFTRIRDADRFWFENEDNGIFTKEEIAEIRKFTLWDIIVNSTDIEADEIQRDVFHWKQGDPCPQPEQLNATLLEPCNYLEGYDYFSGSELAYIYSCVFLGFVPILCAGAGYCVIKLQNSRRRKLKIKQEAMKNTANTKVSVEKMVAREWLI
LVTVKFGPEASIYTVDRKGEKLRTFNLKHVDVVTVEQSQENYTAKKPYILLRVPNDHDLVLELESNSARRKFVKKLEDFLVLHKKTMTFVESNRDLMLAKAETRERRQKRLEHFFREAYALTFGLRPGERRRRSDASLDGEVMTVMRTSLSKSEFAAALGMKQDDMFVRKMFNIVDKDKDGRISFQEFLETVVLFSRGKTDDKLI
CDNDRNGVIDKGELSEMMRSLVEIARTTSVTDEQVNELIDGMFQDVGLEHKNHLTYEDFKLMMKEYKGDFVAIGLDCKGAKQNFLDTSTNVARMTSFHIEPISDSRRHWMQEKWDCYTTFLEENRQNIFYLFLFYVITIVLFVERFIHYSFMAEHTDLRHIMGVGIAITRGSAASLSFCYSLLLLTMSRNLLTKLKEFPIQQYIP
QFHKIAACTALFFSLLHTVGHIVNFYHVSTQSIENLKCLTKEVHFTSDYRPDITYWLFQTITGVTGVMLFVTMCIIFAFAHPTIRKKAYKFFWNAHSLYVVLYALCLVHGLARLTGAPRFWLFFIGPGIVYTLDKIVSLRTKYMALDVIETDLLPSDVIKIKFYRPPNLKYLSGQWVRLSCTEIKPEEMHSFTLTSAPHENFLS
GPWTWKLRNYFDPCNYNPDDQPKIRIEGPFGGGNQDWYKFEVAVMVGGGIGVTPYASILNDLVFGTSTNRYSGVACKKVYFLWICPSHKHFEWFIDVLRDVEKKDVTNVLEIHIFITQFFHKFDLRTTMLYICENHFQRLSKTSMFTGLKAVNHFGRPDMSSFLKFVQKKHSYVSKIGVFSCGPRPLTKSVMSACDEVNKSRKWI
FENFG
>7227:000fd5 {"pub_gene_id":"Duox", "pub_og_id":"EOG090R008X", "og_name":"Similarity:Contains FAD-binding FR-type domain.","level":33392, "description":"Similarity:Contains FAD-binding FR-type domain."}
MSVPSAPHQRAESKNRVPRPGQKNRKLPKLRLHWPGATYGGALLLLLISYGLELGSVHCYEKMYSQTEKQRYDGWYNNLAHPDWGSVDSHLVRKAPPSYSDGVYAMAGANRPSTRRLSRLFMRGKDGLGSKFNRTALLAFFGQLVANEIVMASESGCPIEMHRIEIEKCDEMYDRECRGDKYIPFHRAAYDRDTGQSPNAPREQ
WIDGSFIYSTSEAWLNAMRSFHNGTLLTEKDGKLPVRNTMRVPLFNNPVPSVMKMLSPERLFLLGDPRTNQNPAILSFAILFLRWHNTLAQRIKRVHPDWSDEDIYQRARHTVIASLQNVIVYEYLPAFLGTSLPPYEGYKQDIHPGIGHIFQAAAFRFGHTMIPPGIYRRDGQCNFKETPMGYPAVRLCSTWWDSSGFFADTS
GLASQISEREDPVLCSDVRDKLFGPMEFTRRDLGALNIMRGRDNGLPDYNTARESYGLKRHKTWTDINPPLFETQPELLDMLKEAYDNKLDDVDVYVGGMLESYGQPGEFFTAVIKEQFQRLRDADRFWFENERNGIFTPEEIAELRKITLWDIIVNSTDVKEEEIQKDVFMWRTGDPCPQPMQLNATELEPCTYLEGYDYFSG
YVCVFLGFVPILCAGAGYCVVKLQNSKRRRLKIRQEALRAPQHKGSVDKMLAREWLHANHKRLVTVKFGPEAAIYTVDRKGEKLRTFSLKHIDVVSVEESATNHIKKKPYILLRVPSDHDLVLELESYGARRKFVKKLEDFLLLHKKEMTLMEVNRDIMLARAETRERRQKRLEYFFREAYALTFGLRPGERRRRSDASSDGEVI
SLSKAEFAAALGMKPNDMFVRKMFNIVDKDQDGRISFQEFLETVVLFSRGKTDDKLRIIFDMCDNDRNGVIDKGELSEMMRSLVEIARTTSLGDDQVTELIDGMFQDVGLEHKNHLTYQDFKLMMKEYKGDFVAIGLDCKGAKQNFLDTSTNVARMTSFNIEPMQDKPRHWLLAKWDAYITFLEENRQNIFYLFLFYVVTIVLF
YSFMAEHTDLRHIMGVGIAITRGSAASLSFCYSLLLLTMSRNLITKLKEFPIQQYIPLDSHIQFHKIAACTALFFSVLHTVGHIVNFYHVSTQSHENLRCLTREVHFASDYKPDITFWLFQTVTGTTGVMLFIIMCIIFVFAHPTIRKKAYNFFWNMHTLYIGLYLLSLIHGLARLTGPPRFWMFFLGPGIVYTLDKIVSLRTK
IDTDLLPSDVIKIKFYRPPNLKYLSGQWVRLSCTAFRPHEMHSFTLTSAPHENFLSCHIKAQGPWTWKLRNYFDPCNYNPEDQPKIRIEGPFGGGNQDWYKFEVAVMVGGGIGVTPYASILNDLVFGTSTNRYSGVACKKVYFLWICPSHKHFEWFIDVLRDVEKKDVTNVLEIHIFITQFFHKFDLRTTMLYICENHFQRLSK
LKAVNHFGRPDMSSFLKFVQKKHSYVSKIGVFSCGPRPLTKSVMSACDEVNKTRKLPYFIHHFENFG
>7460:002ad8 {"pub_gene_id":"GB51481", "pub_og_id":"EOG090R008X", "og_name":"Similarity:Contains FAD-binding FR-type domain.","level":33392, "description":"Uncharacterized protein  "}
MTRRRPPRSDSNWIYLLILLLWLVLPIKTGVVHSYADKQRYDGWYNNLAHPDWGSIDSRLIRKMPAAYSDGVYMLAGQDRPSPRKLSQLFMQGDDGLPSVKNRTALFAFFGQLVTSEIIMASESGCPIEYHRIDVDKCDPVFDKECQGNKYIPFRRADYDRQTGRSPNSPREQINKVTSWIDGSFVYSSSEAWANTMRSFKNGS
RKFPVRNTMRAPLFNHAVPHVMRMLSPERLYLLGDPRTNQHPPLLALGILFYRWHNVIAARIQLENPTMSDEDIFQKARRVVIGTLQNIILYEYIPILLNEDLPPYTGYKSDLHPGISHIFQSAAFRFGHTLIPPGLYRRDENCEYRRTNTDQPAIRLCSTWWDSNEVLTNSTIEELLMGMTSQIAEKEDNLLGTDIRNNLFGPI
DLGALNIMRGRDNGLPDYNTARAHFKLPRKKTWNEINPELFNKNPSLLRTLVEIHSNNLNNMDVYVGGMLESSAGPGELFSTVIKEQFLRLRDSDRFWFENEENGIFTRSEIADIRRITLWDVIVNATGIPADSIQRKVFTWEEGDPCPQPYQLNSTMLEPCVPLORYDYFEGSELVYIYACVFLGFVPILCAGAGYGLVKLQN
```

www.orthodb.org

OrthoDB

10 orthologous groups returned

OrthoDB

Your search for **immunity** at Endopterygota level returned 10 groups

Bookmark OrthoDB@Endopterygota | Get All Fasta | Get All as Tab delimited

3. Get ALL gene/protein information of the genes from the selected species for ALL 10 of the search result orthologous groups

```
pub_og_id       og_name level_taxid     organism_taxid  organism_name   int_prot_id     pub_gene_id     description
EOG090R008X     Similarity:Contains FAD-binding FR-type domain. 33392   7070    Tribolium castaneum     7070:00073c     TC002498        Putative uncharacterized protein
EOG090R008X     Similarity:Contains FAD-binding FR-type domain. 33392   7091    Bombyx mori     7091:001565     BGIBMGA005478
EOG090R008X     Similarity:Contains FAD-binding FR-type domain. 33392   7165    Anopheles gambiae       7165:002379     DUOX    Similarity:Contains FAD-binding FR-type domain.
EOG090R008X     Similarity:Contains FAD-binding FR-type domain. 33392   7227    Drosophila melanogaster 7227:000fd5     Duox    Similarity:Contains FAD-binding FR-type domain.
EOG090R008X     Similarity:Contains FAD-binding FR-type domain. 33392   7460    Apis mellifera  7460:002ad8     GB51481 Uncharacterized protein
EOG090R00T7     Similarity:Contains HECT (E6AP-type E3 ubiquitin-protein ligase) domain.        33392   7070    Tribolium castaneum     7070:0016b5     TC007799        Putative un
GLEAN_07799
EOG090R00T7     Similarity:Contains HECT (E6AP-type E3 ubiquitin-protein ligase) domain.        33392   7091    Bombyx mori     7091:003707     BGIBMGA014088
EOG090R00T7     Similarity:Contains HECT (E6AP-type E3 ubiquitin-protein ligase) domain.        33392   7165    Anopheles gambiae       7165:0021b3     AGAP009516;gambif1;GPRGBB3
(Rel-like) domain.
EOG090R00T7     Similarity:Contains HECT (E6AP-type E3 ubiquitin-protein ligase) domain.        33392   7227    Drosophila melanogaster 7227:000f9c     FBgn0031384     Similarity:
ubiquitin-protein ligase) domain.
EOG090R00T7     Similarity:Contains HECT (E6AP-type E3 ubiquitin-protein ligase) domain.        33392   7460    Apis mellifera  7460:000f03     GB44030 Uncharacterized protein
EOG090R02QF     Protein kinase C        33392   7070    Tribolium castaneum     7070:003a76     TC033289
EOG090R02QF     Protein kinase C        33392   7070    Tribolium castaneum     7070:003ce1     TC033980
EOG090R02QF     Protein kinase C        33392   7165    Anopheles gambiae       7165:002ada     AGAP011988      AGC-kinase, C-terminal
EOG090R02QF     Protein kinase C        33392   7227    Drosophila melanogaster 7227:003246     aPKC    Protein kinase C
EOG090R02QF     Protein kinase C        33392   7460    Apis mellifera  7460:001ccd     GB47743
EOG090R03S4     Nuclear cap-binding protein subunit 1   33392   7070    Tribolium castaneum     7070:0001c8     TC000568        Putative uncharacterized protein
EOG090R03S4     Nuclear cap-binding protein subunit 1   33392   7091    Bombyx mori     7091:00062a     BGIBMGA001579
EOG090R03S4     Nuclear cap-binding protein subunit 1   33392   7165    Anopheles gambiae       7165:000409     Cbp80   80 kDa nuclear cap-binding protein
EOG090R03S4     Nuclear cap-binding protein subunit 1   33392   7227    Drosophila melanogaster 7227:000753     Cbp80;FBgn0022942       cap binding protein 80
EOG090R03S4     Nuclear cap-binding protein subunit 1   33392   7460    Apis mellifera  7460:00125a     GB44934 Uncharacterized protein
EOG090R04J6     Arsenite-resistance protein 2   33392   7070    Tribolium castaneum     7070:000a61     TC003562        Putative uncharacterized protein
EOG090R04J6     Arsenite-resistance protein 2   33392   7091    Bombyx mori     7091:000d97     BGIBMGA003480
EOG090R04J6     Arsenite-resistance protein 2   33392   7165    Anopheles gambiae       7165:002500     Ars2    Arsenite-resistance protein 2 homolog
EOG090R04J6     Arsenite-resistance protein 2   33392   7227    Drosophila melanogaster 7227:0013c5     Ars2;FBgn0033062        Arsenite-resistance protein 2
EOG090R04J6     Arsenite-resistance protein 2   33392   7460    Apis mellifera  7460:000ba0     GB43113 Uncharacterized protein
EOG090R04SW     Similarity:Contains 1 RHD (Rel-like) domain.    33392   7070    Tribolium castaneum     7070:001660     TC007697        Dorsal
```

www.orthodb.org

OrthoDB

10 orthologous groups returned

OrthoDB

Your search for **immunity** at Endopterygota level returned 10 groups

Bookmark OrthoDB@Endopterygota | Get All Fasta | Get All as Tab delimited ?

Group EOG090R04SW at Endopterygota level Similarity:Contains 1 RHD (Rel-like) domain.	194 genes in 98 species
Group EOG090R0F9M at Endopterygota level Nuclear cap-binding protein subunit 2	90 genes in 85 species
Group EOG090R04J6 at Endopterygota level Arsenite-resistance protein 2	100 genes in 98 species
Group EOG090R03S4 at Endopterygota level Nuclear cap-binding protein subunit 1	102 genes in 99 species
Group EOG090R0FIQ at Endopterygota level Peptidoglycan recognition protein	348 genes in 97 species
Group EOG090R07LX at Endopterygota level Protein kinase domain	122 genes in 95 species
Group EOG090R008X at Endopterygota level Similarity:Contains FAD-binding FR-type domain.	110 genes in 100 species

Super Short Summary Info

Summary Gene & Species Counts

www.orthodb.org

OrthoDB

Expanded PGRP orthologous group

Group EOG090R0FIQ at Endopterygota level

View Fasta | View Tab Delimited

Peptidoglycan recognition protein

Functional descriptions

GO Molecular Function	114 genes with GO:0008270: zinc ion binding 114 genes with GO:0008745: N-acetylmuramoyl-L-alanine amidase activity 59 genes with GO:0042834: peptidoglycan binding
GO Cellular Component	58 genes with GO:0005887: integral component of plasma membrane 58 genes with GO:0005576: extracellular region
InterPro Domains	80 genes with IPR015510: Peptidoglycan recognition protein 80 genes with IPR002502: N-acetylmuramoyl-L-alanine amidase domain 78 genes with IPR006619: Peptidoglycan recognition protein family domain, metazoa/bacteria 71 genes with IPR017331: Peptidoglycan recognition protein, PGRP-S

Evolutionary descriptions

Phyletic Profile	348 genes in 97 species (out of 102) single copy in 15 species, multi-copy in 82 species
Evolutionary Rate	1.05
Gene Architecture	Median Protein Length 190 (std. 55.9) Median Exon Count 2 (std. 3.25)

www.orthodb.org

OrthoDB

© R.M.Waterhouse

Orthologs by organism

☑ Selected species only

	Organism | Protein ID | UniProt | Description	AAs	Exons	InterPro
	Anopheles gambiae			
1	AGAP000536 peptidoglycan recognition protein (short) »	200	17	
2	AGAP001212 peptidoglycan recognition protein (Long) »	278!	19	
3	AGAP006342 peptidoglycan recognition protein (short) »	188	13	
4	AGAP006343 peptidoglycan recognition protein (short) »	188	22	
	Drosophila melanogaster			
1	PGRP-SA (Q9VYX7) Protein semmelweis »»	203	4	IPR015510 17331 02502 06619
2	FBgn0033327;PGRP-SC1a;PGRP-SC1b (A0A0B4LEX8) Peptidoglycan-recogniti... »»	185	2	IPR015510 17331 02502 06619
3	FBgn0037906;PGRP-LB (A0A0B4K741) Peptidoglycan recognition protein LB »»	255!	8	IPR017331 02502 06619 15510
4	FBgn0043575;PGRP-SC2 (Q2XY98) PGRP-SC2 »»	184	1	IPR015510 17331 02502 06619
5	FBgn0043576;PGRP-SC1a;PGRP-SC1b (Q2XY86) PGRP-SC1a »»	185	1	IPR017331 15510 02502 06619
6	FBgn0043577;PGRP-SB2 (M9PFJ1) PGRP-SB2 »»	191	3	IPR017331 06619 02502 15510
7	PGRP-SB1 (Q70PY2) Peptidoglycan-recognition protein SB1 »»	190	2	IPR017331 15510 02502 06619
	Apis mellifera			
1	GB47805 Uncharacterized protein »	194	4	
2	GB51741 Uncharacterized protein »	189	4	
	Bombyx mori			
1	BGIBMGA007987 »	195	4	
2	BGIBMGA008038 Peptidoglycan recognition protein »	218	4	
3	BGIBMGA012866 »	!128	3	
	Tribolium castaneum			
1	TC010611 Putative uncharacterized protein »	195	4	
2	TC013620 Putative uncharacterized protein »	188	4	
3	TC015689 Putative uncharacterized protein GLEAN_15689 »	207	2	

Expand ...

Sibling Groups

Group	Overlap	InterPro domains
EOG090R0IO9	27%	IPR015510 02502 06619 17331

Expanded gene annotation (incl. search term)

7 PGRP-SB1 (Q70PY2) Peptidoglycan-recognition protein SB1 190 2 IPR017331 15510 02502 06619

upkws: extracellular region; immune response; innate immune response; microtubule associated complex; N-acetylmuramoyl-L-alanine amidase activity; peptidoglycan binding; peptidoglycan catabolic process; zinc ion binding

flybase: PGRP-SB1 The gene PGRP-SB1 is referred to in FlyBase by the symbol Dmel\PGRP-SB1 (CG9681, FBgn0043578). It is a protein_coding_gene from Drosophila melanogaster. It has one annotated transcript and one polypeptide. Gene sequence location is 3L:16727299..16727989. It has the cytological map location 73C1. Protein features are: N-acetylmuramoyl-L-alanine amidase domain; Peptidoglycan recognition protein; Peptidoglycan recognition protein family domain, metazoa/bacteria; Peptidoglycan recognition protein, PGRP-S. Its molecular function is described by: N-acetylmuramoyl-L-alanine amidase activity; zinc ion binding; peptidoglycan binding. It is involved in the biological process described with: defense response; immune response; peptidoglycan catabolic process. 5 alleles are reported. No phenotypic data is available. No phenotypic class data is available. Summary of modENCODE Temporal Expression Profile: Temporal profile ranges from a peak of very high expression to a trough of very low expression. Peak expression observed in adult female stages.

Ensembl: FBgn0043578 PGRP-SB1 [Source:FlyBase gene name;Acc:FBgn0043578]

UniProt: Q70PY2 Peptidoglycan-recognition protein SB1; Catalytic Activity:Hydrolyzes the link between N-acetylmuramoyl residues and L-amino acid residues in certain cell-wall glycopeptides.; Function:N-acetylmuramyl-L-alanine amidase involved in innate immunity by degrading bacterial peptidoglycans (PGN), preferentially DAP-type PGNs. Probably plays a scavenger role by digesting biologically active PGN into biologically inactive fragments.; Similarity:Belongs to the N-acetylmuramoyl-L-alanine amidase 2 family.; Tissue Specificity:In larvae, it is mainly expressed in fat body.

CTD: 39870

GenomeRNAiadatabaseforcellbasedRNAiphenotypes: 39870

FlyBasegeneCGID: CG9681

FlybaseAnnotationID: FBan0009681

ExpressionAtlas: FBgn0043578

FlyBase: FBgn0043578

FlybaseGene: FBgn0043578

GO **Cellular Component:** extracellular region; microtubule associated complex; integral component of plasma membrane

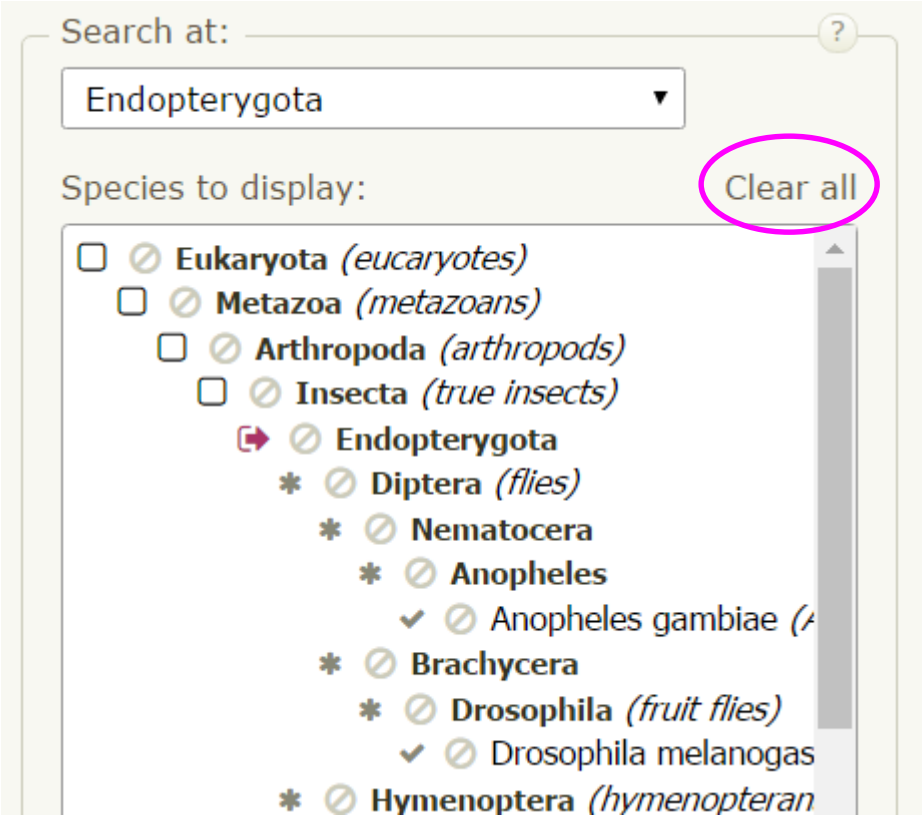
GO **Biological Process:** immune response

GO **Molecular Function:** zinc ion binding; N-acetylmuramoyl-L-alanine amidase activity; peptidoglycan binding

Entrez: PGRP-SB1

www.orthodb.org

OrthoDB



www.orthodb.org

OrthoDB

Select species: ?

Search species by name:

- Eukaryota 588 *(eucaryotes) e.g. S.cerevisiae, C.elegans, M.oryzae, coelaca*
  - Metazoa 330 *(metazoans) e.g. C.elegans, coelacanth, black-legged tick,*
    - Vertebrata 172 *(vertebrates) e.g. coelacanth, platypus, X.tropicalis, e*
      - Tetrapoda 146 *(tetrapods) e.g. platypus, X.tropicalis, elephant, pig,*
        - Mammalia 84 *(mammals) e.g. platypus, elephant, pig, T.truncat*
          - Eutheria 80 *(placentals) e.g. elephant, pig, T.truncatus, cattle*
            - Euarchontoglires 37 *e.g. M.musculus, P.troglodytes, H.sap*
              - Primates 20 *e.g. P.troglodytes, H.sapiens*
                - Cercopithecoidea 8
                - Hominoidea 6 *(apes) e.g. P.troglodytes, H.sapiens*
                - Strepsirrhini 3 *(prosimians)*
                - Aotus nancymaae *(Mas night monkey)*
                - Callithrix jacchus *(white-tufted-ear marmoset)*
                - Tarsius syrichta *(Philippine tarsier)*
              - Glires 14 *(Rodents and rabbits) e.g. M.musculus*
              - Galeopterus variegatus *(Sunda flying lemur)*
              - Tupaia belangeri *(northern tree shrew)*
              - Tupaia chinensis *(Chinese tree shrew)*

www.orthodb.org

OrthoDB

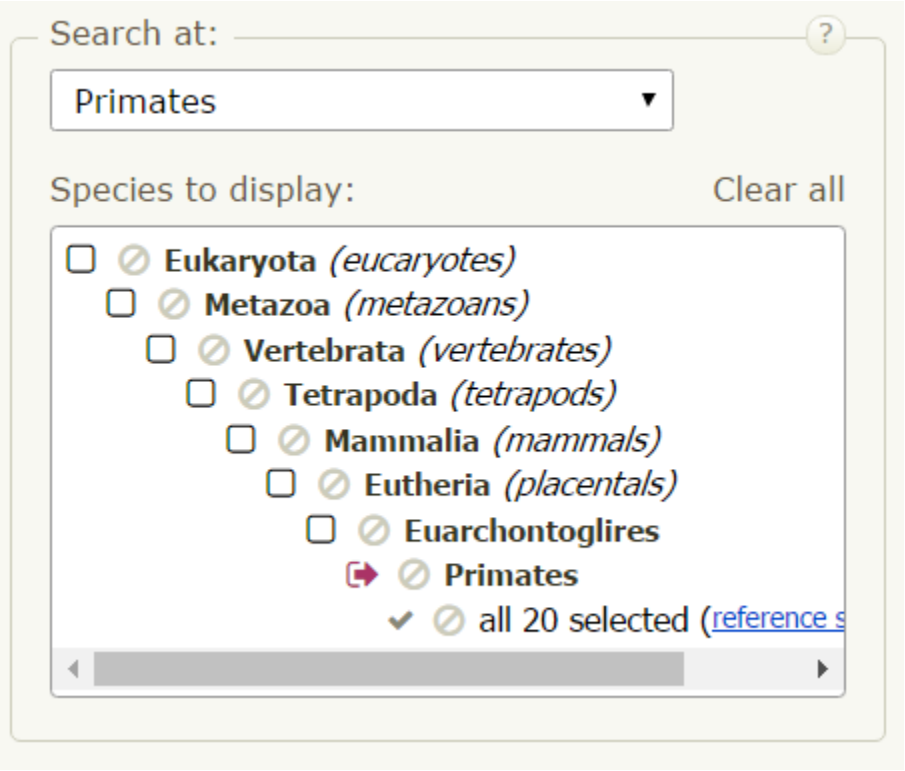
Search species by name:

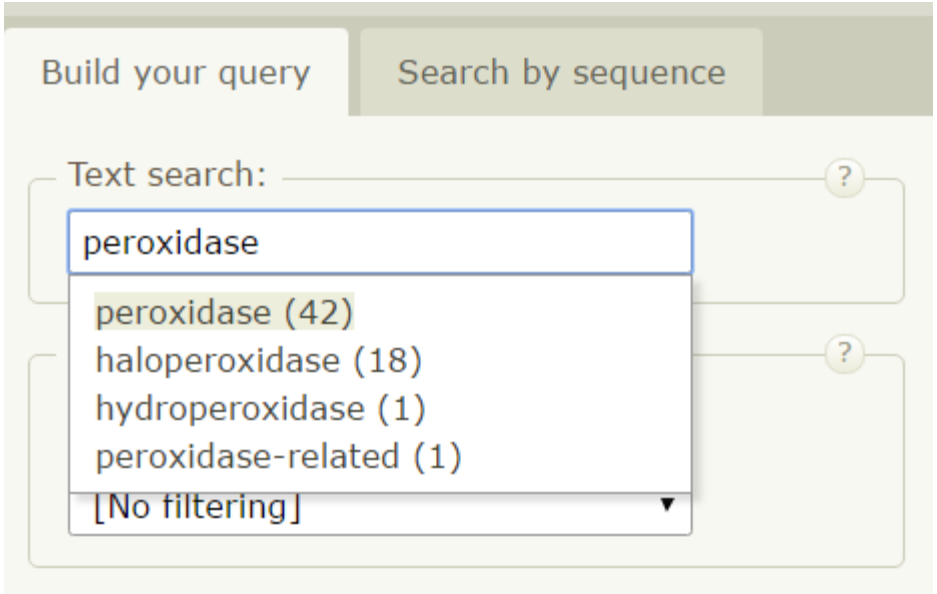
- Eukaryota 588 *(eucaryotes) e.g. S.cerevisiae, C.elegans, M.oryzae, coelac*
  - Metazoa 330 *(metazoans) e.g. C.elegans, coelacanth, black-legged tick*
    - Vertebrata 172 *(vertebrates) e.g. coelacanth, platypus, X.tropicalis,*
      - Tetrapoda 146 *(tetrapods) e.g. platypus, X.tropicalis, elephant, pi*
        - Mammalia 84 *(mammals) e.g. platypus, elephant, pig, T.trunca*
          - Eutheria 80 *(placentals) e.g. elephant, pig, T.truncatus, catt*
            - Euarchontoglires 37 *e.g. M.musculus, P.troglodytes, H.sa*
              - Primates 20 *e.g. P.troglodytes, H.sapiens*
                - Cercopithecoidea 8
                - Hominoidea 6 *(apes) e.g. P.troglodytes, H.sapiens*
                - Strepsirrhini 3 *(prosimians)*
                - Aotus nancymaae *(Mas night monkey)*
                - Callithrix jacchus *(white-tufted-ear marmoset)*
                - Tarsius syrichta *(Philippine tarsier)*
              - Glires 14 *(Rodents and rabbits) e.g. M.musculus*
              - Galeopterus variegatus *(Sunda flying lemur)*
              - Tupaia belangeri *(northern tree shrew)*
              - Tupaia chinensis *(Chinese tree shrew)*

www.orthodb.org

OrthoDB

Search for term 'peroxidase'





www.orthodb.org

OrthoDB

OrthoDB

Your search for **peroxidase** at Primates level returned 42 groups

Bookmark OrthoDB@Primates | Get All Fasta | Get All as Tab delimited ?

Group EOG090O0MYR at Primates level — 20 genes in 19 species
Glutathione peroxidase

Group EOG090O0NO9 at Primates level — 19 genes in 19 species
Glutathione peroxidase

Group EOG090O0OZ9 at Primates level — 19 genes in 19 species
Glutathione peroxidase

Group EOG090O00A0 at Primates level — 20 genes in 20 species
Leucine-rich repeat

Group EOG090O013L at Primates level — 39 genes in 20 species
Similarity:Contains FAD-binding FR-type domain.

Group EOG090O03EV at Primates level — 19 genes in 18 species
thyroid peroxidase

Group EOG090O0NAS at Primates level — 16 genes in 16 species
Glutathione peroxidase

www.orthodb.org

OrthoDB

Phyloprofile:

[No filtering]

[No filtering]

Phyloprofile:

[No filtering]

[No filtering]
Present in all species
Present in >90% species
Present in >80% species

Search at:

Submit

OrthoDB

Your search for **peroxidase** at Primates level returned 15 groups that span all species

Bookmark OrthoDB@Primates | Get All Fasta | Get All as Tab delimited

Group EOG090O00A0 at Primates level — 20 genes in 20 species
Leucine-rich repeat

Group EOG090O013L at Primates level — 39 genes in 20 species
Similarity:Contains FAD-binding FR-type domain.

www.orthodb.org

OrthoDB

Phyloprofile:

Present in all species

[No filtering]

- [No filtering]
- Single-copy in all species
- Single-copy in >90% species
- Single-copy in >80% species

Submit

OrthoDB

Your search for **peroxidase** at Primates level returned 9 groups that span all species and single-copy in all species

Bookmark OrthoDB@Primates | Get All Fasta | Get All as Tab delimited

Group EOG090O00A0 at Primates level — 20 genes in 20 species
Leucine-rich repeat

Group EOG090O012A at Primates level — 20 genes in 20 species
peroxidasin

Group EOG090O07PD at Primates level — 20 genes in 20 species
Prostaglandin G/H synthase 2

Group EOG090O0SFQ at Primates level — 20 genes in 20 species
Synuclein

Expand peroxidasin (EOG090O012A)

Group EOG090O00A0 at Primates level — 20 genes in 20 species

Leucine-rich repeat

Group EOG090O012A at Primates level — View Fasta | View Tab Delimited

peroxidasin

Functional descriptions

Functional Category	M: Cell wall/membrane/envelope biogenesis O: Posttranslational modification, protein turnover, chaperones
GO Molecular Function	11 genes with GO:0005201: extracellular matrix structural constituent 9 genes with GO:0016491: oxidoreductase activity 9 genes with GO:0020037: heme binding

www.orthodb.org

© R.M.Waterhouse

```
>30608:001056 {"pub_gene_id":"ENSMICG00000004588", "pub_og_id":"EOG090O012A", "og_name":"peroxidasin ","level":9443, "description":"peroxidasin "}
YLYKNEIQSIDRQAFQGLASLEQLYLHFNQIETLDPESFQHLPKLERLFLHNNRITHLVPGTFSHLESMKRLRLDSNALRCDCEILWLADLLKTYAQSGNAQAAATCEYPRRIQGRSVATVTPEELDCERPRITSEPQDADVTSGNTVYFTCRAEGNPKPEIIWLRNNNELSMT
TDSRLNLLDDGTLMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGQSVTLECSAAGHPPPRISWTRGDGSPLPADPRVNTTPSGGLYIQNVVQGDSGEYACFASNSVESVQATAFIIVQALPEFTLTPQDRAVIEGQTVDFQCEAKGNPQPV
IAWTKGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXYHYNDLVSPQYLHLIANLSGCTAHRRVNNCSDMCFHQRYRTHDGTCNNLQHPMWGASLTAFERLLKAVYENGFNTPRGINPHRLYNGHALPAPRLVSTTLIGAEAVTPD
AQFTHMLMQWGQFLDHDLDSTVVALSQARFSDGQHCSSACSSDPPCFSVMIPPNDPRARGGGGGARCMFFVRSSPVCGSGMTSLLMNSVYPREQINQLTSYIDASNVYGSTDHEARGIRDLASHRGLLRQGIVQRSGKPLLPFAAGPPTECMRDENESPIPCFLAGDHRANEQL
GLTSMHTLWFREHNRIAAELLALNPHWDGDTIYHETRKIVGAEVQHITYQHWLPKVLGEVGMKMLGEYRGYDPGVNAGIFNAFATAAFRFGHTLVNPVLHRLDENFQAIPQGPLPLHKAFFSPFRIVNEGGIDPLLRGLFGVAGKMRVPSQLLNTELTERLFSMAHTVALDLAA
INIQRGRDHGIPPYHDYRVYCNLSAAHTFEDLKNEIKNPEIREKLKRLYGSTLNIDLFPALMVEDLVPGSRLGPTLMCLLSAQFRRLRDGDRLWYENPGVFSPAQLTQIKQTSLARILCDNADNITRVQRDVFRVAEFPHGYGDCGEVPRVDLRVWQDCCEDCRTRXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXAGKHGTPLGNLTSAFREHAKGPGTNDFREFVQEMQKTITDLRKQIKKLESRLSATECLDASGEPHANHAKWKTDACTLCECRDGQVTCFVEACQPATCPAPVRVAGACCPVCSQNAEEDTP
>30611:0036f0 {"pub_gene_id":"PXDN", "pub_og_id":"EOG090O012A", "og_name":"peroxidasin ","level":9443, "description":"Similarity:Contains LRRCT domain."}
MAARPTGLARRCLLALTLCCCWGALAVVAQKPGAGCPSRCLCFRTTVRCMHLLLETVPAVAPQTSILDLRFNRIREIQPGAFRRLRNLNTLLLNNNQIKRIPSGAFEDLENLKYLYLYKNEIQSIDREAFKGLASLEQLYLHFNQIETLDPESFQHLPKLERLFLHNNRITHLV
PGTFNHLESMKRLRLDSNALHCDCEILWLADLLKTYAQSGNAQAAATCEYPRRIQGRSVATVTPEELDCERPRITSEPQDADVTSGNTVYFTCRAEGNPKPEIIWLRNNNELSMKTDSRLNLLDDGTLMIRNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQP
QNTEVLVGESVTLECSATGHPLPRISWTRGDRTPLPADPRVNITPSGGLYIQNVVQEDSGEYTCFASNSVESVHATAFIIVQEALPEFTVTPQDRVVIEGQTVDFQCEAKGNPQPVIAWTKGGVRQEILRSSVVTGGHGCLPVTPSRRHGVGTGDYPSVTLSGSQKVVAHLTVQ
PRVTPVFASVPSDMTVEVGTNVQLPCSSQGEPEPAITWNKDGVQVTESGKFHISPEGFLTINDVGTADAGRYECVARNTIGSASVSVVLSVNVPDVSRNGDPFVATSIVEAIATVDRAINSTRTHLFDSRPRSPNDLLALFRYPRDPQTSGSLSQPGGVFTRSLTFLCARFGHW
SHAPWKLSPMAGYHYNDLVSPQYLHLIANLSGCTAHRRVNNCSDMCFHQRYRAHDGTCNNLQRPMWGASLTAFERLLKSVYENGFNTPRGINPHRLYNGHALPAPRLVSTTLIGTEAVTPDEQFTHMLMQWGQFLDHDLDSTVVALSQARFSDGQHCSSACSSDPPCFSVLIPP
NDPRARSGARCMFFVRSSPVCGSGMTSLLMNSVYPREQINQLTSYIDASNVYGSTEHEARGIRDLASHRGLLRQGIVQRSGKPLLPFAAGPPTECMRDESESPIPCFLAGDHRANEQLGLTSMHTLWFREHNRIAAELLRLNPHWDGDTIYHEARKLVGAQMQHITYQHWLPKV
LGEVGMKMLGEYRGYDPGVNAGIFNAFATAAFRFGHTLVNPVLYRLDENFQPIAQGHLPLHKAFFSPFRIVNEGGIDPLLRGLFGVAGKMRVPSQLLNMELTERLFSMAHTVALDLAAINIQRGRDHGIPPYHDYRVYCNLSAAHTFEDLKNEIKNPEIREKLKRLYGSTLNID
LFPALVVEDLVPGSRLGPTLMCLLSTQFRRLRDGDRLWYENPGVFSPAQLTQIRQTSLARILCDNSDNITRVQRDVFRVAEFPHGYGSCDEVPRVDLRVWQDCCEDCRTRGQFNVFSHHFRGRRSVEFSYQEDKPAKKRQPREMLSFGKHGPRLSNATSALGKHSNGLGTNDFK
EFVLEMQKTITDLRKQVRAAGSASFPLGTRPASCTRDTQPRAWEGLQWAPSAARDLKCRNGQVTCFVEACQPVACLAPVKTEGACCPVCSQNTAEGKP
>37293:003afe {"pub_gene_id":"105726757", "pub_og_id":"EOG090O012A", "og_name":"peroxidasin ","level":9443, "description":"peroxidasin"}
MGDLRFNRIREIQPGAFRRLRNLNTLRSVCILIQPFMLKPSPCVNAVLLRLKFDLLSVLMLPNVSCFSYLYKNEIQSIDRQAFKGLASLEQLYLHFNQIETLDPESFQHLPKLERLFLHNNRITHLVPGTFNHLESMKRLRLDSNTLHCDCEILWLADLLKTYAESGNAQAAAT
CEYPRRIQGRSVATITPEELNCERPRITSEPQDADVTSGNTVYFTCRAEGNPKPEIIWLRNNNELSMKTDSRLNLLDDGTLMIQNTQETDQGVYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTLECSATGHPPPRISWTRGDHTPLPVDPRVNITPSG
GLYIQNVVQEDSGEYACSATNSIDSVHATAFIIVQALPQFTVTPQDRVVIEGQTVDFQCEAKGNPPPVIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPRVTPVFASIPSDTTVEVGANVQLPCSSQGEPEPAITWNKDGVQVTESG
KFHISPEGFLTINDVGPADAGRYECVARNTIGAASVSMVLSVNVPDVSRNGDPFVATSIVEAIATVDRAINSTRTHLFDSRPRSPNDLLALFRYPRDPYTVEQARAGEIFERTLQLIQEHVQHGLMVDLNGTSYHYNDLVSPQYLNLIANLSGCTAHRRVNNCSDMCFHQKYRT
HDGTCNNLQHPMWGASLTAFERLLKSVYENGFNTPRGINPHRLYHGHALPMPRLVSTTLIGAESVTRDEQFTHMLMQWGQFLDHDLDSTVVALSQARFSDGQHCSSACSNDPPCFSVMIPHNDSRARSGARCMFFVRSSPVCGSGMTSLLMNSVYPREQINQLTSYIDASNVYG
STEHEARSIRDLASHRGLLRQGIVQRSGKPLLPFAAGPPTECMRDENESPIPCFLAGDHRANEQLGLTSMHTLWFREHNRIAAELLQLNPHWDGDTIYYETRKIVGAEIQHITYQHWLPKILGEVGMRALGEYRGYDPGVNAGIFNAFATAAFRFGHTLIHPLLYRLDENFQPI
AQGHIPLHKAFFSPFRIVNEGGIDPLLRGLFGVAGKMRVPSQLLNTELTERLFSMAHTVALDLAAINIQRGRDHGIPPYHDYRVYCNLSAAHTFEDLKNEIKNPEIREKLQRLYGSTLNIDLFPALMVEDLVPGSRLGPTLMCLLSTQFKRLRDGDRLWYENPGVFSPAQLTQI
KQTSLARILCDNADNITRVQSDVFRVAEFPHGYGSCGEIPRVDLRVWQDCCQGACQGRHACSFFIPGVGDCLAEASCVPAEASGACCMQRRCPGTLGPSSPQRRHVSPQRRQVPAACSDVVLAHRVPLPRRGVRCPRRGVRCLLHAATLSWRTGSLFPAEASGVPAEASGACCM
QRRCPGAPGPSSPQRRQVSPQRHQVPAACSDGVLVHRVPLPRRGVRCPRRGIRCRLRAATVSWCTGSLFPAEASGVPAEASGACCMQRRCPGAPGPSSPQRRQVSPQRHQRRQVPAACSDGVLAHPVPLPRRGVRCPRRGIMCPRRGVRCLLRAATVSWRTRSLFSAEASVACC
SWHWFLWLNSCFAFLTPSLG
>379532:0041e3 {"pub_gene_id":"105824647", "pub_og_id":"EOG090O012A", "og_name":"peroxidasin ","level":9443, "description":"peroxidasin"}
MAARPASPARRCLLAIVLCCGWGALAVVAPKPGAGCPSRCLCFRTTVRCMHLLLEAVPAVAPQTSILDLRFNRIREIQPGAFRRLRNLNTLLLNNNQIRRIPGGAFEDLENLKYLYLYKNEIQAIDRQAFQGLASLEQLYLHFNQIETLDPESFQRLPKLERLFLHNNRITHLV
PGTFNHLESMKRLRLDSNALRCDCEILWLADLLKAYAQSGNAQAAATCEYPRRVQGRSVATVTPEELDCERPRITSEPQDADVTSGNTVYFTCRAEGNPKPEIIWLRNNNELSMTTDSRLNLLDDGTLMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPAFVIQP
QNTEVLVGESVTLECSAAGHPPPRISWTRGDGSPLPVDPRVNITPSGGLYIQNVVQGDSGEYTCFASNSVESVHATAFIIVQALPEFTVTPQDRVVIEGQTVDFQCEAKGNPQPVIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR
```

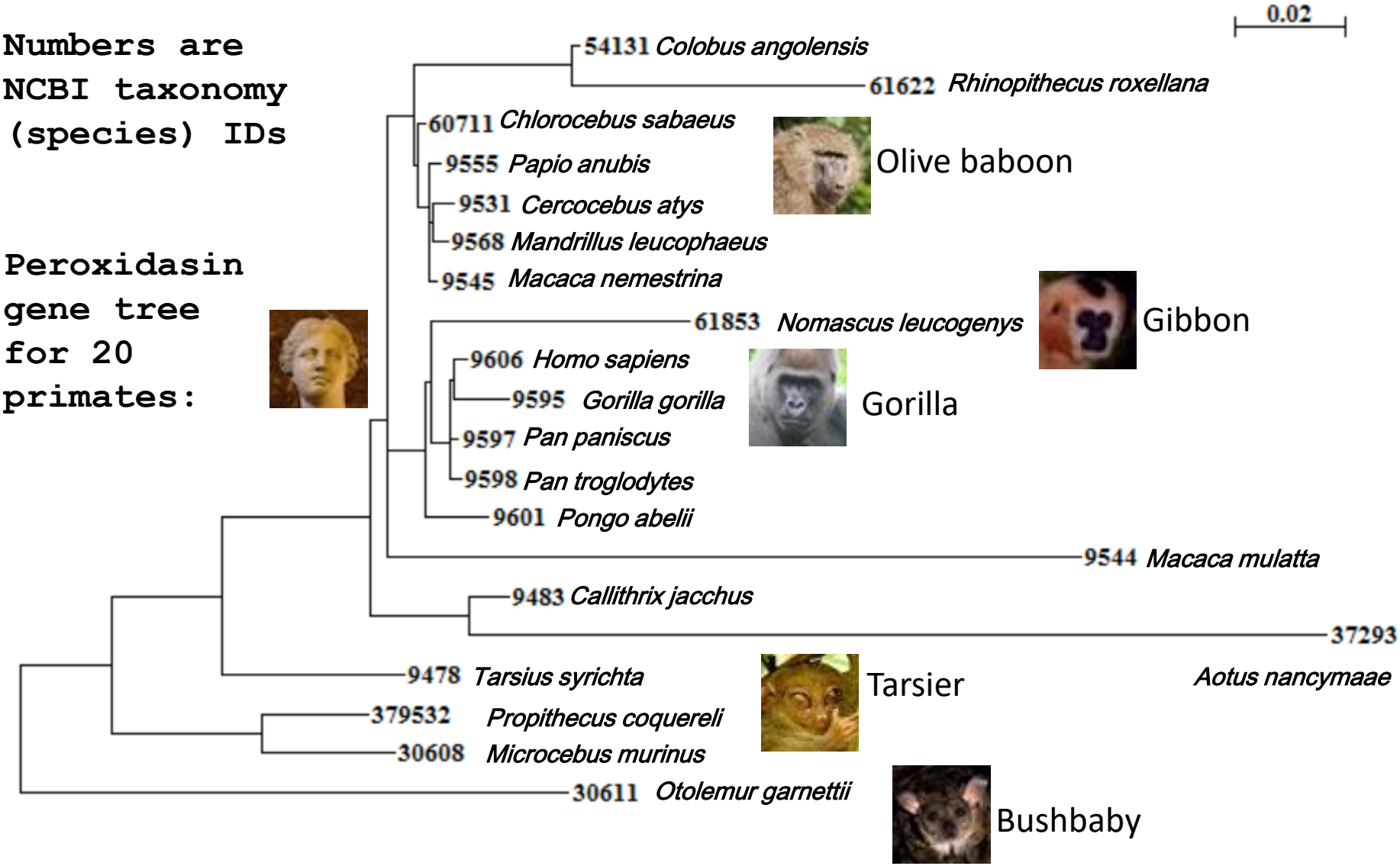
www.orthodb.org

OrthoDB

Njplot: http://doua.prabi.fr/software/njplot

Numbers are NCBI taxonomy (species) IDs

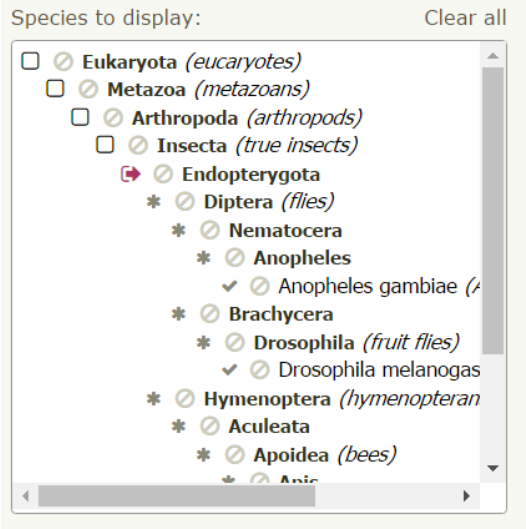
Peroxidasin gene tree for 20 primates:



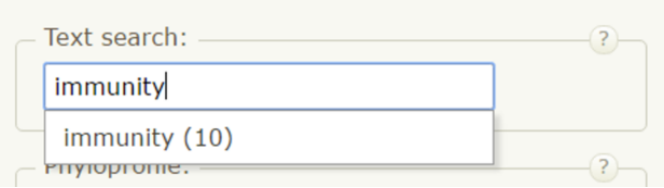
www.orthodb.org

OrthoDB

Species selection

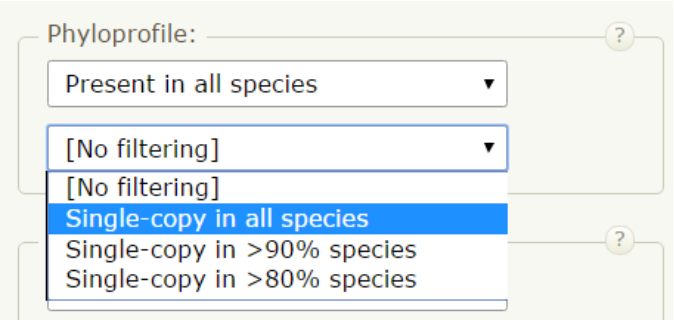


Keyword searching



Annotation browsing

Phyloprofile filtering



Data downloading

UniL SIB 3-5 September 2018

Link for OrthoDB quiz

https://goo.gl/forms/Kfcbfnz4NinymV7y2